Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 19 (2017), Issue 6   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Web Use for Symptom Appraisal of Physical Health Conditions: A Systematic Review (e202)
Julia Mueller, Caroline Jay, Simon Harper, Alan Davies, Julio Vega, Chris Todd. 5

A Review of the Theoretical Basis, Effects, and Cost Effectiveness of Online Smoking Cessation Interventions in the Netherlands: A Mixed-Methods Approach (e230)
Kei Cheung, Ben Wijnen, Hein de Vries. 101

Web-Based Interventions for Weight Loss or Weight Loss Maintenance in Overweight and Obese People: A Systematic Review of Systematic Reviews (e229)
Angela Sorgente, Giada Pietrabissa, Gian Manzoni, Federica Re, Susan Simpson, Sara Perona, Alessandro Rossi, Roberto Cattivelli, Marco Innamorati, Jeffrey Jackson, Gianluca Castelnuovo. 115

Researching Mental Health Disorders in the Era of Social Media: Systematic Review (e228)
Akkapon Wongkoblap, Miguel Vadillo, Vasa Curcin. 325

Viewpoint

Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop (e232)
Susan Michie, Lucy Yardley, Robert West, Kevin Patrick, Felix Greaves. 27

Original Papers

Defining and Predicting Patterns of Early Response in a Web-Based Intervention for Depression (e206)
Wolfgang Lutz, Alice Arndt, Julian Rubel, Thomas Berger, Johanna Schröder, Christina Späth, Björn Meyer, Wolfgang Greiner, Viola Gräfe, Martin Hautzinger, Kristina Fuhr, Matthias Rose, Sandra Nolte, Bernd Löwe, Fritz Hohagen, Jan Klein, Steffen Moritz. 40

Using Feedback to Promote Physical Activity: The Role of the Feedback Sign (e192)
Jan-Niklas Kramer, Tobias Kowatsch. 56

Young People, Adult Worries: Randomized Controlled Trial and Feasibility Study of the Internet-Based Self-Support Method “Feel the ViBe” for Adolescents and Young Adults Exposed to Family Violence (e204)
Karin van Rosmalen-Nooijens, Sylvie Lo Fo Wong, Judith Prins, Toine Lagro-Janssen. 67

Effectiveness of a Web-Based Intervention in Reducing Depression and Sickness Absence: Randomized Controlled Trial (e213)
Till Beiwinkel, Tabea Eißing, Nils-Torge Telle, Elisabeth Siegmund-Schultze, Wulf Rössler. 87

Web-Based Alcohol Intervention: Study of Systematic Attrition of Heavy Drinkers (e217)
Theda Radtke, Mathias Ostergaard, Richard Cooke, Urte Scholz. 131

Informing the Design of “Lifestyle Monitoring” Technology for the Detection of Health Deterioration in Long-Term Conditions: A Qualitative Study of People Living With Heart Failure (e231)
Sarah Hargreaves, Mark Hawley, Annette Haywood, Pamela Enderby. 143

Mobile Text Messaging to Improve Medication Adherence and Viral Load in a Vulnerable Canadian Population Living With Human Immunodeficiency Virus: A Repeated Measures Study (e190)
Elizabeth King, Karen Kinvig, Jonathan Steif, Annie Qiu, Evelyn Maan, Arianne Albert, Neora Pick, Ariane Alimenti, Mary Kestler, Deborah Money, Richard Lester, Melanie Murray. 155

Tailored Communication Within Mobile Apps for Diabetes Self-Management: A Systematic Review (e227)
Heidi Holmen, Astrid Wahl, Milada Cvancarova Småstuen, Lis Ribu. 170

Tracing the Potential Flow of Consumer Data: A Network Analysis of Prominent Health and Fitness Apps (e233)
Quinn Grundy, Fabian Held, Lisa Bero. 186

MyGeneFriends: A Social Network Linking Genes, Genetic Diseases, and Researchers (e212)
Alexis Allot, Kirsley Chennen, Yannis Nevers, Laetitia Poidevin, Arnaud Kress, Raymond Ripp, Julie Thompson, Olivier Poch, Odile Lecompte.
1 9 5

What Motivates Young Adults to Talk About Physical Activity on Social Network Sites? (e226)
Ni Zhang, Shelly Campo, Jingzhen Yang, Petya Eckler, Linda Snetselaar, Kathleen Janz, Emily Leary. 209

Scaling Up Psychological Treatments: A Countrywide Test of the Online Training of Therapists (e214)
Christopher Fairburn, Elizabeth Allen, Suzanne Bailey-Straebler, Marianne O'Connor, Zafra Cooper. 219

Using Social Media to Characterize Public Sentiment Toward Medical Interventions Commonly Used for Cancer Screening: An Observational Study (e200)
Omar Metwally, Seth Blumberg, Uri Ladabaum, Sidhartha Sinha. 226

Frequencies of Private Mentions and Sharing of Mammography and Breast Cancer Terms on Facebook: A Pilot Study (e201)
Marco Huesch, Alison Chetlen, Joel Segel, Susann Schetter. 236

Too Far to Care? Measuring Public Attention and Fear for Ebola Using Twitter (e193)
Liza van Lent, Hande Sungur, Florian Kunneman, Bob van de Velde, Enny Das. 247

The Role of Web-Based Health Information in Help-Seeking Behavior Prior to a Diagnosis of Lung Cancer: A Mixed-Methods Study (e189)
Julia Mueller, Caroline Jay, Simon Harper, Chris Todd. 257

Dr Google Is Here to Stay but Health Care Professionals Are Still Valued: An Analysis of Health Care Consumers’ Internet Navigation Support Preferences (e210)
Kenneth Lee, Kreshnik Hoti, Jeffery Hughes, Lynne Emmerton. 273

Classifying Chinese Questions Related to Health Care Posted by Consumers Via the Internet (e220)
Haihong Guo, Xu Na, Li Hou, Jiao Li. 282

Interactive Voice Response System: Data Considerations and Lessons Learned During a Rectal Microbicide Placebo Adherence Trial for Young Men Who Have Sex With Men (e207)
Jose Bauermeister, Rebecca Giguere, Cheng-Shiun Leu, Irma Febo, Ross Cranston, Kenneth Mayer, Alex Carballo-Diéguez. 294

Social Media Use and Access to Digital Technology in US Young Adults in 2016 (e196)
Andrea Villanti, Amanda Johnson, Vinu Ilakkuvan, Megan Jacobs, Amanda Graham, Jessica Rath. 302

What Predicts Online Health Information-Seeking Behavior Among Egyptian Adults? A Cross-Sectional Study (e216)
Mayada Ghweeba, Antje Lindenmeyer, Sobhi Shishi, Mostafa Abbas, Amani Waheed, Shaymaa Amer. 316

Access to Electronic Personal Health Records Among Patients With Multiple Chronic Conditions: A Secondary Data Analysis (e188)
Alexandra Greenberg, Angela Falisi, Lila Finney Rutten, Wen-Ying Chou, Vaishali Patel, Richard Moser, Bradford Hesse. 342

A Qualitative Analysis of How Online Access to Mental Health Notes Is Changing Clinician Perceptions of Power and the Therapeutic Relationship (e208)
Lauren Denneson, Risa Cromer, Holly Williams, Maura Pisciotta, Steven Dobscha. 356

The Second Victim Phenomenon After a Clinical Error: The Design and Evaluation of a Website to Reduce Caregivers' Emotional Responses After a Clinical Error (e203)
José Mira, Irene Carrillo, Mercedes Guilabert, Susana Lorenzo, Pastora Pérez-Pérez, Carmen Silvestre, Lena Ferrús, Spanish Second Victim Research Team. 363

The Gap in Medical Informatics and Continuing Education Between the United States and China: A Comparison of Conferences in 2016 (e224)
Jun Liang, Kunyan Wei, Qun Meng, Zhenying Chen, Jiajie Zhang, Jianbo Lei. 377

The Effectiveness of Information Technology-Supported Shared Care for Patients With Chronic Disease: A Systematic Review (e221)
Laura Kooij, Wim Groen, Wim van Harten. 385

Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research (e218)
Laura Sbaffi, Jennifer Rowley. 396

Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review (e195)
Su Golder, Shahd Ahmed, Gill Norman, Andrew Booth. 413

A Medical Student-Delivered Smoking Prevention Program, Education Against Tobacco, for Secondary Schools in Germany: Randomized Controlled Trial (e199)
Titus Brinker, Andreas Owczarek, Werner Seeger, David Groneberg, Christian Brieske, Philipp Jansen, Joachim Klode, Ingo Stoffels, Dirk Schadendorf, Benjamin Izar, Fabian Fries, Felix Hofmann. 432

Integrated Detection and Prediction of Influenza Activity for Real-Time Surveillance: Algorithm Design (e211)
Armin Spreco, Olle Eriksson, Örjan Dahlström, Benjamin Cowling, Toomas Timpka. 443

"You Sort of Go Down a Rabbit Hole...You're Just Going to Keep on Searching": A Qualitative Study of Searching Online for Pregnancy-Related Information During Pregnancy (e194)
Julie Prescott, Lynn Mackie. 454

Improving Consensus Scoring of Crowdsourced Data Using the Rasch Model: Development and Refinement of a Diagnostic Instrument (e222)
Christopher Brady, Lucy Mudie, Xueyang Wang, Eliseo Guallar, David Friedman. 465

“Click for Closer Care”: A Content Analysis of Community Pharmacy Websites in Four Countries (e205)
Sandra Zwier. 478

Designing, Prototyping and Evaluating Digital Mindfulness Applications: A Case Study of Mindful Breathing for Stress Reduction (e197)
Bin Zhu, Anders Hedman, Shuo Feng, Haibo Li, Walter Osika. 495

Letter to the Editor

The Research Topic Defines “Noise” in Social Media Data – a Response from the Authors (e165)
Yoonsang Kim, Jidong Huang, Sherry Emery. 509

Corrigenda and Addenda

Metadata Correction: Importance-Performance Analysis of Personal Health Records in Taiwan: A Web-Based Survey (e209)
Hsiao-Hsien Rau, Yi-Syuan Wu, Chi-Ming Chu, Fu-Chung Wang, Min-Huei Hsu, Chi-Wen Chang, Kang-Hua Chen, Yen-Liang Lee, Senyeong Kao, Yu-Lung Chiu, Hsyien-Chia Wen, Anis Fuad, Chien-Yeh Hsu, Hung-Wen Chiu. 511

Review

Web Use for Symptom Appraisal of Physical Health Conditions: A Systematic Review

Julia Mueller[1,2,3], MSc; Caroline Jay[3], PhD; Simon Harper[3], PhD; Alan Davies[3], BSc; Julio Vega[3], MSc; Chris Todd[1,2], PhD

[1]School of Health Sciences, University of Manchester, Manchester, United Kingdom
[2]Manchester Academic Health Science Centre, Manchester, United Kingdom
[3]School of Computer Science, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Julia Mueller, MSc
School of Health Sciences
University of Manchester
Kilburn Building, LF1
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 1612756239
Fax: 44 1613067894
Email: julia.mueller@manchester.ac.uk

Abstract

Background: The Web has become an important information source for appraising symptoms. We need to understand the role it currently plays in help seeking and symptom evaluation to leverage its potential to support health care delivery.

Objective: The aim was to systematically review the literature currently available on Web use for symptom appraisal.

Methods: We searched PubMed, EMBASE, PsycINFO, ACM Digital Library, SCOPUS, and Web of Science for any empirical studies that addressed the use of the Web by lay people to evaluate symptoms for physical conditions. Articles were excluded if they did not meet minimum quality criteria. Study findings were synthesized using a thematic approach.

Results: A total of 32 studies were included. Study designs included cross-sectional surveys, qualitative studies, experimental studies, and studies involving website/search engine usage data. Approximately 35% of adults engage in Web use for symptom appraisal, but this proportion varies between 23% and 75% depending on sociodemographic and disease-related factors. Most searches were symptom-based rather than condition-based. Users viewed only the top search results and interacted more with results that mentioned serious conditions. Web use for symptom appraisal appears to impact on the decision to present to health services, communication with health professionals, and anxiety.

Conclusions: Web use for symptom appraisal has the potential to influence the timing of help seeking for symptoms and the communication between patients and health care professionals during consultations. However, studies lack suitable comparison groups as well as follow-up of participants over time to determine whether Web use results in health care utilization and diagnosis. Future research should involve longitudinal follow-up so that we can weigh the benefits of Web use for symptom appraisal (eg, reductions in delays to diagnosis) against the disadvantages (eg, unnecessary anxiety and health care use) and relate these to health care costs.

(J Med Internet Res 2017;19(6):e202) doi:10.2196/jmir.6755

KEYWORDS

Online health information; health information seeking; Internet; symptom appraisal; Web search; search strategies

Introduction

The Web has become an important resource for lay information about health, with almost three-quarters of the population in developed countries accessing the Web to research health topics [1,2]. The Web is accessed to obtain information on general health topics, such as weight management, and by patients to obtain information on their diagnosed condition [3,4]. Moreover, the Web is accessed to assess and evaluate symptoms and their causes [1].

The way the Web is used by patients who have obtained a specific diagnosis from a health care professional is likely to differ from the way it is used in the absence of a professional diagnosis when appraising symptoms. Postdiagnosis, individuals have specific medical terms they can use as search terms. Most focus their Web search on treatment options, illness management, and prognosis [3,4]. When appraising symptoms with the aim of diagnosing them, on the other hand, most individuals have only symptoms and lay medical knowledge to guide their search, and symptoms are sometimes vague and difficult to describe.

Web use for symptom appraisal may have important implications, although it is unclear whether it plays a beneficial or detrimental role in health care delivery. For example, some evidence suggests it could lead to unnecessary anxiety about health and increase use of health service resources [5]. Other findings imply it could enhance patient empowerment and help patients prepare for consultations [6]. Thus, Web use for symptom appraisal may lead to either wasting or more efficient use of resources. For example, Web use for symptom appraisal may cause anxiety about health by making individuals falsely believe they have a serious condition when they do not. On the other hand, it may encourage people with warning signs to present to health services, promoting earlier diagnosis. It could also falsely reassure people that symptoms are not serious, thus preventing earlier diagnosis. This is particularly important for potentially life-threatening or debilitating conditions, which are easier to treat when detected early, such as cancer [7], heart disease [8], or glaucoma [9]. Understanding Web use for symptom appraisal is also relevant for less serious conditions, such as a common cold, because it could lead people with mild symptoms to present to health services when this is not necessary or it could help people identify the symptoms that can be treated at home. As these examples highlight, Web use for symptom appraisal may have important implications for health care utilization.

To leverage the potential for reducing strain on health care resources and promoting earlier diagnosis, we need to understand the current role of the Web in help seeking and symptom evaluation, and the strategies people use to access information, taking differing contexts into account. Because these questions cannot be addressed in a single study, a systematic review is required, involving a thorough and comprehensive search of the literature, critical appraisal of individual studies, and extraction and synthesis of relevant findings.

This systematic review addresses the following five review questions:

1. What proportion of different populations (eg, general, specific disease, or demographic groups) use the Web to appraise symptoms?
2. Which symptoms are likely to be researched online?
3. How is Web use for symptom appraisal conducted (search strategies)?
4. What are the behavioral consequences of Web use for symptom appraisal?
5. What are the emotional consequences of Web use for symptom appraisal?

Methods

A protocol was developed by the research team based on the review questions and an initial broad search of the available literature, using the Centre for Reviews and Dissemination’s guidance for undertaking reviews in health care [10] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [11].

Eligibility Criteria

Study Focus

We included studies that addressed use of the Web to appraise symptoms (ie, to research symptoms and their potential causes). This could include both actual symptoms and symptoms in fictional scenarios. This did not have to be the primary focus of the study; some reference to Web use for symptom appraisal was sufficient. If studies examined health-related Web use in general, they were screened during full-text review and excluded if no specific reference to symptom appraisal was made. Studies that analyzed anonymous logs were included if they examined symptom-related searches. We included only studies that focused on human behavior; studies that evaluated the performance of Web-based tools were excluded.

Populations

Studies on Web use for symptom appraisal of any physical health conditions were included. Studies examining mental health/psychiatric conditions were excluded to focus the scope of the review. Studies on Web use by health professionals were excluded. Studies from any country were included, as long as the publication was written in English.

Study Design

Our initial scoping suggested a scarcity of research in this area, thus we did not limit included studies to any particular design. Nonempirical studies (eg, theoretical papers and literature reviews) were excluded.

Publication Types

Full paper, English-language publications were included, regardless of the original language of the research.

Information Sources

We searched PubMed, EMBASE, PsycINFO, ACM Digital Library, SCOPUS, and Web of Science for relevant publications up to September 30, 2016. To minimize publication bias, grey literature was explored by searching OpenGrey, an open-access database containing more than 700,000 bibliographical references of grey literature. We also searched the British Library Integrated catalog, which contains reports, conference abstracts, and theses. Finally, authors in the field were contacted to inquire about any unpublished material, if two or more of their papers were among the included studies, or if their paper was judged as particularly relevant to the review (eg, if examining Web use for symptom appraisal was the primary focus of the study).

Search

The terms Internet, Web, online, search engine, Google, help seeking, health information seeking, symptom, and diagnosis were entered into the databases (note Google was used as a search term because this is by far the most widely used search engine worldwide [12]). An example search strategy is provided in Multimedia Appendix 1.

Study Selection

The study selection process followed the guidelines provided in the PRISMA statement [11]. Search terms were entered into the databases and all returned studies were imported into a single Mendeley file. Three independent reviewers assessed the studies for eligibility. Studies were first screened by titles and abstracts. Selected studies were then screened for inclusion by reading full texts. Reference lists of included studies were handsearched for further eligible studies. We also handsearched journals if they contained two or more articles included in this review or if the general journal topic area was particularly relevant to the review, to ensure inclusion of studies not yet loaded on electronic databases. Any discrepancies between the reviewers were discussed until consensus was reached.

Data Collection Process and Data Items

From each study, any information regarding use of the Web for symptom appraisal was extracted, as well as details on study design, procedure, population, sampling method, entry and inclusion criteria for study participants, sample size, measures, and details of analysis methods (data extraction sheet in Multimedia Appendix 2).

Quality Appraisal

A quality appraisal of selected articles was conducted based on five criteria designed for reviews incorporating mixed study designs [13] (Multimedia Appendix 3). Quality appraisal involved two stages. First, articles were assessed for inclusion in the review using a relatively liberal threshold; articles were scored eligible if they addressed each criterion at least minimally. Criteria were then applied more rigorously using a three-point scoring system (low/medium/high; see Multimedia Appendix 3) and main limitations of each study were identified. This assessment was used to critically appraise studies during synthesis of the findings.

Synthesis of Results

The extracted data were synthesized using Thematic Analysis, which has been identified as one of the main approaches used to review and synthesize qualitative and quantitative evidence [14,15]. Our analysis involved the following steps [16]:

1. Data familiarization: familiarization with the data was achieved by reading all included studies several times and extracting the relevant information into the data extraction sheets.
2. A priori grouping: data from the data extraction sheets were grouped according to the review question they pertained to and summarized in a matrix. Studies were entered into the rows of the matrix, whereas study characteristics, limitations, and review questions were entered into the columns. This matrix enabled us to compare the findings of different studies pertaining to the same review question, taking methodological aspects into account (example matrix shown in Multimedia Appendix 4). This method was adapted from Framework Analysis, which is a specific form of Thematic Analysis [17].
3. Generation of initial codes: the data were initially coded using semantic codes within the NVivo10 environment, using the matrix to compare results across studies.
4. Searching for themes: once all data extracts were coded, codes were sorted into broader, more conceptual categories to create themes.
5. Reviewing themes: finally, we reviewed the data extracts the themes related to, to determine whether the created themes satisfactorily captured the raw data.

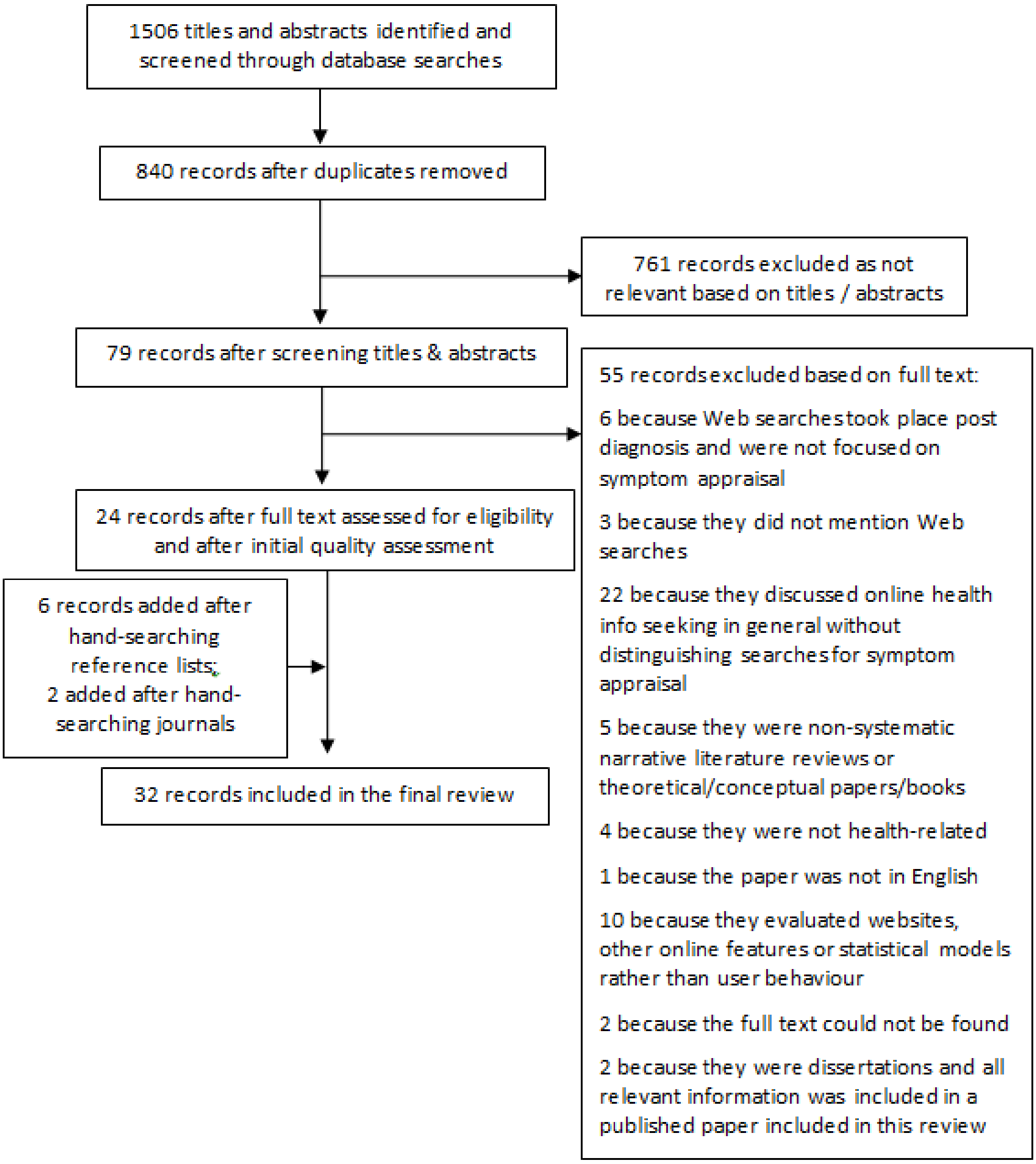
For quantitative studies reporting proportions without confidence intervals, 95% confidence intervals were computed using the asymptotic (Wald) method based on a normal approximation [18] to facilitate comparisons between studies.

Results

Study Selection

Thirty-two studies were identified as eligible for inclusion in the review. The search process is illustrated in Figure 1. The grey literature search yielded no further inclusions. The *Journal of Medical Internet Research*, the *Journal of Health Communication*, *Telemedicine and e-Health*, and the *Journal of the American Medical Informatics Association* were handsearched, resulting in 15 full-text assessments and two further inclusions. Four authors were contacted to enquire about unpublished material. We received one reply, concerning an article we had already identified.

Figure 1. PRISMA diagram for the study identification process.



Study Characteristics

Tables 1-3 provide an overview of study characteristics. Study designs included cross-sectional surveys (n=10, two of these with embedded qualitative interview studies, one with an embedded observational study), qualitative interview and focus group studies (n=4), experimental studies (n=7), two studies evaluating questions posed by users to a health website (n=2), a study evaluating clicks on a specific health website (n=1), and studies involving the analysis of log data from search engines (n=8). Two of these combined a log-based approach with a survey. Most studies were conducted in the United States (n=25).

Table 1. Study design and aim of the studies included in the review (N=32).

Author, date	Study design	Aim
Attfield et al, 2006 [6]	Qualitative interview study; cross-sectional	Explore information seeking of patients before and after consultations, its situational influences, and its impact on patient-provider relationships
Briet et al, 2014 [19]	Quantitative; cross-sectional analysis of website queries	Explore the nature and content of questions and answers on a health website, and to examine the situations of patients asking questions
Cartrightet et al, 2011 [20]	Longitudinal log-based study	Analyze the search activity of users researching health information online and identify goals and patterns of search behavior
Chin, 2009 [21]	Experimental between subjects design: 2×2 (ill–well-defined tasks, younger-older users)	Compare older and younger adults in their performance and search behavior in ill and well-defined tasks
Chin & Fu, 2010 [22]	Experimental between subjects design: 2×2×2 (older-younger adults, parts-systems interface, parts-system task)	Examine differences between older and younger adults in interacting with different online search tasks and interfaces
Cooper et al, 2013 [23]	Qualitative study (focus groups)	Explore how women would evaluate symptoms associated with gynecologic cancers
Cumming et al, 2010 [24]	Cross-sectional Web-based survey study	Evaluate digital storytelling videos (videos of people talking about their own experiences) about help seeking for menopausal symptoms
De Choudhury et al, 2014 [25]	Cross-sectional survey study (quantitative + qualitative data) + longitudinal log-based study	Research the prevalence of health activities on social media and search engines; characterize health activities on the different platforms and describe how people evaluate information obtained from these
Fiksdal et al, 2014 [26]	Qualitative focus group study	To gain a deeper understanding of online health-searching behavior to inform future developments of personalizing information searching and content delivery.
Fox & Duggan, 2013 [1]	Nationwide cross-sectional survey	The Pew Internet & American Life Project is an initiative of the Pew Research Center, a nonprofit “fact tank” that provides information on the issues, attitudes and trends shaping America and the world
Hay et al, 2008 [27]	Mixed-methods survey and interview study	Understand the extent and reasons for online research prior to first appointments for patients in a rheumatology clinic
Keselman et al, 2008 [28]	Cross-sectional qualitative interview and Think Aloud study.	Explore users’ information-seeking difficulties by conceptualizing information seeking as a form of hypothesis testing, and to examine the role of users’ competencies in online information seeking
Lauckner & Hsieh, 2013 [29]	Experimental 2×2 design (position: top-bottom; frequency: high-low)	Does the position and frequency of serious conditions in search results affect perceived severity and susceptibility, and are they related to negative emotional outcomes? Do health literacy and experience with online health seeking moderate these relationships?
Luger, 2014 [30]	Experimental 2×2 design: two different symptom vignettes (mononucleosis or scarlet fever), either Google or WebMD	Explore older adults’ online health seeking to determine the cognitive and diagnostic processes involved
Medlock et al, 2015 [31]	Cross-sectional online survey	To determine which information resources seniors who use the Internet use and trust for health information, which sources are preferred, and which sources are used by seniors for different information needs
Morgan et al, 2014 [32]	Analysis of inquiries posted to a health website	Describe what information people seek from a US website about genetic and rare diseases, and why
Mueller et al, 2016 [33]	Experimental (randomized trial)	To assess the feasibility of testing a symptom appraisal tool for lung cancer symptoms in an online randomized trial
Norr et al, 2014 [34]	Experimental within-subjects design	Investigate whether viewing medical websites adversely affects anxiety sensitivity
North et al, 2011 [35]	Cross-sectional analysis of clicks on a health website and calls to a telephone triage system	Establish what symptoms Internet users tend to look up online and whether telephone triage algorithms could be applied to these
Perez et al, 2015 [36]	Experimental study with Think Aloud	Describe Internet search processes and identify demographic and personal characteristics associated with use of system 1 (does not include hypothesis testing and evidence gathering) and system 2 (includes hypothesis testing and evidence gathering) processing

Author, date	Study design	Aim
Powell et al, 2011 [37]	Cross-sectional survey with embedded qualitative semistructured interviews	Identify the characteristics and motivations of online health information seekers accessing the NHS Direct website
Powley et al, 2016 [38]	Cross-sectional survey and observational study	Evaluate whether patients with inflammatory arthritis and inflammatory arthralgia use the Internet for symptom appraisal and to assess the advice given and diagnoses suggested by the NHS and WebMD symptom checkers
Rice, 2006 [39]	Cross-sectional survey study; secondary analysis of existing dataset	Understand what influences online health seeking, what the reported benefits of online health seeking are, and to identify similarities among online activities
Teriaky et al, 2015 [40]	Cross-sectional survey	Understand how outpatients awaiting initial gastroenterology consultation seek medical information on the Internet and how wait times affect Internet usage
Thomson et al, 2012 [41]	Cross-sectional survey study	Explore characteristics of colorectal cancer patients who used the Web to appraise symptoms prior to diagnosis
White & Horvitz, 2009 [5]	Longitudinal log-based study and cross-sectional survey	(1) Describe escalations that occur when users search for common symptoms and how this escalates to queries about serious conditions, and (2) examine how this persists over several sessions
White & Horvitz, 2009 [42]	Cross-sectional survey study	Explore how lay individuals use the Web to find explanations for symptoms, what activities they pursue, and what their experiences are
White & Horvitz, 2010 [43]	Longitudinal log study using logs from Windows Live toolbar	Predict escalations in searches based on characteristics of websites visited
White & Horvitz, 2010 [44]	Longitudinal log-based study	Establish predictors of when searches for common symptoms lead to health care utilization
White & Horvitz, 2012 [45]	Longitudinal log-based study	Explore how users search for medical concerns and particularly how these concerns impact on future behavior (eg how this influences focus and attention of future searches)
White & Horvitz, 2013 [46]	Longitudinal log-based study	(1) Whether snippets in search results are biased toward serious conditions when symptoms are entered into search engines and 2) how these snippets influence user behavior
Ybarra & Suman, 2006 [47]	National, longitudinal telephone survey	Examine which factors predict whether a Web user is likely to contact a health professional

Table 2. Characteristics of the study populations of studies included in the review (N=32).

Author	Study population	Setting	Sample size
Attfield et al [6]	2 groups of 8 NHS patients: 1 group from a Patient Advice and Liaison Service (PALS) patient panel (43-81 years, mean 64) and one group of MSc students for HCI (25-42 years, mean 31)	UK	16
Briet et al [19]	Users asking hand surgery-related questions from a free online health consultation website	USA (American website; no restriction regarding location of website users)	131 questions
Cartright et al [20]	A set of filtered logs from a toolbar deployed by the Windows Live search engine, containing at least 1 symptom	USA (English-language logs, but no restriction regarding location of users)	2,329,231 actions (=queries issued to a search engine)
Chin [21]	Younger and older adults from a university community	USA	69; 41 younger adults (18-35), 28 older adults (60-83)
Chin & Fu [22]	Younger and older adults from community of a medium-sized city	USA	46, 23 younger (18-28) and 23 older (60-77) adults
Cooper et al [23]	Women aged 40-60 years	USA	132
Cumming et al [24]	Visitors of a UK-based menopause website	UK (UK website; no restriction regarding location of website users)	539
De Choudhury et al [25]	Survey: US adults 18-70 years (census representative sampling); Twitter: 15-month sample of Twitter's Firehose stream, English-language Tweets relating to health; log: data from a major Web search engine	USA (survey with US residents, only English-language log data but not restricted to a certain country)	210 survey respondents; 125,166,549 tweets; 174,605,024 searches
Fiksdal et al [26]	Adult, English-speaking members of the Olmsted County, MN community (where Mayo Clinic is located) and Mayo Clinic patients, employees, and family visitors	USA	19
Fox & Duggan [1]	Adults living in the United States	USA	3014
Hay et al [27]	English-speaking US adults (≥17 years)	USA	120
Keselman et al [28]	Lay individuals (convenience sample)	USA	20
Lauckner & Hsieh [29]	Students from an undergraduate communication course at a large Midwestern university	USA	274
Luger [30]	Older US adults, ≥50 years, community resident, without cognitive impairment, who owned a computer	USA	79
Medlock et al [31]	Members of a local senior (Christian) organization	Netherlands	118
Morgan et al [32]	Random sample of English-language inquiries posted by lay people to the question and answer section of the GARD website and inquiries sent via email	USA (American website but no restrictions on locale of users)	278 inquiries, 68 from 2006 and 210 from 2011
Mueller et al [33]	Adults living in UK with undiagnosed symptoms potentially related to lung cancer	UK	97
Norr et al [34]	Undergraduate students from a large university in the Southern United States.	USA	56
North et al [35]	All symptom assessment callers to Ask Mayo Clinic (telephone triage) and all clicks to specific symptoms on the symptom-checker page of MayoClinic.com	USA	70,370 calls; 2,059,299 clicks
Perez et al [36]	Young adults aged 21-35 with experience of online health information and reported barriers to accessing health services	USA	78
Powell et al [37]	Users of the NHS Direct website	UK	792 for survey, 26 for interviews
Powley et al [38]	Newly presenting patients with either clinically apparent synovitis or a new onset of symptoms consistent with inflammatory arthritis but without clinically apparent synovial swelling attending a secondary care based rheumatology clinic	UK	34

Author	Study population	Setting	Sample size
Rice [39]	US adults: respondents from studies conducted within the Pew Internet and American Life project	USA	13,978 respondents in 2000 who reported health seeking online, 500 of these were telephone interviewed in 2001
Teriaky et al [40]	Patients awaiting appointments at a general gastroenterology clinic in London, ON, Canada	Canada	87
Thomson et al [41]	Newly diagnosed colorectal cancer patients (<6 months)	USA	242
White & Horvitz [5]	Log data related to symptom queries (no mention of restriction by locale) from all major Web search engines (eg, Google, Yahoo!, or Live Survey): randomly selected employees of the Microsoft Corporation who had performed at least 1 health-related online search; survey: Microsoft employees	USA (survey with US residents, no restriction mentioned regarding locale for logs)	Logs: 8732 users with symptom-related queries; survey: 515 participants
White & Horvitz [42]	5000 Microsoft employees were invited via email, from these volunteers were chosen who indicated in a prescreening that they searched the Web for medical information	USA	515 survey respondents
White & Horvitz [43]	Logs from windows live browser toolbar, English-speaking USA relating to 6 basic symptoms	USA (log data issued from US locale)	“Many thousands of logs were mined”
White & Horvitz [44]	Logs from consenting Windows live toolbar users over a 6-month period relating to 3 symptoms: chest pain, muscle twitches, abdominal pain	USA (log data issued from US locale)	700 queries with symptom to HUI transition; 700 queries with symptoms to no HUI transition
White & Horvitz [45]	Logs from consenting Windows live toolbar users over a 3-month period	USA (log data issued from US locale)	169,513 queries
White & Horvitz [46]	Log data related to symptoms queries generated in English-speaking US locale	USA (log data issued from US locale)	2070 symptom queries from 714 users
Ybarra & Suman [47]	Americans living throughout the 50 states and the District of Columbia	USA	Year 1=2104; year 4: 2010, 570 of these were year 1 participants

Table 3. Nature of measures and procedures of studies included in the review (N=32).

Author	Nature of measures and procedure
Attfield et al [6]	Semistructured interviews, eliciting accounts of health information-seeking episodes and how they relate to ongoing health care
Briet et al [19]	Questions and answers to a health website were categorized and analyzed descriptively
Cartright et al [20]	Logs were mined and categorized as either evidence-directed, hypothesis-directed with diagnostic intent, or hypothesis-directed with informational intent, according to defined algorithms
Chin [21]	Participants were randomized to complete either an ill-defined task (find possible causes for a list of symptoms) or well-defined task (find a specific medical term), using a health website; cognitive measures (working memory capacity, processing speed), health literacy measures, medical knowledge measure, search performance for both tasks were measured
Chin & Fu [22]	Participants were given a symptom vignette and asked to find possible causes. Participants were randomized to complete either a parts task (described symptoms based on body parts) or a systems task (described symptoms by functional systems). Tasks were completed either in the parts interface (categorized symptoms based on body parts) or systems interface (categorized symptoms based on functional body systems). Measures included Patients' Medical Background Knowledge, Mental Interface Match Index, Broadness (no. of links), link decision time: time spent reading.
Cooper et al [23]	Discussion in focus groups: which symptoms from a list would be of most concern, why, and what could cause them, what would be their hypothetical response to them, what were actual responses in the past?
Cumming et al [24]	Participants viewed a storytelling video online and then completed a questionnaire evaluating the effect of the video on feeling informed, planned future help seeking, etc
De Choudhury et al [25]	In the survey, participants were asked questions about their experiences using Twitter and search engines to share and seek health information; on the log analysis, tweets and logs were categorized as relating to 4 categories: (1) symptoms of major diseases, (2) benign explanations (nonlife-threatening illnesses), (3) serious illnesses, and (4) disabilities; logs were then analyzed descriptively
Fiksdal et al [26]	Moderators used a semistructured moderator guide to facilitate discussion in focus groups about: (1) participants' perception and understanding of health care information, (2) the process of information collection on the Internet, (3) understanding and usage of information, and (4) implications of health care information for health and well-being
Fox & Duggan [1]	People were contacted via telephone for telephone interviews about online health information seeking
Hay et al [27]	Before their appointment, patients were interviewed about online health information (OHI) seeking, and completed the Wong-Baker-Faces Pain Scale; the consultation was audio-recorded to determine whether OHI was mentioned and then patients completed a satisfaction scale regarding the consultation
Keselman et al [28]	Participants read a hypothetical scenario describing a relative who experienced symptoms typical of stable angina and then discussed possible causes of symptoms from the symptom vignettes in semistructured interviews; then Think Aloud while they researched symptoms on MedlinePlus
Lauckner & Hsieh [29]	The study took place online; participants were presented with a symptom vignette and then with a search engine result page manipulated to show serious conditions either at the top or bottom, and low or high frequency of serious conditions; participants then completed several scales: perceptions of severity and susceptibility using the Risk Behavior Diagnosis scale, history of viewing online health information, their health status, how often they experienced each of the 4 symptoms, and their demographic information, health literacy using the Newest Vital Sign (NVS)
Luger [30]	Participants were presented with 1 of 2 symptom vignettes and asked to diagnose them using Think Aloud, either on Google or WebMD. Measures taken included Think Aloud, self-reported age, gender, ethnicity, education, and income, recent health history, number of hours per week that they used a home computer as well as the number of years that they had owned a home computer, whether or not they had previous experience with the Internet tool to which they were assigned (Google or WebMD's Symptom Checker).
Medlock et al [31]	Participants completed an online questionnaire, which included questions about health information resources used; the Autonomy Preference Index was used to assess information needs and preferences for involvement in health decisions
Morgan et al [32]	A random sample of questions posted to the GARD website were analyzed thematically; collected data included inquiry origin (domestic), type of contact (email and Web-based form), gender, date received at the information center, the specific condition for which they were inquiring, primary language (English), and their reason for inquiry
Mueller et al [33]	Participants first completed a survey about their symptoms and risk factors. They were then randomized to receive the intervention (personalized, theory-based health webpages), or control conditions. Subsequently, participants completed a questionnaire which assessed demographic details, participants' self-reported intention to seek help (scale 1-7), behavioral attitudes and beliefs about help seeking.
Norr et al [34]	Participants first completed the Anxiety Sensitivity Index (ASI), Intolerance of Uncertainty Scale (IUS), and a health anxiety scale (SHAI). Participants were randomized to view either symptom-related websites or general health and wellness control websites. Afterwards, they completed the ASI and SHAI.
North et al [35]	For the MayoClinic website, click data was collected using Google Analytics; for the telephone triage, all completed calls were counted and put into symptom categories based on the algorithm/guideline used during the call.

Author	Nature of measures and procedure
Perez et al [36]	Participants were randomized to one of two symptom scenarios and instructed to search the Internet while using Think Aloud; participants' Internet searches and think-out-loud vocalizations were digitally recorded using screen capture video-recording software
Powell et al [37]	Users of the NHS Direct website completed an online questionnaire survey. A subsample of survey respondents participated in in-depth, semistructured, qualitative interviews by telephone or instant messaging/email.
Powley et al [38]	Patients completed a brief survey on Internet use for symptom appraisal prior to attending clinic; patients were then asked to complete the NHS and WebMD symptom checkers based on their symptoms and their answers and the outcomes were recorded; demographic and disease-related data were obtained from clinic records.
Rice [39]	Respondents were contacted via telephone for telephone interviews asking about online health seeking.
Teriaky et al [40]	Patients awaiting gastroenterology consultation were asked to complete a questionnaire consisting of 16 multiple-choice questions to understand patient use of Web resources for medical information. Abstracted information included patient demographics, level of education, reason for referral, preceding investigations, patient resources utilized, websites browsed, information obtained, reasons for seeking information on the Internet, patient self-diagnosis, and lifestyle changes instituted.
Thomson et al [41]	Semistructured interviews focused on patient sociodemographic and psychological factors, symptom recognition and appraisal, and communication with HCPs, friends, and family.
White & Horvitz [5]	Analysis of logs: Formulated a list of symptoms and associated benign and serious conditions. Recorded all queries to search engines and clicks on result pages, and identified those that included symptoms as search terms. Escalations: Observed increases in medical severity of search terms within a search session. Nonescalations: Search progresses to benign explanation of the symptom; survey: Microsoft employees were sent a survey with open and closed-ended questions regarding participants' medical history and online search behavior
White & Horvitz [42]	Microsoft employees were sent a survey to elicit perceptions of online medical information, experiences in searching for this information, and the influence of the Web on health care concerns and interests. The survey contained "around 70" open and closed questions
White & Horvitz [43]	Cases were identified where queries for symptoms were followed by a query about a related serious condition. Cases where it led to a benign query or no change were termed nonescalations. Using logistic regression, a model was developed to predict escalation using website features of the previously visited page; website features: structural features, title and URL features, firs-person testimonials, page reliability/credibility, commercial intent
White & Horvitz [44]	Log analysis: logs containing symptoms as search terms were filtered, and it was determined whether subsequent searches showed health care utilization intent (HUI). Logistic regression was used to predict HUI based on search characteristics; log entries include a user identifier, a timestamp for each page view, and the URL of the page visited; HUI: queries that indicate searching for contact information for medical facilities
White & Horvitz [45]	Queries were labeled to identify medical and symptoms related queries, and escalations. Subsequently occurring searches were examined. Log entries included a unique user identifier, a timestamp for each page view. Search sessions on Google, Yahoo!, and Bing. Escalation queries were categorized as within-session and between session
White & Horvitz [46]	Log data relating to symptom queries were filtered. Subsequent behavior on the search engine result page was examined, including hovering, cursor movements, clicks, scrolling, as well as bounding boxes of *areas of interest* (AOIs)
Ybarra & Suman [47]	Respondents were contacted via telephone and completed a telephone survey about online health information seeking and help-seeking behavior (seeking help from a health professional or others)

As Table 4 shows, some studies explored Web use regarding current symptoms (n=5) or symptoms that had been experienced previously (n=7), or both (n=1), whereas other studies examined Web use for symptom appraisal by providing participants with a symptom vignette and instructing them to imagine they have these symptoms (n=8). In several studies (n=11), the exact situation of participants was unclear because anonymous data were collected online. Table 4 also highlights the variety and nonspecificity of symptoms examined; most studies (n=15) examined general symptoms and although 10 studies examined specific conditions, only two studies examined similar conditions [27,38]. Finally, most studies (n=20) did not follow up whether participants had received a diagnosis.

Table 4. Symptoms and diagnoses examined in included studies.

Author, date	Were participants symptomatic, asymptomatic, or previously symptomatic[a]?	Type of symptoms examined	Did the study follow up whether Web use was followed by a diagnosis?
Attfield et al [6]	Previously symptomatic	General (any symptoms)	Not assessed
Briet et al [19]	Unclear, participants were users asking questions about symptoms[b]	Hand illness-related symptoms	Not assessed
Cartright et al [20]	Unclear, participants were users issuing symptom-related queries to a search engine[b]	General[c]	Not assessed
Chin [21]	Asymptomatic, participants were presented with a symptom vignette	Symptom vignettes included: pain and stiffness in the body; burning, itching, and sometimes tingling sensation on their body; feeling feverish and chilly after an overseas trip; fatigue, sudden weight gain and difficulty dealing with cold; however, results were not analyzed separately for different symptoms	Not applicable[d]
Chin & Fu [22]	Asymptomatic; participants were presented with a symptom vignette	General (participants received 6 different vignettes with different symptoms, not assessed separately)	Not applicable[d]
Cooper et al [23]	Asymptomatic; participants were presented with a list of symptoms	Symptoms related to gynecologic cancers	Not applicable[d]
Cumming et al [24]	Most symptomatic (448/492), but some asymptomatic (44/492)	Menopausal symptoms	Not assessed
De Choudhury et al [25]	Unclear, participants were users issuing symptom-related Tweets and queries to a search engine[b]	General, logs were filtered for references to symptoms using a comprehensive list of symptoms from the Merck medical dictionary	Not assessed
Fiksdal et al [26]	Previously symptomatic	General (any symptoms)	Not assessed
Fox & Duggan [1]	Previously symptomatic	General (any symptoms)	Participants were asked whether their diagnosis was confirmed by a health professional; 45% said it was confirmed, 35% did not present, 19% said it was not confirmed/inconclusive
Hay et al [27]	Symptomatic; participants were newly diagnosed rheumatology patient	Rheumatoid symptoms	Yes, patients' diagnoses were gathered after the appointment or at follow-up appointment
Keselman et al [28]	Asymptomatic; participants received a symptom vignette	Symptoms typical of stable angina	Not applicable[d]
Lauckner & Hsieh [29]	Asymptomatic; participants received a symptom vignette	Symptom vignettes involved one of four symptoms: headaches, chest pain, muscle twitches, or abdominal pain, but the different symptoms were not analyzed separately	Not applicable[d]
Luger [30]	Asymptomatic; participants received a symptom vignette	Symptom vignettes involved either mononucleosis or scarlet fever	Not applicable[d]
Medlock et al [31]	Previously symptomatic	General (any symptoms)	Not assessed
Morgan et al [32]	Unclear, participants were users issuing symptom-related Tweets and queries to a search engine[b]	Symptoms related to any type of genetic or rare disease	Not assessed
Mueller et al [33]	87 participants were symptomatic, 10 were asymptomatic but searching on behalf of someone else	Symptoms related to lung cancer	Not assessed
Norr et al [34]	Asymptomatic; participants viewed a list of symptoms	General ("websites focused on symptoms of medical conditions")	Not applicable[d]

Author, date	Were participants symptomatic, asymptomatic, or previously symptomatic[a]?	Type of symptoms examined	Did the study follow up whether Web use was followed by a diagnosis?
North et al [35]	Unclear, participants were users searching the MayoClinic website or using a telephone triage[b]	General (any symptoms)	Not assessed
Perez et al [36]	Asymptomatic; participants received a symptom vignette	One of two clinical symptom scenarios: (1) fever, mild headache, dry cough, and myalgia, suggestive of influenza, and (2) fever, severe headache, and stiff neck, suggestive of meningitis	Not applicable[d]
Powell et al [37]	Unclear, participants were users of the NHS website[b]	General (any symptoms)	Not assessed
Powley et al [38]	Symptomatic; participants were patients attending a secondary care based rheumatology clinic	Either clinically apparent synovitis or a new onset of symptoms consistent with inflammatory arthritis but without clinically apparent synovial swelling	Yes, rheumatological diagnosis was recorded after consultation
Rice [39]	Previously symptomatic	General (any symptoms)	Not assessed
Teriaky et al [40]	Symptomatic; participants were patients awaiting gastroenterology appointments	Symptoms related to gastroenterology	Not assessed
Thomson et al [41]	Symptomatic; participants were colorectal cancer patients	Symptoms related to colorectal cancer	Yes; all participants were diagnosed with colorectal cancer
White & Horvitz [5]	Logs: Unclear, participants were users issuing symptom-related queries to a search engine[b]; survey: previously symptomatic	Logs related to 3 common symptoms (headache, muscle twitches, and chest pain)	Not assessed
White & Horvitz [42]	Previously symptomatic	General (any symptoms)	Not assessed
White & Horvitz [43]	Unclear, participants were users issuing symptom-related queries to a search engine[b]	Queries related to any of 6 common symptoms: headache, chest pain, muscle twitches, abdominal pain, nausea, and dizziness	Not assessed
White & Horvitz [44]	Unclear, participants were users issuing symptom-related queries to a search engine[b]	Queries related to one of 3 symptoms: chest pain, muscle twitches, and abdominal pain	Not assessed
White & Horvitz [45]	Unclear, participants were users issuing symptom-related queries to a search engine	General[c]	Not assessed
White & Horvitz [46]	Unclear, participants were users issuing symptom-related queries to a search engine[b]	General[c]	Not assessed
Ybarra & Suman [47]	Previously symptomatic	General (any symptoms)	Not assessed

[a] Symptomatic: participants experienced the symptoms at the time of the study; asymptomatic: participants did not have symptoms and were surveyed regarding fictional symptoms; previously symptomatic: participants were surveyed about symptoms they experienced previously.

[b] Participants were users asking questions about symptoms (could be own symptoms or asking on behalf of someone else).

[c] Any queries related to a comprehensive list of symptoms from the Merck medical dictionary.

[d] Patients were not symptomatic.

Quality Assessment and Risk of Bias

Quality assessment of the studies is shown in Multimedia Appendix 3. Subsequently, we use this information to critically appraise evidence from the studies and assess risk of bias.

What Proportion of Different Populations Use the Web to Appraise Symptoms?

Four studies, all surveys, reported the proportion of the study sample that engaged in Web use for symptom appraisal (Table 5).

Table 5. Percentage of people engaging in Web use for symptom appraisal reported by included studies (n=4).

Reference	Study population	Sample size	Reported Web use for symptom appraisal, % (95% CI)
Fox & Duggan [1]	Adults living in the US	3014	35% (33%-37%)
White & Horvitz [42]	US Microsoft employees	515	75% (71%-79%)
Medlock et al [31]	Members of a senior church organization, Netherlands	118	23% (15%-31%)
Thomson et al [41]	Colorectal cancer patients, US	242	25% (20%-31%)

In Fox and Duggan's [1] population-based survey with adults living in the United States, 35% reported going online to attempt self-diagnosis. Participants were sampled to mirror the population in terms of demographics, but disproportionately stratified to increase the incidence of nonwhite respondents. This survey was conducted in the United States and is therefore likely to reflect proportions in Western, high-income countries with high Internet penetration.

White and Horvitz's [42] survey conducted among Microsoft employees found that "three-quarters of subjects" (the authors do not provide absolute numbers; assuming the proportion is 75% of N=515, 95% CI 71%-79%) reported searching for information on symptoms. "Two-thirds" reported researching professionally undiagnosed conditions, at least once a month [42]. It should be noted that this sample was biased toward younger, male respondents with high educational level and socioeconomic status working within an industry that is very Web-oriented.

Medlock et al [31] examined online health information seeking in older people by surveying members of a senior Christian organization. They found that 23% of participants reported using the Web in the past 12 months to determine the cause of symptoms. This shows that, although Web use for symptom appraisal may be less common among older people than in the general population (compared to 33%-37% found by Fox and Duggan [1]), older people do engage in it.

While the previous surveys focused on diagnostic searches for any conditions/symptoms, Thomson et al [41] conducted a survey with colorectal cancer patients and found that 25% of the sample reported prediagnosis Web use for symptom appraisal.

To conclude, Fox and Duggan's [1] study with its large, population-based sample size is most likely to give an accurate proportion for the general population (in a Western, higher income country), although the other included studies give an indication of how this proportion can vary depending on the population being surveyed (ie, depending on sociodemographic variables and disease-related factors). It should also be noted that the confidence intervals are wide in a number of these studies reflecting considerable uncertainty about the true proportion.

Which Symptoms Are Likely to Be Researched Online?

Six studies examined characteristics of symptoms that were searched for online [25-27,35,39,41]. Three of these were survey studies [27,39,41], one was an interview study [26], and two involved analyses of data on usage of online resources such as social media, search engines, and health websites [25,35].

North et al [35] compared users of the Mayo Clinic website with people who used a telephone triage system to appraise their symptoms and found that telephone triage users were more likely to have acute and conspicuous symptoms requiring immediate relief, whereas website users were more likely to research chronic conditions. Hay et al [27] surveyed rheumatology patients and found that some individuals in their study sought help online because they had a history of undiagnosed symptoms. Findings from both of these studies suggest symptoms are researched online when they have been present for a prolonged time.

In their study on colorectal cancer patients, Thomson et al [41] found that neither symptom severity nor stage at diagnosis was related to Web use for symptom appraisal, but Web users were more likely to experience symptoms typically perceived as embarrassing, such as change in bowel habits. Similarly, Choudhury et al [25], who analyzed log data obtained from Twitter and a search engine, found that potentially embarrassing, stigmatized, or sensitive symptoms such as "vaginal bleeding" or "pelvic pain" were more likely to be searched for than tweeted. Furthermore, Rice's [39] population-based telephone survey conducted in the United States concluded that more frequent online health seekers were more likely to look for sensitive health topics that are difficult to talk about than less frequent online health seekers.

Finally, in Fiksdal et al's [26] focus group study with 19 US adults, participants reported turning to the Web when symptoms were perceived as trivial/nonserious and they wanted to avoid "bothering" health professionals.

In conclusion, it appears Web use for symptom appraisal occurs when symptoms are persistent, have a history of being undiagnosed by health professionals, are potentially embarrassing or stigmatized, and/or when they are perceived as superficial/nonserious.

How Is Web Use for Symptom Appraisal Conducted (Search Strategies)?

Theme 1: Symptom-Based, Condition-Based, and Treatment-Based Searches

Three distinct approaches to searching were identified: (1) symptom-based searches, which used symptoms as search terms; (2) condition-based searches, which involved searches for particular conditions, and (3) treatment-based searches, which involved researching treatments for symptoms without prior research on possible causes.

Log data from search engines suggest the majority (65%) of exploratory health-related searches (ie, those aimed at diagnosing a condition) are symptom-based rather than condition-based [20], and remain symptom-based throughout the search because search sessions tend to start and end with purely symptom-related queries [5,20]. One should bear in mind, however, that log-based studies cannot ascertain searchers' actual intentions and motivations. The authors assume occurrences of certain search terms signal certain intentions (eg, a symptom and the term "cause" signals diagnostic intent); however, articles did not report any prior validation of these algorithms.

An experimental study that observed people (N=79) as they used Google or a symptom-checker tool to diagnose symptom vignettes reported that most users conduct symptom-based searches because most people began their search by entering symptoms and only 24% began by specifying a condition [30].

In an experimental study conducted by Perez et al [36], participants (N=78) were instructed to research the Web as if they were experiencing a given symptom, described in a vignette, while using Think Aloud. Think Aloud (also known as "cognitive interviewing") requires participants to vocalize their thoughts while performing a task [48]. The authors found that 19% of searches were treatment-based and the remainder symptom- or condition-based (the authors did not report these separately). It should be noted that the external validity of vignette-based studies is limited because individuals base their searches on the vignette descriptions rather than actual perceptions or observations, and the search behavior observed is likely to depend on the phrasing of the vignette.

Keselman et al [28] used interview and Think Aloud methods to explore how a convenience sample of 20 lay individuals interpreted a symptom vignette using the American consumer health information service MedlinePlus. They concluded that some participants conducted condition-based searches and some participants used a symptom-based approach. Additionally, they identified a group of participants who used a condition-based approach, but began their search with a broader hypothesis, such as "heart disease," and then attempted to narrow down their search. Different barriers seemed to play a role in the different search strategies: condition-based searchers were prone to confirmation bias, seeking out information that confirmed their hypothesis and terminating the search before reviewing further hypotheses. Those starting with a broad hypothesis often terminated the search without coming to a conclusion. The symptom-based searchers struggled to find the relevant results due to the lack of specificity of their search terms [28].

Overall, it seems most Web use for symptom appraisal searches are symptom-based and both log-based studies, which have high external validity, and experimental studies, which have high internal validity, confirm this finding. No validation was reported for the algorithms used for the log-based studies, however, and experimental and qualitative studies used to observe search behavior have limited generalizability to real-world contexts.

Theme 2: Selection of Search Terms

Keselman et al [28] examined the search behavior of 20 adults using Think Aloud and discovered that participants often ignored symptoms mentioned in the vignette if they perceived them as irrelevant and exempted these from their search terms. The authors termed this "selective perception bias." The participants also tended to ignore aspects of duration of the symptoms and had difficulty discerning acute from chronic symptoms. However, these findings stem from a single study with only 20 adults using fictional scenarios, thus further validation is required.

Theme 3: Age Differences

Three studies reported on age differences in search behavior [21,22,30]. Chin et al [21] (N=69) compared the search performance of younger and older adults while performing either a well-defined task (searching for a specific medical term on a website) or an ill-defined task (using the website to diagnose a set of symptoms). The study found that older adults performed better in the ill-defined task, whereas younger participants performed better in the well-defined task [21].

In another study (N=46), Chin and Fu [22] presented older and younger adults with different interfaces of the same website: one interface categorized symptoms based on the body parts they occurred in and the other interface categorized symptoms according to functional systems (eg, respiratory system). Younger adults tended to click on significantly more links within one category, suggesting they followed the interface of the website, whereas older adults clicked significantly more between-category links regardless of the interface. Chin and Fu [22] conclude this was due to older adults using their existing medical knowledge rather than the interface to guide their search, which is supported by their finding that older adults performed better in a medical knowledge task [22].

Luger at al [30] explored the search behavior of adults aged 50 years and older (N=79) in a Think Aloud study and found that participants who accurately diagnosed the condition presented in a symptom vignette were slightly younger (mean 61.72, SD 6.17 years) than those who were inaccurate (mean 65.51, SD 7.54 years), although no inferential statistics were reported.

Thus, there are some indications that older adults perform differently in Web searches for symptom appraisal than younger adults, possibly due to medical knowledge. However, the available studies used small sample sizes, thus inferences to the wider population may not be appropriate.

Theme 4: Selecting Information

Several studies examined how users select information from their search results. We identified four subthemes relating to selection of information.

Number of Search Results Viewed

Lauckner and Hsieh [29] reported that participants in their laboratory-based, experimental study with undergraduate students (N=274) viewed approximately four links on results pages, which was the number of results visible above the "fold" (ie, users would need to scroll down to view more results). Corroborating this finding, Keselman et al [28] discovered that participants in their qualitative study (N=20) often ignored relevant links while trying to diagnose a symptom vignette if these were located below the fold. Luger et al [30] found that older adults (N=79) in their experimental, laboratory-based study tended to view approximately six conditions on the WebMD symptom-checker tool after entering a set of symptoms, although the authors do not clarify whether this was the number visible above the fold.

Thus, the top results returned by search engines will have maximum impact on symptom appraisal, whereas those located below the fold may have little to no effect. Because these findings all relate to laboratory-based studies, however, further investigation in naturalistic settings would be beneficial.

Process of Elimination

In their study using Think Aloud with 79 adults aged 50 years and older, Luger et al [30] found that 91% of participants used a "process of elimination," whereby the symptoms described in the vignette were compared against those listed for a given condition and the condition was discarded as a hypothesis if it included symptoms not mentioned in the vignette. This finding suggests that a common search strategy is to narrow down the hypotheses by discarding those with symptoms not matching one's own. However, the majority of the sample was highly educated (all had some university education), therefore generalizability to the wider population is unclear.

Source Credibility

In Luger et al's [30] study with adults aged 50 years and older (N=79), source credibility was mentioned by only 25% of the sample. They also found that one-third viewed user-generated content such as discussion boards, which are not quality controlled. White and Horvitz [46] found in their log-based study that consumer sites such as MayoClinic.com or WebMD (both well-known American corporations and health websites) are positively related to click-through rates following searches that contain symptoms as search terms, suggesting widely known, established health websites are likely to be accessed during Web use for symptom appraisal. White and Horvitz [5] also suggest searches that "escalate" (ie, progress from searching for symptoms to serious conditions) contain more visits to "trusted sources" (eg, governmental websites, websites of health organizations). It is important to note that although the authors claim to research "diagnostic searches," this was identified through the presence of symptoms in search terms and may therefore also include nondiagnostic searches.

Pages Mentioning Serious Illnesses

White and Horvitz [46] filtered logs from a search engine and examined how users issuing symptom queries subsequently interacted with search results pages. They found that users engage more with captions on search results pages that mention serious illnesses, hovering more frequently and longer over these captions and clicking these more often than captions mentioning benign causes. Terminology related to serious illnesses such as "malignant," "severe," and "tumor" significantly increased click probability, whereas terms such as "benign" decreased click probability. Additionally, users were more likely to engage with sites indicating they can help identify causes of symptoms (eg, by mentioning the words "learn" and "causes").

Although we do not know searchers' intentions or how they used the information found, these findings suggest those researching symptoms online are more likely to engage with websites relating to serious causes.

To summarize, Web use for symptom appraisal typically involves inputting information into a search tool and subsequently narrowing down results returned by the search tool. When inputting information, most users appear to choose search terms based on symptoms rather than hypothesized conditions, but users do not appear to utilize all information available (eg, some symptoms may be omitted, as well as the frequency/duration of symptoms). Furthermore, there is some limited evidence that older adults perform differently in Web searches for symptom appraisal than younger adults, and that this may be due to medical knowledge. Once a selection of results is provided by the search tool, users tend to narrow results down by taking into account the results' position on the results page, the degree of seriousness of the condition, the credibility of the source, and the extent of overlap between the listed and the experienced symptoms.

Behavioral Consequences of Web Use for Symptom Appraisal

Theme 1: Increased Help Seeking

In Fox and Duggan's [1] population-based survey (N=3014), 46% (95% CI 44.22%-47.78%) of online self-diagnosers claimed that information found online led them to think they needed the attention of a health professional. Thomson et al [41] found in their survey with colorectal cancer patients that 25% (95% CI 19.54%-30.46%) of online self-diagnosers were reportedly persuaded by the information found online to see a health professional. This suggests the proportion of people encouraged to seek medical help based on Web use for symptom appraisal may be significantly lower among colorectal cancer patients than in the general population surveyed by Fox and Duggan [1].

Using logistic regression with a survey sample of more than 2000 Americans aged 12 years and older, Ybarra et al [47] found that online self-diagnosers were 2.5 times more likely to report contacting a health professional than online health information seekers who did not try to diagnose symptoms online, suggesting Web use for symptom appraisal is linked to increased health care contact.

Some studies suggest that the mode of presenting information on a website may affect users' decisions to seek medical advice: in a UK-based qualitative study [24], participants reportedly felt encouraged to seek help after viewing an online "digital storytelling" video about urogenital atrophy; 73% who had reportedly been too embarrassed to see a health professional before and 87% who had not wanted to bother their doctor would now seek help. In an online pilot randomized trial, Mueller et al [33] examined whether addition of theory-based components to online health information can increase intention to seek help. The theory-based components appeared to significantly increase intention to seek medical help, although the sample size in this pilot study (N=97) was too small to allow firm conclusions.

Using log-based search engine data, White and Horvitz [44] examined search behaviors related to health care utilization intent. Health care utilization intent was assumed to be present when users conducted searches for health care practitioners/clinics near their geographical area. They found that users who displayed certain search behavior—such as visiting websites that mention serious conditions before benign ones—were more likely to show health care utilization intent subsequently. This suggests online search behavior following symptom queries is related to subsequent health care contact, although mechanisms of causality are unclear based on this data. Furthermore, the authors do not report whether/how their algorithms were validated; thus, it is unclear whether their proxy measure of health care utilization intent is valid.

By observing how patients attending a rheumatology clinic completed the NHS and WebMD symptom-checker tools, Powley et al [38] found indications that symptom-checker tools provide information that can propagate unnecessary help seeking. Of 34 patients, 15 were inappropriately advised to seek help from emergency services rather than primary care, indicating potential issues with the algorithms used in symptom-checker tools. One should bear in mind, however, that this study does not allow any conclusion on whether real users would follow this advice or not.

Theme 2: Decreased Help Seeking

In Powell et al's [37] interview study of users (N=26) of the NHS Direct website (the official website of the UK National Health Service), some participants reportedly used online health information as a form of "demand management," to identify trivial symptoms not warranting medical attention. Similarly, some participants in Fiksdal et al's [26] focus group study reported using the Web to avoid "bothering" health professionals with trivial symptoms. In both studies, it was not followed up whether users had correctly or incorrectly classified symptoms as trivial. Finally, in Attfield et al's [6] interview study, participants reported sometimes being reassured by Web searches that help seeking was not necessary. Evidence for this theme stems only from qualitative studies, thus, generalizability is uncertain.

Theme 3: Communication with Health Professionals

Fox and Duggan [1] found that 53% of online self-diagnosers reportedly discussed the health information found online with a health professional. Two qualitative studies by Cooper et al [23] and Attfield et al [6] found that patients used the Web to appraise symptoms in order to prepare for consultations by preparing questions, collating relevant information, and enhancing their knowledge in order to understand the advice received.

Two studies found indications that Web use for symptom appraisal is related to reduced communication with a health professional [27,41]. Hay et al [27] found that new rheumatology patients who engaged in Web use for symptom appraisal were significantly less likely to want to challenge their health professionals' advice than those who did not. The authors note that study participants were concerned about evoking the impression of questioning health professionals' advice. Thomson et al [41] found that Web use for symptom appraisal was significantly related to feeling hesitant about discussing symptoms with a health professional. However, direction of causality is unclear. It is possible that information found online dissuaded individuals from communicating with health professionals or that people chose to research symptoms online because they were reticent about discussing their symptoms.

From the preceding findings, we can conclude that Web use for symptom appraisal is used to inform the decision of whether to present to health services and that online self-diagnosers are more likely than other health information seekers to contact a health professional. This can potentially be increased, where appropriate, with novel methods such as "digital storytelling" or theory-based components. Some evidence also suggests that online health information can potentially reduce help seeking by calming users' fears. It is unclear, however, what proportion of users feel encouraged or discouraged to seek help appropriately (ie, what proportion of users who feel encouraged to seek help actually have a condition warranting medical attention, and what proportion of users who feel discouraged to seek help actually do not need medical attention). Furthermore, it is unclear whether those engaging in Web use for symptom appraisal are more or less likely to seek medical advice than those experiencing the same symptoms without researching online because this comparison was not made in any of the included studies. Web use for symptom appraisal can also play a role in communication with health professionals by influencing how individuals prepare for consultations and prompting discussion of online health information.

Emotional Consequences of Web Use for Symptom Appraisal

In White and Horvitz's [42] survey among 515 Microsoft employees, 38.5% reported that online health information had made them feel anxious in the past, and 50.3% reported Web use for symptom appraisal had made them feel less anxious. The survey sample was biased toward younger, more educated, and information technology-literate respondents.

Powell et al [37], who examined the motivations of users of the NHS Direct website using semistructured interviews, found that participants sought health information online to obtain reassurance about symptoms. The majority nevertheless subsequently sought medical help, although sometimes with less urgency and anxiety.

Teriaky et al [40] surveyed patients (N=87) awaiting appointments at a general gastroenterology clinic and asked those who reported using the Web prior to their consultations whether this had changed their anxiety levels. In all, 77% experienced no change, 21% experienced an increase, and 2% a decrease. One should note that this sample consists of those who decided to report to health services (and who admitted Web use for symptom appraisal); there may be a larger proportion of users who felt calmed by their searches and therefore did not present to health services.

Lauckner et al [29] found in their experimental laboratory-based study (N=274) that presenting search engine results relating to serious conditions before benign conditions and a higher frequency of results relating to serious conditions was related to negative emotional outcomes such as fear. These findings suggest a causal relationship between exposure to search results during Web use for symptom appraisal and increases in anxiety.

Another experimental study conducted by Norr et al [34], however, found no difference in the anxiety levels of their participants (N=56) following review of either (1) websites containing information on causes of symptoms or (2) websites on general health and wellness (eg, exercise, healthy diet) without reference to medical conditions or symptoms.

Therefore, some evidence suggests there is a relationship between Web use for symptom appraisal and health anxiety. Findings from experimental studies were mixed regarding causal relationships.Surveys and interviews indicate there is a potential for calming effects and decreases in anxiety, and that the proportion who report feeling calmed by Web use for symptom appraisal is higher than those reporting anxiety. It is also possible that those who engage in Web use for symptom appraisal are more anxious about their health generally. It is unclear when anxiety is warranted because participants' actual diagnoses were not followed up, and comparisons to those who did not research symptoms was lacking.

Discussion

This is the first systematic review and synthesis of the literature available on Web use for symptom appraisal. Our main findings were:

1. Approximately 35% of the general population engage in Web use for symptom appraisal, but the proportion can vary considerably (25%-75%) depending on the population under study.
2. Symptoms tend to be researched online when they are long term, potentially embarrassing/stigmatized, have been presented to health services previously with inconclusive outcomes, and/or when they are perceived as trivial.
3. Searches tend to be based on symptoms rather than hypothesized conditions; users seem to focus on particular symptoms while disregarding other symptoms and aspects such as frequency and duration.
4. Once a selection of results is returned by the search tool, people use specific techniques to narrow results down (eg, taking into account the position on the results page or the credibility of the source).
5. Evidence indicates that online information is used to inform the decision of whether to contact health services and is related to (increased and decreased) anxiety, but the precise impact cannot be discerned due to lack of follow-up and appropriate comparison groups.

Subsequently, we discuss whether Web use for symptom appraisal should be viewed as an asset or a liability in health care delivery based on currently available evidence, and make recommendations for the improvement of online health information.

Web Use for Symptom Appraisal: Assistance or Hindrance to Health Promotion?

Criticisms of online self-diagnoses include concern over unnecessary anxiety and health care utilization [5,49]. Our review confirms that the Web can increase anxiety and health care contact among users [5,29,42,44], but reveals that there is insufficient evidence to conclude this occurs unnecessarily.

First, it is important to note limitations of approaches used to examine relationships between Web use for symptom appraisal and health anxiety or help-seeking behavior. Cross-sectional surveys cannot show direction of causality. It is possible that using the Web to appraise symptoms causes anxiety, or that anxiety triggers Web use for symptom appraisal, or that a third factor influences both. Furthermore, the surveys that reported on anxiety among online self-diagnosers were biased toward certain demographic [42] or patient groups [40], and did not use validated measures of anxiety levels.

Log-based studies, which evaluate behavior based on search engine log data, do not allow firm conclusions regarding users' actual behaviors and motivations. For example, White and Horvitz [5] found that users who begin their searches for symptoms often progress to researching serious conditions, but it is not clear whether users are anxious or using the information to reassure themselves. The authors assume certain search terms signal certain intentions (eg, a symptom and the term "cause" signals diagnostic intent), but no action was described to determine the validity of these assumptions. Thus, insights from log-based studies are limited.

Experimental research shows that users asked to research certain symptoms may report feeling anxious following Web searches [29], but it is not possible to infer whether this anxiety would be unwarranted in a naturalistic setting (ie, if symptoms were actually present).

Using the Web to appraise symptoms may also decrease anxiety in some cases [37,42] and Web searches are sometimes used to identify alternatives to health care utilization [6,37]. Individuals describe using online information to evaluate mild/superficial symptoms to avoid wasting health professionals' time [26]. This shows that Web use for symptom appraisal can also decrease anxiety and help seeking. It is possible that Web use for symptom appraisal discourages help seeking for trivial symptoms, thus reducing pressure on health care resources. However, it is also possible that Web use for symptom appraisal leads to complacency and prevents help seeking when it is actually necessary.

There are also indications that Web-based information can help individuals recognize their symptoms as signs of serious conditions [41]. Indeed, our review highlights that online health information is an important resource when obtaining information from health professionals is difficult (eg, when symptoms are embarrassing or stigmatized) [25,39,41] or when previous visits to health care have been ineffective [27,35]. This suggests there is potential for the Web to be an assistance to health care.

Finally, it should be noted that worry can also have positive effects on health behaviors [50]. The Protection Motivation Theory suggests that fear will increase intention to perform a certain behavior if the individual feels able to perform the behavior and believes that the behavior will reduce the threat [51]. Similarly, according to the Health Belief Model, fear should result in recommended health behavior if perceived benefits of the recommended behavior are high and barriers are low [51]. Overall, this indicates anxiety induced through online health information can enhance recommended health behaviors if information is presented in a way that enables concrete action and decision making.

A limitation we discovered across different methodologies was the lack of follow-up on participants' help-seeking behavior and diagnoses. Without this information, we cannot discern whether individuals' self-diagnoses and decisions regarding help-seeking behavior are appropriate or not. We also cannot determine long-term impacts on health care utilization. Furthermore, essential comparison groups are generally lacking. For example, it would be necessary to compare those who research symptoms online with those who do not (rather than surveying only online self-diagnosers), and to compare those who present to health services with those who do not (rather than surveying only patients presenting in clinic) to determine impacts of Web use.

Recommendations to Improve Online Health Information

Based on the findings of this review, we suggest changes to health websites, Web apps, and search engines such that they can provide useful information to those researching symptoms.

Our analyses reveal that users tend to search inductively based on symptoms. Search engines and symptom-checker tools need to ensure users are directed to useful information when symptoms are entered. The review also shows that searchers tend to omit dimensions such as duration and frequency of symptoms in their search terms [28], and that symptom-based searchers struggle to find relevant results due to lack of specificity of their search terms. This suggests it is important that users are directed to useful terms to narrow their search and prompted to provide information on duration and frequency of symptoms to improve specificity of searches. In support of this, recent research suggests that incorporation of query expansion techniques into information retrieval systems can improve the search effectiveness of search engines for diagnostic symptom searches [52].

Our review also reveals that online health information can impact on the decision to seek help and on communication with health professionals. Health websites and apps need to ensure they provide useful information to support searchers in their decisions and health care interactions. Health websites providing symptom information should, for example, provide clear guidelines on when medical advice should be sought (eg, if a symptom has a certain quality or duration) and how help should be sought (eg, immediately via emergency services or within the next week via primary care).

Strengths and Limitations

As the review includes a diversity of study types and methods, a quantitative synthesis or meta-analysis was not possible. However, traditional forms of systematic review that do not make use of all forms of evidence often do not take differing contexts into account, limiting their use to policy makers and practitioners [15]. More inclusive forms of review that combine findings from different study designs allow a richer, more holistic understanding of the phenomenon under study [15]. We were able to combine real-world insights from observational studies, such as analyses of search engine log data with data from more controlled, experimental settings, thereby improving external and internal validity. Furthermore, by incorporating findings from large, population-based studies as well as smaller interview-based studies, we were able to gain an understanding of the impact of Web use for symptom appraisal at the population level, while also obtaining more detailed reports of peoples' perceptions and experiences. Moreover, by including studies that cover a broad range of populations (eg, different conditions/symptom profiles, age groups, socioeconomic status), we have shown how Web usage can differ depending on context.

In this review, we considered a diversity of symptoms and conditions; when more research in this area becomes available, it would be useful to carry out more focused reviews because the nature of the symptom is likely to influence Web use online [25-27,35,39,41].

Finally, it should also be noted that this review did not examine Web use for mental health symptoms. Web use for symptoms related to mental health and its impact on help seeking represent an important field of study and should be assessed in a separate review of the literature.

Conclusions and Future Work

This systematic review indicates that the Web can disseminate information to those worried about symptoms and can affect their decisions to present to health services. It also suggests Web use for symptom appraisal can impact on how patients prepare for consultations with health care professionals. Thus, we can conclude that Web use for symptom appraisal has the potential to influence the timing of help seeking and the communication between patients and health care professionals during consultations.

At present, limitations of the reviewed studies mean it is not clear when the Web plays a beneficial role in health care delivery and when it is detrimental. Web use for symptom appraisal has been linked to increased as well as decreased anxiety and health care contact. However, the evidence does not show when this is warranted because most studies did not follow up whether participants ultimately sought help following their Web searches and whether they received a diagnosis. Furthermore, comparison groups are lacking to determine the effects of Web use for symptom appraisal.

We need longitudinal research that follows up whether participants seek help and are ultimately diagnosed following Web searches, and compare Web searchers to non-Web searchers. These data can then be used to weigh the benefits of Web use for symptom appraisal (eg, reductions in delays to diagnosis and avoidance of unnecessary health care use) against the disadvantages (eg, unnecessary anxiety and health care use) and relate these to health care costs. Research should focus on real-world samples of people experiencing symptoms and could involve novel methods of tracking behavior, such as analysis of search engine log data and mobile geotracking as used in some of the included studies to follow people over time. These studies have the advantage of high external validity and large sample sizes. However, the algorithms used to analyze these data should first be tested extensively for reliability and validity before further work to evaluate cost effectiveness can meaningfully be conducted. Moreover, further experimental studies would allow a detailed analysis of search behavior. Future research could examine how the different search strategies identified here—symptom-based, condition-based, and treatment-based—relate to cognitive biases and link this to theory.

Acknowledgments

This project was funded by a Medical Research Council (MRC) Doctoral Training Partnership studentship (Ref: 1354978) and by a President’s Doctoral Scholar Award from the University of Manchester.

Conflicts of Interest

None declared.

Authors' Contributions

JM drafted the review protocol, and all coauthors critically revised the protocol. JM, AD, and JV independently completed the study selection procedure and critically discussed all studies for inclusion in the review. JM extracted and analyzed the data and wrote the first draft of the manuscript. CT, SH, and CJ contributed to discussion of the content and analysis method. All coauthors reviewed and edited the manuscript before submission.

Multimedia Appendix 1

Example search strategy.

[PDF File (Adobe PDF File), 18KB - jmir_v19i6e202_app1.pdf]

Multimedia Appendix 2

Data extraction sheet.

[PDF File (Adobe PDF File), 24KB - jmir_v19i6e202_app2.pdf]

Multimedia Appendix 3

Quality appraisal of included studies.

[PDF File (Adobe PDF File), 103KB - jmir_v19i6e202_app3.pdf]

Multimedia Appendix 4

NVivo Matrix Screenshot.

[PNG File, 191KB - jmir_v19i6e202_app4.png]

References

1. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health online 2013 URL: http://pewinternet.org/Reports/2013/Health-online.aspx [accessed 2016-07-08] [WebCite Cache ID 6iqsuJyWn]
2. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
3. Dickerson SS, Reinhart A, Boemhke M, Akhu-Zaheya L. Cancer as a problem to be solved: internet use and provider communication by men with cancer. Comput Inform Nurs 2011 Jul;29(7):388-395. [doi: 10.1097/NCN.0b013e3181f9ddb1] [Medline: 20975535]
4. Kirschning S, von Kardorff E. The use of the Internet by women with breast cancer and men with prostate cancer-results of online research. J Public Health 2007 Jul 10;16(2):133-143. [doi: 10.1007/s10389-007-0134-0]
5. White RW, Horvitz E. Cyberchondria. ACM Trans Inf Syst 2009 Nov 01;27(4):1-37. [doi: 10.1145/1629096.1629101]
6. Attfield SJ, Adams A, Blandford A. Patient information needs: pre- and post-consultation. Health Informatics J 2006 Jun;12(2):165-177. [doi: 10.1177/1460458206063811] [Medline: 17023406]

7. Radzikowska E, Roszkowski-Śliż K, Głaz P. The impact of timeliness of care on survival in non-small cell lung cancer patients. Pneumonol Alergol Pol 2012;80(5):422-429 [FREE Full text] [Medline: 22926903]
8. Duprez D. Early detection of cardiovascular disease-the future of cardiology? E-J ESC Counc Cardiol Pract 2006;4(19):1 [FREE Full text]
9. Topouzis F, Anastasopoulos E. Glaucoma-the importance of early detection and early treatment. Eur Ophthalmic Rev 2007;13(4):16. [doi: 10.17925/EOR.2007.00.00.13]
10. Centre for Reviews and Dissemination. CRD's Guidance for Undertaking Reviews in Health Care. York: Centre for Reviews and Dissemination, University of York; 2009.
11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [Medline: 19622552]
12. eBiz. 2016. Top 15 most popular search engines URL: http://www.ebizmba.com/articles/search-engines [accessed 2016-07-13] [WebCite Cache ID 6iyL0tgeK]
13. Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol 2006;6:35 [FREE Full text] [doi: 10.1186/1471-2288-6-35] [Medline: 16872487]
14. Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. J Health Serv Res Policy 2005 Jul;10 Suppl 1:6-20. [doi: 10.1258/1355819054308576] [Medline: 16053580]
15. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [Medline: 15667704]
16. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
17. Spencer L, Ritchie J, O'Connor W, Morrell G, Ormston R. Analysis in practice. In: Qualitative Research Practice: A Guide For Social Science Students and Researchers. London: Sage Publications; 2013:295-346.
18. Brown LD, Cai TT, Dasgupta A. Interval estimation for a binomial proportion. Stat Sci 2001;16(2):101-133 [FREE Full text]
19. Briet JP, Hageman MG, Blok R, Ring D. When do patients with hand illness seek online health consultations and what do they ask? Clin Orthop Relat Res 2014 Apr;472(4):1246-1250 [FREE Full text] [doi: 10.1007/s11999-014-3461-9] [Medline: 24442841]
20. Cartright M, White R, Horvitz E. Intentions and attention in exploratory health search. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. 2011 Presented at: SIGIR '11; Jul 24-28, 2011; Beijing, China p. 65-74. [doi: 10.1145/2009916.2009929]
21. Chin J, Fu W, Kannampallil T. Adaptive information search: age-dependent interactions between cognitive profiles and strategies. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI '09; Apr 4-9, 2009; Boston, MA p. 1683-1692. [doi: 10.1145/1518701.1518961]
22. Chin J, FU WT. Interactive effects of age and interface differences on search strategies and performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10; Apr 10-15, 2010; Atlanta, GA p. 403-412. [doi: 10.1145/1753326.1753387]
23. Cooper CP, Polonec L, Stewart SL, Gelb CA. Gynaecologic cancer symptom awareness, concern and care seeking among US women: a multi-site qualitative study. Fam Pract 2013 Feb;30(1):96-104 [FREE Full text] [doi: 10.1093/fampra/cms040] [Medline: 22948336]
24. Cumming G, Currie H, Moncur R, Lee A. Web-based survey on the effect of digital storytelling on empowering women to seek help for urogenital atrophy. Menopause Int 2010;16(2):51-55. [Medline: 20729493]
25. De Choudhury M, Ringel Morris M, White RW. Seeking and sharing health information online: comparing search engines and social media. 2014 Presented at: SIGCHI Conf Hum Factors Comput Syst; Apr 26-May 1, 2014; Toronto, ON. [doi: 10.1145/2556288.2557214]
26. Fiksdal AS, Kumbamu A, Jadhav AS, Cocos C, Nelsen LA, Pathak J, et al. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. J Med Internet Res 2014;16(10):e224 [FREE Full text] [doi: 10.2196/jmir.3341] [Medline: 25348028]
27. Hay MC, Cadigan RJ, Khanna D, Strathmann C, Lieber E, Altman R, et al. Prepared patients: internet information seeking by new rheumatology patients. Arthritis Rheum 2008 Apr 15;59(4):575-582. [doi: 10.1002/art.23533] [Medline: 18383399]
28. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]
29. Lauckner C, Hsieh G. The presentation of health-related search results and its impact on negative emotional outcomes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13; Apr 27-May 2, 2013; Paris p. 333-342. [doi: 10.1145/2470654.2470702]
30. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]

31. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
32. Morgan T, Schmidt J, Haakonsen C, Lewis J, Della RM, Morrison S, et al. Using the internet to seek information about genetic and rare diseases: a case study comparing data from 2006 and 2011. JMIR Res Protoc 2014 Feb 24;3(1):e10 [FREE Full text] [doi: 10.2196/resprot.2916] [Medline: 24565858]
33. Mueller J, Davies A, Harper S, Jay C, Todd C. Widening access to online health education for lung cancer. In: Proceedings of the 13th Web for All Conference. 2016 Presented at: W4A '16; Apr 11-13, 2016; Montreal, QC p. 35. [doi: 10.1145/2899475.2899495]
34. Norr AM, Capron DW, Schmidt NB. Medical information seeking: impact on risk for anxiety psychopathology. J Behav Ther Exp Psychiatry 2014 Sep;45(3):402-407. [doi: 10.1016/j.jbtep.2014.04.003] [Medline: 24818986]
35. North F, Varkey P, Laing B, Cha SS, Tulledge-Scheitel S. Are e-health web users looking for different symptom information than callers to triage centers? Telemed J E Health 2011;17(1):19-24. [doi: 10.1089/tmj.2010.0120] [Medline: 21214306]
36. Perez SL, Paterniti DA, Wilson M, Bell RA, Chan MS, Villareal CC, et al. Characterizing the processes for navigating Internet health information using real-time observations: a mixed-methods approach. J Med Internet Res 2015;17(7):e173 [FREE Full text] [doi: 10.2196/jmir.3945] [Medline: 26194787]
37. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
38. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Aug;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
39. 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]
40. Teriaky A, Tangri V, Chande N. Use of internet resources by patients awaiting gastroenterology consultation. Turk J Gastroenterol 2015 Jan;26(1):49-52 [FREE Full text] [doi: 10.5152/tjg.2015.6043] [Medline: 25698271]
41. Thomson MD, Siminoff LA, Longo DR. Internet use for prediagnosis symptom appraisal by colorectal cancer patients. Health Educ Behav 2012;39(5):583-588 [FREE Full text] [doi: 10.1177/1090198111423941] [Medline: 21990571]
42. White RW, Horvitz E. Experiences with web search on medical concerns and self diagnosis. AMIA Annu Symp Proc 2009;2009:696-700 [FREE Full text] [Medline: 20351943]
43. White RW, Horvitz E. Predicting escalations of medical queries based on web page structure and content. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2010 Presented at: SIGIR '10; Jul 19-23, 2010; Geneva p. 769-770. [doi: 10.1145/1835449.1835607]
44. White RW, Horvitz E. Web to world: predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. AMIA Annu Symp Proc 2010:882-886 [FREE Full text] [Medline: 21347105]
45. White RW, Horvitz E. Studies of the onset and persistence of medical concerns in search logs. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2012 Presented at: SIGIR '12; Aug 12-16, 2012; Portland, OR p. 265-274. [doi: 10.1145/2348283.2348322]
46. White RW, Horvitz E. Captions and biases in diagnostic search. ACM Trans Web 2013 Oct 01;7(4):1-28. [doi: 10.1145/2486040]
47. 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]
48. Lewis C. Using the “Thinking Aloud” Method in Cognitive Interface Design. Yorktown Heights, NY: IBM TJ Watson Research Center; 1982.
49. White RW, Horvitz E. From web search to healthcare utilization: privacy-sensitive studies from mobile data. J Am Med Inform Assoc 2013 Jan 1;20(1):61-68 [FREE Full text] [doi: 10.1136/amiajnl-2011-000765] [Medline: 22661560]
50. Chae J. Online cancer information seeking increases cancer worry. Comput Hum Behav 2015 Nov;52(C):144-150. [doi: 10.1016/j.chb.2015.05.019]
51. Conner M, Norman P. Predicting and Changing Health Behaviour: Research and Practice with Social Cognition Models. Berkshire, UK: Open University Press; 2015.
52. Palotti J, Zuccon G, Goeuriot L, Kelly L, Hanbury A, Jones G. CLEF eHealth Evaluation Lab 2015, Task 2: Retrieving Information About Medical Symptoms. 2015. URL: http://ceur-ws.org/Vol-1391/inv-pap9-CR.pdf [accessed 2017-05-28] [WebCite Cache ID 6qnp2RoKL]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 06.10.16; peer-reviewed by A Marcu, CJ Lee; comments to author 22.01.17; revised version received 06.03.17; accepted 23.03.17; published 13.06.17.

Please cite as:
Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C
Web Use for Symptom Appraisal of Physical Health Conditions: A Systematic Review
J Med Internet Res 2017;19(6):e202
URL: http://www.jmir.org/2017/6/e202/
doi:10.2196/jmir.6755
PMID:28611017

Viewpoint

Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop

Susan Michie[1], MA, MPhil, DPhil; Lucy Yardley[2,3], BSc, MSc, PhD, CPsychol; Robert West[4], BSc, PhD; Kevin Patrick[5,6], MD, MS; Felix Greaves[7,8], BMBCh, PhD

[1]Centre for Behaviour Change, Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom
[2]Department of Psychology, University of Southampton, Southampton, United Kingdom
[3]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
[4]Health Behaviour Research Centre, University College London, United Kingdom
[5]Department of Family Medicine and Public Health, The Qualcomm Institute - Calit2, University of California, San Diego, CA, United States
[6]The Center for Wireless and Population Health Systems, The Qualcomm Institute - Calit2, University of California, San Diego, CA, United States
[7]Public Health England, Skipton House, London, United Kingdom
[8]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Susan Michie, MA, MPhil, DPhil
Centre for Behaviour Change
Department of Clinical, Educational and Health Psychology
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 2076795930
Fax: 44 2074364276
Email: s.michie@ucl.ac.uk

Abstract

Devices and programs using digital technology to foster or support behavior change (digital interventions) are increasingly ubiquitous, being adopted for use in patient diagnosis and treatment, self-management of chronic diseases, and in primary prevention. They have been heralded as potentially revolutionizing the ways in which individuals can monitor and improve their health behaviors and health care by improving outcomes, reducing costs, and improving the patient experience. However, we are still mainly in the age of promise rather than delivery. Developing and evaluating these digital interventions presents new challenges and new versions of old challenges that require use of improved and perhaps entirely new methods for research and evaluation. This article discusses these challenges and provides recommendations aimed at accelerating the rate of progress in digital behavior intervention research and practice. Areas addressed include intervention development in a rapidly changing technological landscape, promoting user engagement, advancing the underpinning science and theory, evaluating effectiveness and cost-effectiveness, and addressing issues of regulatory, ethical, and information governance. This article is the result of a two-day international workshop on how to create, evaluate, and implement effective digital interventions in relation to health behaviors. It was held in London in September 2015 and was supported by the United Kingdom's Medical Research Council (MRC), the National Institute for Health Research (NIHR), the Methodology Research Programme (PI Susan Michie), and the Robert Wood Johnson Foundation of the United States (PI Kevin Patrick). Important recommendations to manage the rapid pace of change include considering using emerging techniques from data science, machine learning, and Bayesian approaches and learning from other disciplines including computer science and engineering. With regard to assessing and promoting engagement, a key conclusion was that sustained engagement is not always required and that for each intervention it is useful to establish what constitutes "effective engagement," that is, sufficient engagement to achieve the intended outcomes. The potential of digital interventions for testing and advancing theories of behavior change by generating ecologically valid, real-time objective data was recognized. Evaluations should include all phases of the development cycle, designed for generalizability, and consider new experimental designs to make the best use of rich data streams. Future health economics analyses need to recognize and model the complex and potentially far-reaching costs and benefits of digital interventions. In terms of governance, developers of digital behavior interventions should comply with existing regulatory frameworks, but with consideration for emerging standards around information governance, ethics, and interoperability.

(J Med Internet Res 2017;19(6):e232) doi:10.2196/jmir.7126

KEYWORDS

health behavior; psychological theory; mobile applications; behavioral medicine; mHealth; eHealth

Introduction

Programs and devices using digital technology (digital interventions) have great potential to improve population health and the efficiency and reach of health care delivery. Mobile apps, SMS (short message service) messages, wearable and ambient sensors, social media, and interactive websites can improve health by supporting behaviors involved in disease prevention, self-management of long-term conditions, and delivery of evidence-based health care practice. Such interventions also have potential to do harm if they provide inappropriate advice, involve interactions that undermine desired behaviors, inappropriately share data, or are used instead of more effective behavior change interventions.

Many of these digital interventions seek to foster or support behavior change on the part of health care professionals, patients, or the general public. The challenges involved in developing, evaluating, and implementing effective digital behavior change interventions (DBCIs), and preventing the use of counterproductive ones, have only just begun to be delineated, let alone met [1]. Some of the challenges are similar to those faced by other behavior change interventions, but many are unique, including those of pace of development, engagement with the intervention, measurement of effectiveness and cost effectiveness, and compliance with regulatory, ethical, and security requirements. These challenges are set out in more detail in Table 1.

Table 1. Challenges for developing and evaluating digital interventions targeting behavior change.

Topics	Challenges
Pace and efficiency	Rapid technological change and iterative development cycles make it necessary to continually update and adapt interventions.
	Existing development and evaluation cycles are slow and unsuited to dynamic systems and rapidly changing contexts.
	Efficient, continuing relationships between academics and intervention developers are needed for implementation, continued development, and evaluation.
Engagement	Engagement with digital interventions is often too limited to support behavior change.
	Engagement is multidimensional and cannot be evaluated simply by DBCI[a] usage.
	Engagement with DBCIs may be unequal between different groups and at risk of reinforcing disparities or inequalities.
Theory	Often, there is a lack of clarity around the mechanisms through which DBCIs have their effect.
	Methods of characterizing intervention components, mode of delivery, and contexts that characterize their essential features are required but limited.
Evaluation of effectiveness	Controlling the testing environment is made problematic by the ready availability of alternative interventions.
	It is difficult to specify comparator interventions or control conditions that allow meaningful evaluation of the intervention of interest.
	Better methods for structuring and analyzing very large, dynamic, and heterogeneous data sets are needed.
	Reach and engagement can be low.
	The complex multi-component nature of interventions requires an iterative design and testing cycle.
Evaluation of cost-effectiveness	There is a lack of techniques for economic and cost-effectiveness evaluation across the digital development, deployment, and delivery cycle.
	Funding mechanisms are not aligned with the digital model of development, implementation, iterative improvement, and evaluation.
Regulation, ethics, and information governance	There are competing commercial and ethical demands on data ownership and intellectual property.
	There are emerging and different standards around ethical or institutional review in the biomedical, psychological, and digital development communities.
	There are uncertain quality standards and regulatory processes for digital interventions (with standards either in development or inappropriately adapted from other contexts).

[a] DBCI: Digital behavior change interventions.

There are also unique opportunities. For example, the type and amount of data that can be collected creates unprecedented potential to test and advance theories; understanding more about human behavior will enable the development of more effective DBCIs [2]. Rising to these challenges and making the most of opportunities will require the expertise and collaboration of a wide range of academic disciplines such as behavioral, computer, and engineering sciences and user-centered design.

Given the explosion of development and use of DBCIs aimed at improving health, there is a need for recommendations for designing, evaluating, and implementing digital interventions in health care. Such recommendations are needed to (1) identify the scientific principles relevant to developing effective DBCIs, making digital research more efficient and future interventions safer and more effective and (2) to support key disciplines, health care professionals, patients, and the public to work together more effectively to advance research methods and the understanding and techniques of behavior change through digital technology.

To this end, an international workshop of experts in relevant fields was convened to consider the challenges, opportunities, and strategies for advancement and to formulate principles for developing and evaluating DBCIs. The workshop led to a series of publications [3-7]. This paper discusses key recommendations arising from the workshop and subsequent discussion.

Methods

An international expert consensus-building two-day workshop, supported by the United Kingdom’s Medical Research Council and the National Institutes of Health and the Robert Wood Johnson Foundation of the United States was held in September 2015 in London. The 42 participants from four countries were selected to include those who develop, evaluate, use, and fund DBCIs for both research and practical purposes. Participants included health care professionals, population health researchers (eg, systematic reviewers, behavioral scientists, and health economists), and intervention developers. The workshop proposal (led by SM) and steering group are shown in Multimedia Appendix 1.

The primary deliverable from this workshop was a set of journal articles that would summarize key issues for research in DBCIs and a synoptic paper setting out some key recommendations.

Six topics were identified through discussion amongst the participants. Leaders for each topic were identified and participants were asked to sign up to one or more topics on the basis that they would actively contribute to writing the papers. The topics were (1) the pace of technological development, (2) understanding and promoting effective engagement with users, (3) advancing models and theories, (4) evaluation strategies, (5) economic evaluation, and (6) regulation and governance.

The writing groups met by teleconference before the workshop to write draft papers that were circulated to the whole group in advance of the workshop. These drafts provided the structure and basis of the discussions at the workshop. A formal consensus process was not used, but the structured and open discussions did not reveal any fundamental disagreements about the nature of the recommendations, while at the same time supporting their refinement and specificity. Scribes were appointed from the writing groups to take and circulate notes of relevant points for each paper and the proceedings were audio-recorded. The writing groups developed five articles after the workshop, now published in the American Journal of Preventive Medicine (AJPM) [1,3-7].

The five articles informed the recommendations presented in this article. All authors were workshop participants (SM was PI for the project and LY and KP were Editors of the *AJPM* Special Section).

Results

The key recommendations for developing and evaluating digital interventions are summarized in Table 2.

Achieving Rapid and Efficient Development

Behavior change interventions are moving rapidly from their historical roots in interpersonal counseling and mass communication to the increasingly pervasive world of ubiquitous personal mobile and social technologies. Thus, the methods by which we deploy, evaluate, and improve these DBCIs are moving from relatively data-poor, infrequent, and typically post-hoc assessments to methods that incorporate continuous measurement of the intervention effects in real time [3]. Related to these issues, the following recommendations emerged from the discussion.

Consider Adopting Methods From Engineering and Other Data-Intensive Domains in the Development Cycle

Rather than using deployment-evaluation cycles in which successive measures are collected every few weeks or months, a new type of rapid feedback approach is possible. Theories, models, and methods to support this approach can be found in systems engineering and related sectors, for example, control theory [8], use of factorial or fractionated evaluation designs [9], and system optimization strategies [10] (please refer to [2] and Multimedia Appendix 2 for more details).

Use Bayesian and Related Approaches to Improve the Predictive Modeling Capabilities of Digital Behavior Change Interventions

In addition to enabling more agile evaluation of interventions as they are deployed, sciences outside the traditional medical, public health, and social or behavioral arenas can inform modeling and prediction when multiple behaviors are addressed, when these behaviors have consequences on other important health-related parameters, and for all of these over time and across populations. As with the previous recommendation, the change in scale of the data now attainable from new technologies is analogous to the changes that happened in the field of meteorology when multi-scale and multilevel sensors, combined with improved computing capabilities, enabled weather prediction models to be rapidly built, tested, improved, and retested [11].

Table 2. Summary of recommendations according to topic.

Achieving rapid and efficient development	Understanding and promoting engagement	Advancing models and theories	Evaluating effectiveness	Evaluating cost-effectiveness	Ensuring regulatory, ethical, and information governance
Consider adopting methods from engineering and other data-intensive domains in the development cycle.	Specify and establish empirically what constitutes “effective engagement” for each DBCI[a], that is, sufficient engagement to achieve the intended outcomes.	Use the large amounts of real-time, ecologically valid data generated by DBCIs to test and advance models and theories of behavior change.	Evaluate at all phases in the development cycle.	At every stage, including concept development, identify all the relevant future costs and benefits.	Ensure compliance with appropriate ethics or institutional review board processes.
Use Bayesian and related approaches to improve the predictive modeling capabilities of DBCIs.	Identify and develop valid and efficient combinations of objective and subjective measures to build and test multidimensional models of engagement.	Develop methods able to efficiently analyze large, complex data sets to test dynamic theoretical propositions and allow personalization of DBCIs.	Design evaluations for generalizability.	Take account of projected uptake as well as reach.	Identify and adhere to regulatory processes that may be required for digital medical devices.
Leverage advances in data science such as machine learning, but ensure that human input is retained as needed.	Develop DBCIs with a person-centered and iterative approach, using mixed methods to progressively refine the DBCI to meet user requirements.	Specify the circumstances in which a proposed mechanism of action of a DBCI will produce a targeted effect and build an ontology to organize knowledge resulting from this.	Use methods of DBCI evaluation that capitalize on their unique characteristics.	Select a modeling framework appropriate for the complexity of the projections.	Ensure compliance with national standards for data handling, sharing, and interoperability, where appropriate.
		Develop DBCIs using a modular approach.	Use features of DBCIs to optimize control and access rich data streams.	Separately evaluate societal, personal, and health care cost-effectiveness.	Provide clear and transparent information on how data from the intervention will be used and shared.
		Support interdisciplinary research collaborations and transdisciplinary thinking.	Choose comparators that minimize contamination.		

[a]DBCI: Digital behavior change interventions.

Leverage Advances in Data Science Such As Machine Learning, But Ensure That Human Input Is Retained As Needed

Machine learning and related approaches are increasingly being used to solve big data challenges, including health behavior assessment and interventions [12]. This is particularly the case in the move beyond “on average” effects to personalized inputs and outputs based upon each individual’s situation, characteristics, and desired outcomes. However, we are in the early phase of this new science, so the optimal balance between computer-driven processes and human input is not yet clear: a mix may be needed with the balance determined by both qualitative and quantitative assessments of outcomes.

Understanding and Promoting Engagement

The novel ways in which interventions can be delivered using digital technology result in new ways of engaging with them. Face-to-face behavior change support typically requires users to attend a set number of therapeutic or coaching sessions, whereas users of DBCIs can access support when they feel it is necessary. Analysis of this very different pattern of engagement requires careful consideration of the relationship between the “micro” level of immediate engagement with the digital dimension of the intervention and the “macro” level of engagement with longer-term behavior change (see next section) [13].

Specify and Empirically Establish What Constitutes “Effective Engagement” for Each Digital Behavior Change Intervention, That Is, Sufficient Engagement to Achieve the Intended Outcomes

Acknowledgement of the complex relationship between engagement with the behavioral and technological aspects of the intervention challenges the common assumption that engagement can be measured simply by technology usage. Behavior change may or may not require sustained or in-depth engagement with the digital intervention; hence, technology usage correlates with behavioral outcomes. However, the association is often not strong. For some users and contexts, just one in-depth period of engagement with the DBCI may be sufficient to initiate new habits or teach new skills, whereas for other types of behavior change or other users, brief but timely context-triggered prompts may be needed long-term, whenever the behavior is required [14]. It is therefore important to empirically establish what the “effective” engagement required to achieve behavior change is, in a particular intervention context and for any particular user, as this is likely to differ for different types of interventions and target behaviors.

Identify and Develop Valid and Efficient Combinations of Objective and Subjective Measures to Build and Test Multidimensional Models of Engagement

Measuring effective engagement requires a multi-dimensional, mixed method approach, combining objective assessment of technology usage, behavior, and reactions to the intervention with reports of subjective and offline experiences of users. DBCIs promise exciting new opportunities to collect detailed objective longitudinal data about the antecedents of behavior change, but much work is necessary to develop and validate reliable, non-intrusive means of assessing and analyzing user behavior and its context. Qualitative methods are resource intensive but provide vital complementary insights into user views and behaviors [15]—for example, when not engaging with the digital dimension of the intervention.

Develop Digital Behavior Change Interventions With a Person-Centered and Iterative Approach, Using Mixed Methods to Progressively Refine Them to Meet User Requirements

To promote engagement with DBCIs, a “user-centered” [16] or “person-based” [17] approach is essential to ensure that interventions are responsive to users’ needs and preferences. These approaches are useful in the development of any intervention, but are particularly important for developing DBCIs; human therapists can adjust their advice in real time based on user reactions, whereas the content and delivery of DBCIs must be pre-adapted during development to anticipate a range of user reactions. This involves carrying out iterative qualitative research and stakeholder consultation throughout the design and development process. Tailoring interventions to different needs and preferences can improve engagement, but a well-designed intervention that allows choice is often accessible and engaging for a wide range of users. Adding human support is also known to promote engagement with many interventions [18]. However, as tailoring and human support both increase the cost of interventions, it is important to establish when and how interventions need to be tailored to the individual or supplemented by human support.

Advancing Models and Theories of Behavior Change

Digital technology makes it much easier to collect data in real time and places less reliance on self-report when it comes to recording behavior and taking physiological or physical measurements of study participants or their environment. DBCIs generate large amounts of real-time, ecologically valid data that form digital traces that can be aggregated, connected, and organized to gain greater understanding of how and why behavior changes within an individual over time and how that is influenced by internal physiological and psychological states and the external world. To realize the potential of these data for understanding and changing behavior, a number of recommendations emerged from the discussion (see [4]).

Use the Large Amounts of Real-Time, Ecologically Valid Data Generated by Digital Behavior Change Interventions to Test and Advance Models and Theories of Behavior Change

Data should be collected at a level of granularity that enables the testing and advancement of models and theories of behavior change, accounting for individual variation and changes over time. These data should be used to build dynamic theories of human behavior, modeling not just causal and mediating relationships, but accounting for how effects vary across individuals, contexts, and over time. Theories and models should be continually tested by DBCI-generated data and the results used to systematically refine models and theories.

Develop Methods to Efficiently Analyze Large, Complex Data Sets to Test Dynamic Theoretical Propositions and Allow Personalization of Digital Behavior Change Interventions

Achieving the promise of DBCIs for advancing behavior change theories requires methods that allow vast amounts of complex data to be analyzed and interpreted. For theories and models to be useful in guiding data analysis and interpreting findings, they should be as precise, quantitative, and testable as possible [19]. This allows the building of idiographic models of behavior change and the personalization of DBCIs, that is, the tailoring of the DBCI content and delivery to individuals. It allows DBCIs to adapt, as data are gathered about how the person responds in different contexts. It also enables DBCIs to intervene at opportune moments (sometimes referred to as ecological momentary interventions [20] or “just-in-time” adaptive interventions [14,21].

Specify the Circumstances in Which a Digital Behavior Change Intervention’s Proposed Mechanism of Action Will Produce a Targeted Effect and Build an Ontology to Organize Knowledge Resulting From This

DBCIs represent a qualitative leap in our ability to answer the question posed by researchers, practitioners, and policy-makers: “What works for whom in what settings to change what behaviors, and how?” To optimally organize knowledge about the circumstances (when, where, for whom, and in what state for that person) in which a proposed mechanism of action for a DBCI will produce a targeted effect requires coherent structures and uniform terminologies to describe constructs and their inter-relationships [2,22]. Such knowledge-organizing structures are called “ontologies” [22,23]. Work is beginning to develop an ontology of behavior change interventions. The Human Behaviour Change Project is applying artificial intelligence to refine an ontology for analyzing the up-to-date world literature to answer the question “What works, how well, with what degree of exposure, for whom, in what settings, with what behaviors, and why?” [24].

Develop Digital Behavior Change Interventions Using a Modular Approach

To optimize the effectiveness of DBCIs, theory should be applied to their development and the data generated should be analyzed in terms of the underlying theoretical propositions in order to test and advance theory. This can be done by a modular approach to DBCI development in which modules represent one or more specific theoretical propositions [14]. An example of this approach is “Drink Less,” a mobile app [25] to reduce excessive alcohol consumption in which five theoretically distinct modules were tested in a full factorial trial (Normative feedback, Feedback and self-monitoring, Identity change, Action planning, and Cognitive bias re-training) [26]. This allows the linking of theory to intervention content, the investigation of which techniques and mechanisms is doing the “heavy lifting” in effective interventions and enables the cumulative process of advancing theory and developing more effective DBCIs. Testing and advancing theory is strengthened by supplementing these quantitative data analytic strategies with qualitative research methods.

Support Interdisciplinary Research Collaborations and Transdisciplinary Thinking

Ontologies help both to organize evidence and strengthen interdisciplinary collaboration in developing and evaluating DBCIs informed by models and theories of behavior change. Such collaborations, for example between behavioral scientists, engineers, and computer scientists, will bring unique expertise to bear upon the challenges described in this paper, but also, in the right circumstances, generate new understandings and knowledge that cross disciplinary boundaries, with the whole greater than the sum of the parts [27].

Evaluating Effectiveness

Evaluating effectiveness can be challenging for any behavior change intervention because of problems obtaining valid outcome measures and sufficient numbers of participants who are representative of the population of interest, fidelity in terms of delivery of the intervention, varying levels of engagement with the intervention, loss to follow up, and context sensitivity of the findings. Evaluating DBCIs presents a particular blend of challenges, some similar to other behavior change intervention modalities and some different. Here, we highlight some key recommendations (see [5]).

Evaluate at all Phases in the Development Cycle

Evaluation must be built into the development cycle from the initial concept, and then carried through to prototypes and the final implemented version. This involves concept testing, user testing, factorial experiments, randomized controlled trials (where feasible and appropriate), and testing against baseline after implementation (A-B testing) [2]. For an extensive list with definitions of options for DBCI evaluation, please see Multimedia Appendix 2, drawn from a published guide to developing and evaluating DBCIs in health care [2].

Design Evaluations for Generalizability

Given the high context sensitivity of DBCI effectiveness, evaluations need to be set up in such a way that inference beyond the strict testing conditions can be justified. Generalization may be to different types of populations, the same type of population at a later date, or a different implementation of the DBCI. For example, rapidly changing fashions with regard to design and use of interactive components means that studies need to be designed to allow plausible generalizations beyond specific DBCI features. Qualitative research has a contribution to make in identifying and informing aspects of delivery and context that increase understanding of generalizability.

Use Agile Methods of Digital Behavior Change Intervention Evaluation That Capitalize on Unique Characteristics

The continuous and data-intensive nature of DBCIs, combined with the rapidly changing technologies that support them, challenge traditional research designs such as randomized controlled trials (RCTs) with locked-down interventions and research approaches that involve lengthy recruitment, enrolment, and study periods. While RCTs will continue to have value, they should be complemented with adaptive research designs, A/B testing, N-of-1 studies, and other research methods that yield insights in a shorter time frame or in ways that reflect the granular nature of the intervention effects [28]. DBCIs also provide the opportunity to engage the user directly in the design of the intervention, including tailoring of preferences based upon context and changing life circumstances.

Use Features of Digital Behavior Change Interventions to Optimize Control and Access Rich Data Streams

DBCIs can deliver complex personalized interventions with high fidelity, and engagement with their components can be assessed automatically. This provides an unrivalled opportunity to undertake factorial experiments to assess the effectiveness of components, but presents major challenges in terms of conceptualizing and analyzing very large temporally structured data streams. Analytical methods need to be devised to address this challenge, but their value will be limited without a theoretical underpinning to structuring and aggregating data.

Choose Comparators That Minimize Contamination

For many DBCIs, such as smoking cessation or weight management apps, highly developed products are readily available to study participants. Therefore, researchers face a difficult choice between a comparator that has sufficient active components to be credible and deter searching for alternatives and one that has so much active content that the “true” effect of the intervention cannot be measured. In practice, for many DBCIs, we have to accept that we may never be able to assess their full impact, only their impact relative to another active intervention.

Evaluating Cost-Effectiveness

Rigorous economic evaluation is required to inform decision makers about allocation of their scarce resources. A central motivation for DBCIs is that they may be cost effective—and certainly cheaper than their face-to-face counterparts—because they can rapidly increase scale at minimal cost. Generating accurate evidence of cost effectiveness will be essential in gaining support from health system payers, but it requires additional consideration beyond conventional analysis (see [7]).

At Every Stage, Including Concept Development, Identify all the Relevant Future Costs

As with all interventions, economic evaluation should be considered from the start of intervention development to ensure that all cost data are collected. However, existing economic approaches (eg, ISPOR guidance [29]) will need to be adapted to take into account the way digital health interventions are delivered, covering the whole life cycle of the intervention. This may include development, implementation, update, and eventual obsolescence. While scaling interventions may be cheap, with a low additional unit cost, development costs can be high at the start of the process and need to be fully included in calculations. The rapid cycle of development for many DBCIs, with multiple versions of a product that evolves with small iterative and incremental improvements, is different from that of the relatively fixed products common in the pharmaceutical and medical devices industries. The cost of these iterations must be captured and included across the lifespan of the intervention. In addition, as many DBCIs may be more effective when supplemented with human interactions, evaluations may need to consider these less cheaply scaled human costs.

Take Projected Uptake and Reach into Account

Economic evaluations will need to understand issues of uptake, reach, and retention of the intervention, as these will be central factors in any projection of benefit. Consideration will also need to be given to the effects of interventions around those immediately engaged, including effects on wider social networks. For example, given the typically high development costs of a DBCI, the cost-effectiveness can depend critically on getting a large number of users. Economic evaluations need to be able to draw on evidence as to the likely uptake of a DBCI to assess this, which may include costs incurred in promoting the DBCI to help drive uptake.

Select a Modeling Framework Appropriate for the Complexity of the Projections

DBCIs are complex interventions by nature; they are responsive, with multiple components. Economic appraisals will need to recognize this complexity—in terms of interventions, outcomes, and causal pathways [30]—and use appropriate methods [31] to assess them. More complex interventions may require the use of sophisticated modeling techniques, including agent-based approaches, which capture dynamic interactions between the intervention, the population it is applied to, and the wider environment.

Separately Evaluate Societal, Personal, and Health Care Cost-Effectiveness

Finally, cost effectiveness analysis must be diligent in measuring effect across its forms. This might include benefit (1) to an individual, for example, in improved quality of life, (2) to society, for example, in improved productivity of the workforce, and (3) to the health care system, for example, in reduced referrals to conventional face-to-face settings or hospital admissions. Ideally, disaggregation of these separate components should be attempted, if possible.

Ensuring Regulatory, Ethical, and Information Governance

Beyond the practical and measurable aspects of development and evaluation, DBCIs present new questions in terms of ethics and regulation as they challenge existing frameworks and rules often designed in a pre-digital era.

Ensure Compliance With Rules on Ethics or Institutional Review Board Processes

The diversity of disciplines involved in the development of new DBCIs is challenging as the expectations and approaches of different cultures involved may not be immediately aligned. The notion of a minimum viable product and “fail fast, fail often” might be encouraged in an engineering world view, but it stands in sharp contrast to the “first do no harm” spirit of the health care profession and the risk-averse nature of clinical governance and patient safety. Developers may need to adjust to the more stringent regulatory frameworks of the biomedical sector and local ethics or institutional review processes. At the same time, regulators will need to make improvements in terms of speed and responsiveness in order to be able to meet the needs of fast evolving technologies. This was emphasized in the United States in a 2016 report from the National Academies of Science, Engineering, and Medicine [32], which advocated reflecting the contemporary realities of new research contexts, including accessibility and use of personal data.

Identify and Adhere to Regulatory Processes That May Be Required for Digital Medical Devices

DBCIs will need to navigate the complex landscape of regulation and governance. Many of these interventions fall at the blurred edges of existing regulatory frameworks, which evolved in a pre-digital and less software-intensive era. Depending on their nature, many DBCIs will be classified as medical devices. A lack of consistency of approach between countries means that those developing interventions will need to carefully monitor and engage with the emerging frameworks of the jurisdictions they work within. Regulatory organizations include the United Kingdom’s Medicines & Healthcare Products Regulatory Agency, which has issued specific guidance on apps [33] and the US Federal Trade Commission, which has released similar advice [34]. In addition, although several frameworks for assessment exist, there is diversity in views about what the necessary domains of quality are for DBCIs. Developers will need to understand this variation and create the appropriate evidence to match these specifications. Developing assessment frameworks in the United Kingdom include the National Health Service (NHS) app assessment model [35] and the European Union mHealth assessment guidelines [36].

Ensure Compliance With National Standards for Data Handling, Sharing, and Interoperability, Where Appropriate

There are unique challenges of information governance, as DBCIs may produce large amounts of personal, identifiable, and potentially valuable data. As data is often collected directly from mobile devices, this may include geographical information or information related to contacts and social networks. Those developing interventions will have to comply with local legislation that varies in scope and detail. Relevant legislation includes the Health Insurance Portability and Accountability Act (HIPAA) [37] in the United States, the Data Protection Act [38] (together with the Caldecott principles) in the United Kingdom, and the General Data Privacy Regulation in the EU [39]. In addition, DBCIs may have more benefit if they are able to formally interact with health care delivery systems and, in particular, with electronic health records, requiring compatibility with local (and often highly variable) standards.

Provide Clear and Transparent Information on How Data From the Intervention Will Be Used and Shared

Beyond the essential regulatory frameworks, there are wider societal questions that have not yet been definitively answered regarding acceptability of data collection and sharing with competing commercial and ethical demands on data ownership. Some interventions are provided on a fee-for-use basis. Others have more complicated business models where a product is provided for free, but data is made available to the provider. There may be a trade-off that the public is willing to make between sharing their own data, effectiveness of interventions, and cost. As the threshold of acceptability is still being understood—and as this threshold may vary between populations and individuals—developers will need to tread cautiously and transparently. In addition, ownership of the data produced by DBCIs is not always clear and may vary by jurisdiction. These data have the potential to be a valuable resource for health care research and service planning [40], but the public may not be aware of these potential uses and may have reservations about how the data are used [41]. Developers should provide clarity about how and with whom they share their data. They may also wish to make sure that the business model for provision of their intervention is openly reported.

Discussion

Future Research

The approaches used to address the compelling health problems of our times should be based in state-of-the-art science in medicine, public health, and the social and behavioral sciences. They must also reflect advances in other disciplines such as engineering, data science, systems science, human-computer interaction, and communications technologies, in particular if these promise to improve the means by which we both understand the determinants of health and improve health behaviors. Innovations in digital health are now emerging on several fronts that demonstrate the utility of methods drawn from these fields. This should not surprise anyone who is engaged as a digital citizen in the increasingly smart and connected world in which we now live. The mobile in our pockets, purses, or backpacks is becoming both an increasingly important window on the potential of digital health and an enabler thereof. Thus, there is a need to consider several implications of these developments and optimally shape them as they go forward.

The ways in which users engage with digital interventions are changing as fast as the technologies for delivering them, and it is important that our methods of conceptualizing and assessing engagement keep pace. Earlier, PC-based digital interventions often replicated a series of therapeutic or coaching “sessions,” and engagement was defined simply as completing sufficient sessions to receive the full therapeutic program. However, mobile users of just-in-time adaptive interventions may just engage briefly at the crucial time-point with precisely the required behavior change support. The “effective” engagement required for behavior change support is therefore likely to differ, depending on the users and their contexts, and can only be determined by analyzing complex patterns of relationships between usage, user experiences, and outcomes. A challenge for future research into DBCIs is to find the most valid and efficient combinations of methods of measuring engagement with the intervention, both on the Internet and offline. Technology can provide detailed, unobtrusive assessment of behavior and its context, while complementary qualitative methods are crucial to fully understand and interpret user experiences. These qualitative approaches are central to participatory user-centered design, which is the key to developing and evaluating DBCIs in order to ensure that they are engaging and effective [16,17].

With respect to theory, DBCIs provide an unprecedented opportunity to test and advance our understanding of how behavior changes. However, this requires the application of more “agile” science [42]. This includes designing DBCIs on a modular basis with a clear understanding of which behavior change techniques [43] constitute that module and how they link to the theoretically postulated mechanisms of action. Collaborations between behavioral scientists and computer scientists are enabling theories of change to be specified in precise detail as sets of constructs and specified relationships between them. Computational models can define and test not just interactions between modules, mechanisms of action, and behaviors, but also how these vary across individuals and populations, settings, and time. This is spawning a fruitful area of “ontology” development, ontologies being well specified structures for organizing knowledge [22]. By combining this with computational algorithms and machine learning, the promise is that we can advance theoretical understanding of behavior change so that we can answer with confidence the questions put by health practitioners and policy makers: “What works, how well, for whom, in what settings, for what behaviors, and why?” [24].

With respect to evaluation, one key effectiveness question for policy makers and practitioners is whether purchasing or commissioning a given intervention will improve outcomes for their population compared with current practice or another innovation that they may be considering. Another is whether digital interventions can reduce the cost of health care provision without a significant loss of effectiveness. Ideally, this would involve pragmatic randomized trials (individual or cluster) comparing the different forms of care. However, the analysis in this paper indicates that this is often impracticable. Moreover, there are so many unknowns in the development of these interventions that a much more iterative cycle of development and testing is needed before getting to the point of a pivotal evaluation. Even when an intervention has been found to be effective in a pivotal evaluation, there are no guarantees that it will continue to be effective in a rapidly changing context. This means that the evaluation strategy for DBCIs needs to be fully integrated into every phase of development and implementation and needs to take advantage of the wide range of methodologies that are available, both experimental and observational, while also taking into account the exposure as well as effect in those exposed. The evaluation needs to take advantage of statistical techniques, particularly Bayesian analyses that promote ongoing accumulation of evidence.

With respect to the economic evaluation, policy makers and service providers will be hungry for the potential to deliver interventions at scale with only small increases in marginal cost, often in comparison with face-to-face interventions. This creates a requirement for evaluators to adapt existing economic approaches and develop new ones to accurately capture cost across the full digital development cycle, measure the reach of the intervention, and consider the complexity of the intervention to allow meaningful comparison. In evaluating both effectiveness and cost effectiveness, policy makers and organizations that consider purchasing these services will need to compare them with conventional delivery models and evaluation approaches. Thus, developers and evaluators will need to ensure that they produce evidence that allows this comparison to be made.

With respect to ethical, regulatory, and policy issues, standards for DBCIs are still emerging, and people developing and testing them will have to pick a cautious path through rules not often designed with this new category of interventions in mind. They will need to ensure that new, agile approaches to testing DBCIs demonstrate that they meet the exacting expectations of the clinical and biomedical community. Possibilities to develop common standards for intra-operability between interventions, with electronic health records and health care delivery systems, and in common behavioral data repositories should be explored.

The advice provided here is specific to DBCIs, but it should also build on the good practice of reporting general mHealth interventions published elsewhere to ensure adequate description of the technical components of the intervention [44,45].

Conclusions

DBCIs present unique methodological challenges. New techniques and approaches are becoming available that offer opportunities to accelerate both the development and evaluation of these interventions, taking advantage of the speed and volume of data they generate, their potential adaptability, reach, and cost. These new methods have implications for changes within other fields of health care.

Acknowledgments

This article is an output from a two-day international workshop on how to create, evaluate, and implement effective digital health and behavior change interventions held in London in September 2015. The proposal and Steering Group are in Multimedia Appendix 1. The key themes and working papers of the workshop formed the basis of five articles in the American Journal of Preventive Medicine, cited in this article. The authors of this article attended the workshop and agreed to develop the key recommendations from the workshop into a peer-reviewed publication. SM produced the initial draft and all the authors contributed to writing the article. SM is guarantor. We would like to thank Dr. David Crosby, Programme Manager of the MRC-NIHR Methodology Research Programme, for his support in developing the workshop that underpinned this work.

This article is an output from an international workshop supported by the Medical Research Council (MRC), National Institute for Health Research (NIHR), the Methodology Research Programme (PI Susan Michie), and the Robert Wood Johnson Foundation (PI Kevin Patrick). RW was funded by Cancer Research, UK.

Conflicts of Interest

Susan Michie is Director of the Centre for Behaviour Change, UCL, which has received funds from industry and government agencies. Robert West has undertaken research and consultancy for, and received travel funds from, companies that develop and manufacture smoking cessation medications (Pfizer, GSK, and J&J). He is an unpaid advisor to the UK’s National Centre for Smoking Cessation and Training. His salary is funded by Cancer Research, UK. Felix Greaves is employed by Public Health England. Kevin Patrick and Lucy Yardley have none.

Multimedia Appendix 1

How to create, evaluate, and implement effective digital healthcare interventions: development of guidance.

[PDF File (Adobe PDF File), 54KB - jmir_v19i6e232_app1.pdf]

Multimedia Appendix 2

Options for testing DBCIs.

[PDF File (Adobe PDF File), 28KB - jmir_v19i6e232_app2.pdf]

References

1. Yardley L, Choudhury T, Patrick K, Michie S. Current issues and future directions for research into digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):814-815. [doi: 10.1016/j.amepre.2016.07.019] [Medline: 27745680]
2. West R, Michie S. A Guide to the Development and Evaluation of Digital Behaviour Change Intervention. Great Britain: Silverback Publishing; 2016.
3. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
4. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832. [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
5. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
6. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
7. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and undertaking a health economics study of digital health interventions. Am J Prev Med 2016 Nov;51(5):852-860. [doi: 10.1016/j.amepre.2016.05.007] [Medline: 27745685]
8. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. Drug Alcohol Depend 2007 May;88(Suppl 2):S31-S40 [FREE Full text] [doi: 10.1016/j.drugalcdep.2006.10.020] [Medline: 17169503]
9. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. Am J Prev Med 2014 Oct;47(4):498-504 [FREE Full text] [doi: 10.1016/j.amepre.2014.06.021] [Medline: 25092122]
10. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
11. Shuman F. History of numerical weather prediction at the National Meteorological Center. Weather Forecast 1989;4(3):286-296. [doi: 10.1175/1520-0434(1989)004<0286:HONWPA>2.0.CO;2]
12. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas 2014 Nov;35(11):2191-2203 [FREE Full text] [doi: 10.1088/0967-3334/35/11/2191] [Medline: 25340969]
13. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
14. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34S:1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
15. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
16. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
17. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
18. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
19. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of Behaviour Change Theories: An Essential Resource for Researchers, Policy Makers and Practitioners. Great Britain: Silverback Publishing; 2014.
20. Patrick K, Intille SS, Zabinski MF. An ecological framework for cancer communication: implications for research. J Med Internet Res 2005 Jul 01;7(3):e23 [FREE Full text] [doi: 10.2196/jmir.7.3.e23] [Medline: 15998614]
21. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014 Jul;13(3):13-17. [doi: 10.1109/MPRV.2014.46]
22. Larsen KR, Michie S, Hekler EB, Gibson B, Spruijt-Metz D, Ahern D, et al. Behavior change interventions: the potential of ontologies for advancing science and practice. J Behav Med 2016 Aug 01;40(1):6-22. [doi: 10.1007/s10865-016-9768-0] [Medline: 27481101]
23. Arp R, Smith B, Spear A. Building Ontologies with Basic Formal Ontology. Cambridge: The MIT Press; Jul 2015.

24. Human Behaviour Change Project (HBCP). UCL. The human behaviour-change project website URL: http://www.ucl.ac.uk/human-behaviour-change [accessed 2016-11-25] [WebCite Cache ID 6mHchILVO]
25. Drink Less Alcohol. 2016. URL: http://www.drinklessalcohol.com [accessed 2016-11-29] [WebCite Cache ID 6mNVNeRcx]
26. Garnett C, Crane D, Michie S, West R, Brown J. Evaluating the effectiveness of a smartphone app to reduce excessive alcohol consumption: protocol for a factorial randomised control trial. BMC Public Health 2016 Jul 08;16:536 [FREE Full text] [doi: 10.1186/s12889-016-3140-8] [Medline: 27392430]
27. Christmas S, Michie S, West R. Thinking about behaviour change: an interdisciplinary dialogue. Great Britain: Silverback Publishing; 2015.
28. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
29. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-an ISPOR good research practices task force report. Value Health 2015 Mar;18(2):161-172 [FREE Full text] [doi: 10.1016/j.jval.2015.02.001] [Medline: 25773551]
30. Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, et al. Complex interventions and their implications for systematic reviews: a pragmatic approach. J Clin Epidemiol 2013 Nov;66(11):1209-1214. [doi: 10.1016/j.jclinepi.2013.06.004] [Medline: 23953085]
31. Shiell A, Hawe P, Gold L. Complex interventions or complex systems? implications for health economic evaluation. BMJ 2008 Jun 07;336(7656):1281-1283 [FREE Full text] [doi: 10.1136/bmj.39569.510521.AD] [Medline: 18535071]
32. Committee on Federal Research Regulations and Reporting Requirements. Optimizing the Nation's Investment in Academic Research A New Regulatory Framework for the 21st Century. Washington, DC: The National Academies Press; Jul 27, 2016.
33. Medicines and Healthcare products Regulatory Agency. Guidance: Medical device stand-alone software including apps. London, UK: MHRA; Aug 08, 2014.
34. Federal Trade Commission. FTC. 2016. Mobile health apps interactive tool URL: https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-apps-interactive-tool [accessed 2017-06-08] [WebCite Cache ID 6r427izow]
35. National Information Board. Work stream 1.2: enable me to make the right health and care choices. In: Personalised Health and Care 2020: Using data and technology to transform outcomes for patients and citizens. London, UK: HM Government; Nov 13, 2014.
36. European Commission. Open stakeholder meeting on mHealth assessment guidelines: Presentations and survey. In: EU guidelines on assessment of the reliability of mobile health applications. 2016 Presented at: Open stakeholder meeting on mHealth assessment guidelines; May 4, 2016; Brussels, Belgium URL: https://ec.europa.eu/digital-single-market/en/news/open-stakeholder-meeting-mhealth-assessment-guidelines-presentations-and-survey
37. US Government Publishing Office. Conference Report. In: Health Insurance Portability and Accountability Act of 1996. 1996 Jul 31 Presented at: 104th Congress, 2d session; July 31, 1996; House of Representatives, Washington DC URL: https://www.gpo.gov/fdsys/search/pagedetails.action?granuleId=CRPT-104hrpt736&packageId=CRPT-104hrpt736&bread=true
38. Data Protection Act 1998. London: The London Stationary Office; 1998.
39. European Union. Regulation (EU) of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Brussels: Official Journal of the European Union Commission; 2016.
40. Evans BJ. Barbarians at the gate: consumer-driven health data commons and the transformation of citizen science. Am J Law Med 2016;42(4):1-32 [FREE Full text]
41. Kostkova P, Brewer H, de LS, Fottrell E, Goldacre B, Hart G, et al. Who owns the data? open data for healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
42. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016 Jun;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
43. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
44. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [Medline: 26988021]
45. Eysenbach G, Consort- EHEALTH Group. Consort-eHealth: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

DBCI: digital behavior change intervention
EU: European Union
HIPAA: Health Insurance Portability and Accountability Act
MRC: Medical Research Council
NHS: National Health Service
PI: Principal investigator
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 14.12.16; peer-reviewed by M Graham, L Tarzia, M Allman-Farinelli; comments to author 18.02.17; revised version received 28.02.17; accepted 22.04.17; published 29.06.17.

Please cite as:
Michie S, Yardley L, West R, Patrick K, Greaves F
Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop
J Med Internet Res 2017;19(6):e232
URL: http://www.jmir.org/2017/6/e232/
doi:10.2196/jmir.7126
PMID:28663162

Original Paper

Defining and Predicting Patterns of Early Response in a Web-Based Intervention for Depression

Wolfgang Lutz[1], PhD; Alice Arndt[1], MSc; Julian Rubel[1], PhD; Thomas Berger[2], PhD; Johanna Schröder[3], PhD; Christina Späth[4], Dipl Psych; Björn Meyer[5], PhD; Wolfgang Greiner[6], PhD; Viola Gräfe[6], MSc; Martin Hautzinger[7], PhD; Kristina Fuhr[7], PhD; Matthias Rose[8], MD; Sandra Nolte[8], PhD; Bernd Löwe[9], MD; Fritz Hohagen[4], MD; Jan Philipp Klein[4], MD; Steffen Moritz[3], PhD

[1]Department of Psychology, University of Trier, Trier, Germany
[2]Departmemt of Psychology, University of Bern, Bern, Switzerland
[3]Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[4]Department of Psychiatry and Psychotherapy, Lübeck University, Lübeck, Germany
[5]GAIA AG, Hamburg, Germany
[6]Department of Health Economics and Health Care Management, Bielefeld University, Bielefeld, Germany
[7]Department of Psychology, Eberhard Karls University Tuebingen, Tuebingen, Germany
[8]Department of Psychosomatic Medicine, Charité University Medical Center, Berlin, Germany, Berlin, Germany
[9]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
Wolfgang Lutz, PhD
Department of Psychology
University of Trier
Am Wissenschaftspark 25-27
Trier,
Germany
Phone: 49 651 201 2883
Fax: 49 651 201 2886
Email: wolfgang.lutz@uni-trier.de

Abstract

Background: Web-based interventions for individuals with depressive disorders have been a recent focus of research and may be an effective adjunct to face-to-face psychotherapy or pharmacological treatment.

Objective: The aim of our study was to examine the early change patterns in Web-based interventions to identify differential effects.

Methods: We applied piecewise growth mixture modeling (PGMM) to identify different latent classes of early change in individuals with mild-to-moderate depression (n=409) who underwent a CBT-based web intervention for depression.

Results: Overall, three latent classes were identified (N=409): Two early response classes (n=158, n=185) and one early deterioration class (n=66). Latent classes differed in terms of outcome (*P*<.001) and adherence (*P*=.03) in regard to the number of modules (number of modules with a duration of at least 10 minutes) and the number of assessments (*P*<.001), but not in regard to the overall amount of time using the system. Class membership significantly improved outcome prediction by 24.8% over patient intake characteristics (*P*<.001) and significantly added to the prediction of adherence (*P*=.04).

Conclusions: These findings suggest that in Web-based interventions outcome and adherence can be predicted by patterns of early change, which can inform treatment decisions and potentially help optimize the allocation of scarce clinical resources.

(J Med Internet Res 2017;19(6):e206) doi:10.2196/jmir.7367

KEYWORDS

patterns of early change; depression; web interventions; psychotherapy research

Introduction

Web-based interventions for individuals with depressive disorders have been a recent focus of research and may be an effective addition to face-to-face psychotherapy or pharmacological treatment. For example, such interventions may be appropriate for individuals who have difficulty accessing psychological treatment or do not want to utilize face-to-face treatment [1-5]. Several studies suggest that some forms of Web-based interventions may be as effective as face-to-face therapy [6], although various methodological limitations of this body of research have also been noted [7]. One limitation of Web interventions is that they are not accepted by all patients and some drop out early or do not adhere to the treatment protocol [8]. Especially in unguided Web interventions, the risk of dropout is high [9,10] and results of studies on prereatment predictors of outcome in Web interventions remain inconsistent [11]. Additionally, not all Web interventions are equal with regard to their quality or evidence base [12]. So far, investigations of Web interventions have mainly focused on treatment efficacy and short-term symptom change in comparison with treatment-as-usual control groups, in which participants were only able to access the Web intervention after a delay of several weeks or months [13,14].

Whereas a good database has been established regarding the general effectiveness of several Web-based interventions for the treatment of psychological problems, there is still a lack of research investigating the process and shape of change [15]. On the other hand, this area of research has a certain tradition in individual therapy. In recent years, interest in the investigation of early change patterns and their relation to outcome has grown. The basic idea behind this research is to use early change of the target behavior (eg, depressive symptoms) to predict treatment outcome [16,17]. Early change patterns have been shown to be associated with outcome across different diagnoses [18,19], different treatment approaches [20,21], and different measures [22].

For example, a recent study investigated early change patterns in patients with panic disorder (n=326), who underwent manualized cognitive behavioral therapy (CBT) [23]. Using growth mixture modeling (GMM), 4 latent subgroups were identified, showing clusters of change trajectories over the first 5 sessions. One of the subgroups consisted of patients whose symptoms decreased rapidly and who also showed the best outcomes (early responders). This information on early response improved treatment prediction by 16.1% over patient intake characteristics. Early change patterns also significantly predicted early dropout. Likewise, a further study focused on early change patterns during low intensity interventions [24]. This study used data from patients with anxiety disorders or depression, who accessed the Improving Access to Psychological Therapies (IAPT) service in the United Kingdom (n=511). This service comprised between 1 and 8 sessions and was often delivered via telephone. Early response, defined as reliable improvement until session 4, was predictive of clinically significant recovery after treatment termination. It was noted that attrition was highest in early sessions, so that early attempts to engage patients should be made.

This example emphasizes the importance of also studying early change patterns in low intensity and Web interventions. Whereas conclusions drawn from efficacy and effectiveness studies are limited to the average patient after treatment, knowledge about early change patterns may help to answer clinical questions. For example, such knowledge could be used to predict whether the treatment in question will work for a particular subgroup or to decide whether users should continue treatment [25,26]. Such questions have become increasingly relevant in the face of the recent implementation of stepped-care models, where patients are matched to a treatment with the option of being “stepped-up” to more intensive care [27,28]. Knowledge about predictive determinants may add to the development of empirically based rules that support clinicians in their decisions [28] and may help to prevent dropout or low adherence. In addition, investigating change patterns promotes the understanding of change processes, which is necessary for treatment development efforts [15].

Although early change patterns are important predictors of treatment outcome [23,29], to date only one study has looked at early change patterns in patients undergoing Web-based interventions. Schibbye and colleagues [30] examined change patterns during a CBT-oriented Web-based intervention, which was provided to patients with panic disorder, social phobia, or depression (n=112) by the Internet Psychiatry Clinic in Sweden. Outcome of the Web intervention was predicted by estimation of early change. The prediction was best when the rating of a disorder-specific measure at week 4 was used.

In the present study, we analyzed data from a multicenter trial testing the efficacy of a CBT-oriented web-based intervention for individuals with mild to moderate depression. Based on the existing literature on individual therapy, we predicted the existence of distinct early patient response clusters in this Web intervention. We further hypothesized that these clusters would add to the prediction of treatment outcome as well as adherence. This study also examined whether initial impairment, attitudes toward Web-based interventions, and email support predict early change patterns.

Methods

Participants and Treatment

This study was conducted from January 2012 to December 2013 and approved by the local ethics committee (DGPs, reference number SM 04_2012). Written informed consent was obtained, and the study was registered at ClinicalTrials.gov (identifier: NCT01636752). Several settings were used to recruit participants: (1) In- and outpatient medical and psychological clinics, (2) internet forums for depression, (3) health insurance companies, and (4) the media (eg, newspaper). Participants were directed to the study’s website. In total, 2020 participants signed up for the study and were screened for inclusion and exclusion criteria. Inclusion criteria consisted of (besides Internet access) mild-to-moderate depressive symptoms defined by scores between 5 and 14 on the Patient Health Questionnaire-9 (PHQ-9) and ages between 18 and 65 years. Participants who fulfilled these criteria were further screened by telephone using the Mini International Neuropsychiatric Interview (M.I.N.I. [31]). Also, a baseline assessment was conducted using several self-report measures (see below). If PHQ-9 scores were above 14, acute suicidality was determined or a lifetime diagnosis of bipolar disorder or schizophrenia was identified in the interview [32], participants were excluded from the study, and professional help was suggested to them. Included participants (n=1013) were then randomized into either an intervention group (IG), in which a CBT-oriented Web-based intervention (Deprexis) was delivered in addition to care as usual (IG; n=509), or into a control group (CG), which solely consisted of care as usual (CG; n=504). During the study, participants in the care as usual group did not receive any Web intervention. The use of other interventions initiated by the participants in the care as usual group was measured during the course of the study. At posttreatment, participants reported having utilized the following treatments during the course of the study: medication, treatment by a psychotherapist, treatment at an outpatient clinic, and treatment at an inpatient clinic. There were no significant differences between participants in the IG and the CG regarding the use of medication (*P*=.54), treatment by a psychotherapist (*P*=.38), treatment at outpatient clinics (*P*=.68), and treatment at inpatient clinics (*P*=.29).

As incentive, all participants were entered into a lottery for 12 iPods after the last assessment. Furthermore, participants in the CG received access to the Web intervention 1 year after baseline assessments.

In addition to the pre- and post-treatment PHQ-9 assessments, which all participants filled out, participants in the IG filled out PHQ-9 assessments every 2 weeks during the course of the study. Furthermore, participants in the IG who had mild symptoms of depression (PHQ-9 scores between 5-9), received the Web-based intervention without any guidance, whereas participants who had moderate depressive symptoms (PHQ scores between 10-14) received the same Web intervention in combination with weekly email support [33,34]. Studies have shown that unguided Web interventions are also effective in the treatment of depression (eg, [35]). However, considering safety and efficacy, more intensive support seemed appropriate for patients with moderate depression.

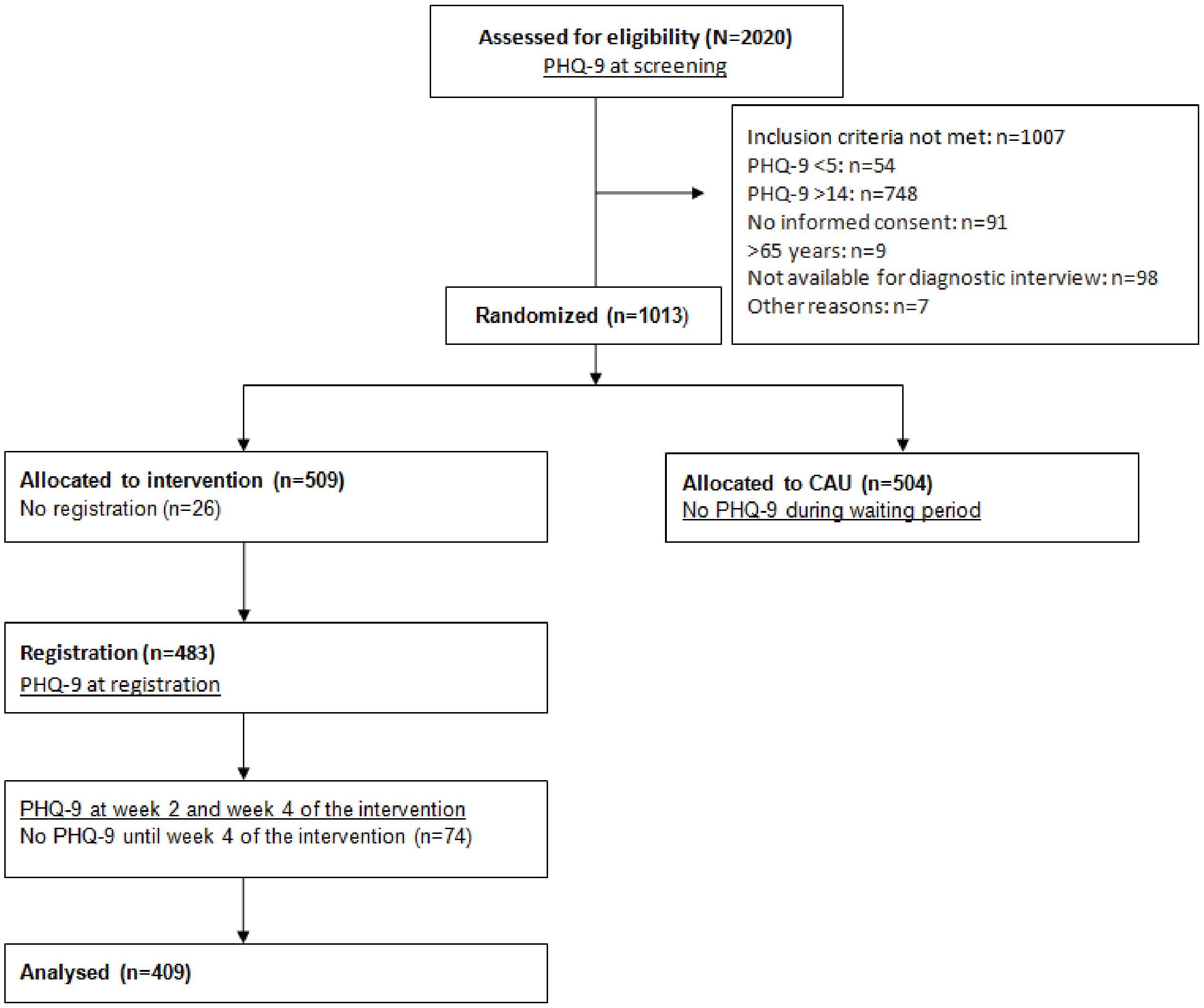
After randomization into the IG, participants had to register on the study's website and were then able to use the Web-based intervention (Deprexis) for a period of 12 weeks. It is based on a cognitive-behavioral approach and consists of 10 modules that are presented in the form of a dialogue or "chat." The modules contain classic CBT elements such as behavioral activation, but also broader therapeutic elements such as mindfulness, emotion-focused interventions, and interpersonal skills. Information as well as advice on the application of the relevant concepts in daily life were combined in the modules, which included text, illustration, and audio. This content was presented in dialogue form, where the user was asked to select one of several response options to the program's explanations. In total, several published randomized controlled trials have provided evidence in support of the program's efficacy, typically with small-to-medium effect sizes [35,32,3].

In this study, our main interest was to examine the change patterns of participants who received the intervention. Therefore, we focused on participants for whom not only pre- and post-assessments were available, but also several PHQ-9 assessments from during participation. This condition was fulfilled by participants in the IG (PHQ-9 every 2 weeks), but not by participants in the CG (pre and post PHQ-9 assessments only). In total, 483 of the 509 participants randomized into the IG registered on the study's website. A total of 409 participants filled out at least one PHQ-9 during the first 4 weeks of the intervention (assessment at week 2 or assessment at week 4, see flowchart in Figure 1) and were therefore included in our study sample. Participants without any assessment during the intervention were excluded, as no meaningful course of change could be modeled for those cases. Participants (n=409) included in the study, and participants who did not register (n=26) or did not complete any assessment during the intervention (n=74) did not differ with regard to age ($F_{2,506}$=1.18, *P*=.31), gender (χ^2_2=2.2, *P*=.34), or initial impairment (PHQ-9 at screening, $F_{2,506}$=1.47, *P*=.23).

On average, 2 weeks passed between screening and registration (standard deviation, SD=1.36). The first assessment took place 2 weeks after registration. Most participants were recruited by online forums (n=82), health insurance companies (n=134), or learned of the study by other means, commonly by news in media (n=236). Other participants learned of the study while in treatment (n=57). Attrition (n=100) did not differ between the recruiting options described previously (χ^2_{12}=18.0, *P*=.12).

On average, participants were 43.16 years old (SD=11.10, range=18-65) and approximately 70% of participants (287/409) were women. Close to 50% of participants had a high-school diploma qualifying for university entrance (204/409). Most participants (n=264) suffered from moderate depression and therefore received the Web intervention as well as additional weekly brief support via email. In this study, early response was estimated based on PHQ-9 score changes between Web-based registration and week 4 of the intervention. In addition to the PHQ-9, other impairment measures and participant attitudes were assessed at screening and posttreatment (see questionnaires below).

Figure 1. Flowchart of participants.



Measures

Diagnostic Interview

Diagnoses were made using the M.I.N.I. [31], and clinician-rated severity of depression was assessed with the 24-item version of the Hamilton Depression Rating Scale (HDRS-24). The M.I.N.I. is a short structured diagnostic interview for Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) and International Classification of Diseases, Tenth Edition (ICD-10) disorders that has been translated into multiple languages. In several studies, it has shown good interrater reliability (eg, [36]). Acute suicidality was assessed based on current suicidal ideation and past suicide attempts. In this study, trained raters (postgraduate students) conducted the interviews via telephone. Before they were permitted to rate trial participants, raters were trained to conduct the interview either face-to-face or via telephone modules and had to demonstrate adequate interrater reliability on an audiotaped interview.

Patient Health Questionnaire-9 (PHQ-9)

The PHQ-9 consists of 9 items that reflect the criteria of depression in DSM-IV [37]. Answers are provided on a 4-point Likert scale (0-“not at all” and 3-“nearly every day”). Thus, total scores range from 0-27, with scores between 5 and 9 indicating mild depression and scores between 10 and 14 indicating moderate depression. The instrument has a good test-retest reliability (r_{tt}=.84) and internal consistency (Cronbach alpha=.86-.89; [37]). To operationalize reliable improvement, the reliable change index (RCI), which reflects the pre- post treatment difference ΔRC large enough to not be attributable to measurement error, was calculated following Jacobson and Truax (see [38]): where r is the reliability of the PHQ-9 (r=.86) and SD the standard deviation of the PHQ-9 intake score (SD=2.37). The RCI score for the PHQ-9 was 2.46 total points.

Short-Form Health Survey-12 (SF-12)

The SF-12 assesses limitations in role functioning with 12 items. It consists of two subscales measuring physical health ($\text{SF-12}_{\text{Physical Health Scale}}$) and mental health ($\text{SF-12}_{\text{Mental Health Scale}}$) [39]. Presence and severity of different impairments over the last 4 weeks are rated. Subscale scores can vary between 0-100, with higher scores indicating less impairment. Reliability is good with a Cronbach alpha of .76 [40] and test-retest correlations of r_{tt}=.76 for the physical component and r_{tt}=.89 for the mental component [39].

Questionnaire for the Evaluation of Psychotherapeutic Progress-2 (FEP-2)

The FEP-2 comprises 40 items and measures 4 dimensions of therapeutic progress and outcome (well-being, symptom distress, incongruence, and interpersonal problems) [41]. Answers are provided on a 5-point Likert scale (1-“never” and 5-“very often”) with higher scores indicating higher impairment. Reliability is high for the global scale (Cronbach alpha=.96; Retest between r_{tt}=.69-.77) and sensitivity to change has been demonstrated [41].

Attitudes Toward Psychological Online Interventions (APOI) Questionnaire

The attitudes toward psychological online interventions (APOI) [42] measures attitudes toward online-based interventions with 16 items. The following subscales are assessed: (1) Confidence in Effectiveness, (2) Skepticism and Perception of Risks, (3) Technologization Threat, and (4) Anonymity Benefits. Answers are provided on a 5-point Likert scale (1-“I disagree entirely” and 5-“I agree entirely”) and subscale scores range from 4-20. Higher values on the APOI total score indicate a more positive attitude toward psychological online interventions (POI). Reliability is good with a Cronbach alpha of .77.

Adherence

Adherence to the intervention was defined as the extent to which participants used the intervention. A number of modules were calculated by summing up all modules that were accessed for at least ten minutes. Usage time was defined as the number of hours participants spent using the Web intervention. At screening, at registration, every 2 weeks during the 12-week Web intervention period, and after the intervention, participants were asked to fill out the PHQ-9. The number of completed PHQ-9 assessments after week 4 of the intervention was used as an additional indicator of adherence.

Data Analytic Strategy

Patterns of early change in depressive symptoms, measured by the PHQ-9 over the first 4 weeks of the Web intervention, were identified using piecewise growth mixture modeling (PGMM) [43]. GMMs are considered a conservative method of identifying early change patterns in comparison with rational definitions such as reliable or clinical significant change [44]. Individual variance of intercepts (intake scores) and slopes (change) are captured in terms of a latent class variable that is added to the growth model [43], which allows the identification of subpopulations of participants with similar growth curves. In contrast to conventional growth models, which assume that there is only one underlying population with a single change pattern, GMMs allow the investigation of an a priori unknown number of latent subpopulations, which can differ with regard to intercepts and slopes (in the case of a linear model) as well as class specific variations around these parameters. In GMM, cases with a missing value in the PHQ-9 over the first 4 weeks were not excluded, but rather all available data was used to estimate growth curves within clusters.

In this study, we applied a PGMM, modeling the change pattern as 2 distinct phases (phase 1: time between screening and registration; phase 2: time between registration and assessment at week 4 of the intervention). Therefore, we used a model with 3 latent growth factors: an intercept indicating initial impairment and 2 slopes (one for each phase of change). To model change before the intervention (phase 1), the first slope loadings, which represent change in phase 1, were fixed to 0 at screening and to 1 at registration and later assessments. To model change during the first 4 weeks of the intervention (phase 2), the second slope loadings, which represent change in phase 2, were fixed to 0 at screening and registration and for the following 2 assessments, the log-linear transformation (base 10) of 2 and 3 were used respectively. According to the Bayesian information criterion (BIC; [45]), the log-linear transformation of factor loadings for the second slope improved the model fit compared with the linear transformation and was therefore used in subsequent analyses (linear: 6826.98, log-linear: 6820.20).

In order to model early response while taking potential spontaneous remission into account, we implemented one categorical latent class factor based on the 3 growth parameters (intercept, first slope, and second slope). Several fit criteria had been discussed to determine the optimal number of latent trajectory classes. In this study, we applied the BIC and the bootstrapped likelihood ratio test (BLRT) as proposed by Nylund et al [46] to determine the optimal number of latent trajectory classes. Thus, the model determination process was 2-fold. To identify the model with the lowest BIC value, the estimation procedure started with a 1-class solution and then one more class was added in each subsequent run. As mixture models are sensitive to class overextraction, in the second step an additional criterion (BLRT) was used to balance against this potential bias. Once the BIC value no longer decreased from a model with k classes to a model with k+1 classes, this solution was then tested against a solution with k-1 classes using the BLRT. If the BLRT revealed a significant *P* value (*P*<.05), the model was chosen as the best solution. If, however, the BLRT was not significant, the model was rejected and the solution with one class less (k−1) was tested against a model with two classes less (k−2). This procedure was repeated until the BLRT resulted in a significant *P* value.

In the final analysis, we fixed the variances around the class-specific slopes to zero in both phases, whereas intercept variances were freely estimated but constrained to be constant between classes. Therefore, heterogeneity in change had to be captured by the difference in mean slopes of different latent classes completely. Thus, in line with our main interest, we forced the estimation procedure to be more sensitive to patterns of change over time rather than to differences of initial levels of impairment. This approach can be seen as a hybrid of models in which all parameters’ variances are fixed to zero (latent class growth models) and in which the free estimation of all parameters is allowed (for similar approaches, also see [22,47]).

As the purpose of this study was to evaluate the impact of early change on overall treatment response, we examined the effect of early change patterns on change from pre- to post-treatment in terms of effect sizes as well as reliable change. To evaluate change on the PHQ-9, within-group effect sizes were calculated by subtracting the PHQ-9 score at post from PHQ-9 at screening and dividing the result by the SD of the PHQ-9 score at screening. As described previously, reliable change criteria were applied to the change scores to classify patients into 3 groups: reliably improved (pre to post improvement larger than the RCI of the PHQ-9, which equals 2.46 total points), reliably deteriorated (pre to post deterioration larger than the RCI), and not reliably changed (pre to post change remained under the RCI).

Subsequently, the identified latent change patterns were used to predict outcome and adherence, while controlling for initial impairment (PHQ-9 at screening, HRSD-24 at screening), patient characteristics (FEP-2, SF-$12_{\text{Physical Health Status}}$ and SF-$12_{\text{Mental Health Status}}$), and attitudes toward the online intervention ($\text{APOI}_{\text{total}}$) in stepwise regression analysis. Finally, we examined whether initial impairment, patient characteristics, and attitudes toward the online intervention predicted early change patterns using analysis of variances (ANOVAs) and multinomial regression analysis.

Results

Patterns of Early Change

Following the 2-old model determination process, the 3-class solution showed the best model fit, as suggested by the BIC and BLRT (see Table 1). As a result, the 3-class solution was used for further analyses.

Table 1. Information criteria, entropy, and *P* values in a bootstrapped likelihood ratio test for up to 4 latent classes in a 2-piece model.

# Classes	BIC[a]	SABIC[b]	AIC[c]	Entropy	BLRT[d] *P* value
1	6855.68	6830.30	6823.57		
2	6782.91	6744.84	6734.75	0.74	<.001
3	6773.41	6722.64	6709.19	0.65	<.001
4	6777.33	6713.87	6697.06	0.66	<.001

[a] BIC: Bayesian information criterion.

[b] SABIC: sample size adjusted BIC.

[c] AIC: Akaike information criterion.

[d] BLRT: bootstrapped likelihood ratio test.

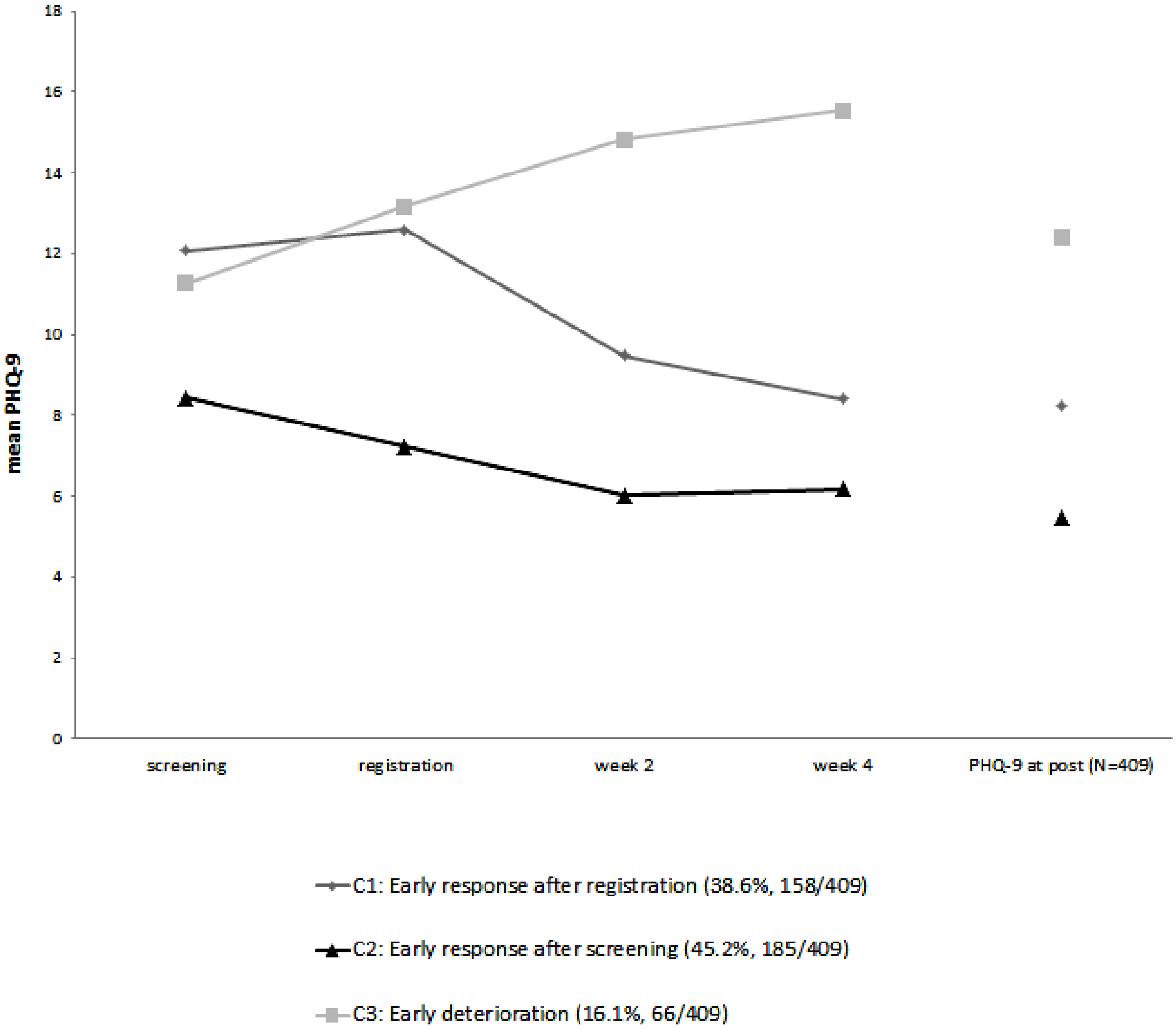
Graphical inspection revealed 2 early response groups and 1 early deterioration group. As shown in Figure 2, patients in classes 1 and 3 were characterized by higher PHQ-9 scores at screening that were above the cut-off score of 9 for clinical samples (C1: mean=12.08, C3: mean=11.27). Class 2 started treatment with lower (mild) depressive symptom severity (C2: mean=8.44).

The first subgroup labeled “early response after registration” (C1: 38.6%, 158/409) showed rapid early decrease in depressive symptom severity after registration. The early change effect size (between screening and week 4) in this group was *d*=1.35, reflecting rapid improvement. In the second subgroup which was labeled “early response after screening” (C2: 45.2%, 158/409), depressive symptoms decreased significantly not only during phase 2, but already during phase 1. The early response effect size within this latent class was large (*d*=0.98). In contrast to these 2 groups, a third subgroup of participants (C3: 16%, 66/409) showed a significant increase of depressive symptoms from screening to registration and from registration to assessment at week 4. This was the only class with a negative early change effect size (*d*=−1.78) and was therefore labeled “early deterioration.”

As some participants received additional treatment during the Web intervention and some were provided with email support, we compared these variables between classes to control for differential influences. The number of patients who were in therapy at the beginning of the Web intervention did not differ across classes (χ^2_2=4.4, *P*=.11), and there was no difference between classes in regard to reported change in additional treatment status at the end of the Web intervention (χ^2_{10}=10.6, *P*=.39). Also, at the beginning of treatment, there was no difference between classes with regard to the number of patients receiving medication (χ^2_2=1.4, *P*=.50). At the end of treatment, classes did not differ with regard to number of patients reporting change in medication (χ^2_2=0.9, *P*=.64) or use of medication ($\chi^2_{2=}$ 4.1, *P*=.13). Also there was no difference regarding use of psychotherapy (χ^2_2=0.2, *P*=.92). Only 9 patients reported being treated in outpatient clinics, and only 5 reported being treated in inpatient clinics, so no meaningful difference between classes could be established.

Furthermore, the number of patients receiving email support during the Web intervention differed significantly between classes (*P*<.001). Whereas almost all participants in C1 (96.2%, 152/158) and C3 (83%, 55/66) received email support, only 31% (57/185) of participants in C2 exceeded the cut-off of 10 on the PHQ-9 at screening and thus received email support. Therefore, email support was included as a predictor variable in the following analyses.

Figure 2. Mean latent growth curves for piecewise growth mixture modeling (PGMM) solution with 3 latent classes within the first 4 weeks and observed mean scores (Patient Health Questionnaire-9, PHQ-9) in the respective classes after the Web intervention.



Patterns of Early Change and Treatment Outcome

Table 2 shows the relative frequency of reliable improvement and pre-post effect sizes on the PHQ-9 depending on class membership. The relationship between reliable improvement of depressive symptoms and class membership was analyzed with a chi-square test, which revealed a significant association (χ^2_9 =74.8, P<.001).

As can be seen in Table 2, 62% (99/158) of participants in C1 (early response after registration) showed reliable change (standardized residual=1.5) and, on average, participants in this group showed the largest pre-post effect size (d=1.63). Only 3% (5/158) showed a reliable negative development at the end of treatment.

In C2 (early response after screening) the rate of reliable improvement (56%, 104/185) and the pre-post effect size (d=1.25) were slightly lower than in C1. The rate of reliable deterioration was 7% (12/185) for this class.

In C3 (early deterioration) only 27% (18/66; standardized residual=−3.00) of participants showed reliable improvement, yet in 39% of cases (26/66; standardized residual=7.2), participants in this class showed reliable deterioration. This was also the only class with a negative pre-post effect size (d=−0.47). All 3 classes had similar rates of lacking reliable change (C1: 34.2%, 54/158; C2: 37.2%, 69/185; C3: 33%, 22/66).

To identify potential relevant predictors beyond impairment at screening (PHQ-9, HRSD-24), we first correlated patient characteristics (FEP-2, SF-$12_{\text{Physical Health Status}}$ and SF-$12_{\text{Mental Health Status}}$) and attitudes toward online interventions ($\text{APOI}_{\text{total}}$) with PHQ-9 postscores. Only initial attitudes toward online interventions (r=−.12, P=.01) were significantly correlated with outcome.

Subsequently, the estimation of the additional predictive power of early change patterns beyond variables at screening was conducted via a stepwise regression analysis. Impairment at screening (PHQ-9) was added into the model first, followed by HRSD-24 and $\text{APOI}_{\text{total}}$. The dummy coded class membership variables were added to the model in the last stage of the analysis (see Table 3). The inclusion of the PHQ-9 score at screening explained 7.6% of the variance of PHQ-9 at post ($P<.001$). Participants with higher scores at the beginning of treatment tended to end with higher scores after treatment. HRSD-24 was also included and significantly increased the amount of explained variance by 3.4% ($P<.001$). Similarly, participants with higher impairment scores tended to end with higher scores. The addition of APOI$_{total}$ significantly increased the amount of explained variance by a further 1.5% ($P=.008$) resulting in a total of 12.5% of explained variance. A higher score at intake, indicating a more positive attitude toward the intervention, was significantly associated with lower PHQ-9 scores after treatment (see Table 3). Email support was not significant and therefore excluded in the next step ($t_{407}=-0.05$, $P=.96$).

Table 2. Relative frequencies of reliable improvement of the Patient Health Questionnaire-9 (PHQ-9), pre-post change on the PHQ-9 (effect sizes) and adherence by patient group of early change.

Sample		Outcome		Adherence		
	n	Reliable improvement on PHQ-9[a] n (%)	Pre-post ES[b] on PHQ-9 (*d*) (95% CI)	Usage time (in hours) mean (SD[c])	Number of modules of the Web intervention mean (SD)	Number of assessments[d] mean (SD)
All patients	409	221 (54)	1.12 (0.94-1.30)	7.89 (4.81)	9.10 (4.38)	2.52 (1.25)
Class 1[e]	158	99 (62)	1.63 (1.39-1.86)	8.36 (4.08)	9.84 (3.90)	2.80 (1.16)
Class 2[f]	185	104 (56)	1.25 (1.04-1.47)	7.32 (4.69)	8.64 (4.52)	2.37 (1.25)
Class 3[g]	66	18 (27)	−0.47 (−1.05 to 0.12)	8.33 (6.39)	8.65 (4.85)	2.27 (1.36)
P value		<.001[h]	<.001[i]	.10[i]	.03[i]	<.001[i]

[a] PHQ-9: Patient Health Questionnaire-9.

[b] ES: effect size.

[c] SD: standard deviation.

[d] Number of assessments: Number of PHQ-9s after week 4.

[e] Class 1: Early response after registration.

[f] Class 2: Early response after screening.

[g] Class 3: Early deterioration.

[h] χ^2 tests were performed, testing the association between class membership and categorized treatment outcome.

[i] 1-way analysis of variances (ANOVAs) were performed, testing the association between class membership and mean d for pre-post change, usage time, number of modules of the Web intervention, and number of assessments.

Adding the dummy coded variables for class membership resulted in a further increase of 21.5% explained variance of treatment outcome ($P \leq .001$). Thus, in total, 34% of variability of PHQ-9 change during the course of treatment was able to be explained by the model that contained initial PHQ-9, HRSD-24, and APOI$_{total}$ scores, as well as early change patterns (see Table 3).

Patterns of Early Change and Adherence

Adherence assessed via the number of modules and number of assessments (number of completed PHQ-9s) was significantly associated with class membership (see Table 2). A 1-way ANOVA revealed significant associations between mean number of modules of the Web intervention and class membership ($F_{2,406}=3.65$, $P=.03$). A post hoc test using Bonferroni correction showed that participants in C1 had accessed significantly more modules of the Web intervention than participants in C2 ($d=0.28$; pooled SDs between clusters were used to calculate between group effect-sizes). Concerning the number of assessments ($F_{3,405}=6.91$, $P=.001$), similar results were found using Bonferroni corrected *P* values: Participants in C1 filled out more assessments than participants in C2 ($d=0.36$) and C3 ($d=0.44$). There was no significant difference between the subgroups regarding usage time ($F_{2,406}=2.32$, $P=.10$).

For adherence measured by the number of modules of the Web intervention used and the number of assessments completed, the predictive power of early change patterns was examined using stepwise regression analysis (see Table 3). In the first step of the analysis, impairment at screening (PHQ-9) was included in the equation. It was significantly associated with number of assessments ($F_{1,406}=15.21$, $P<.001$) and explained 3.6% of variance. HRSD-24 at screening was excluded ($t_{407}=-0.44$, $P=.66$) and neither attitudes toward Web interventions ($t_{407}=-0.03$, $P=.98$) nor email support ($t_{407}=0.83$, $P=.41$) enhanced predictability. In the second step, class membership was entered in the model. The addition of class membership explained an additional 1.5% of variance of number of assessments ($P=.04$). Thus, a total of 5.1% of variability of number of assessments was able to be explained by the final model, which included PHQ-9 at screening and class membership.

Table 3. Stepwise multiple regression analyses predicting outcome and adherence by patient characteristics, email-support, and patterns of early change.

Steps	Predictors[a]	Outcome			Adherence					
		PHQ-9[b] at post			No. of assessments[c]			No. of modules in intervention		
		ΔR^2	Beta	*P*	ΔR^2	Beta	*P*	ΔR^2	Beta	*P*
Step 1		0.125		.008	0.036		<.001	0.024		.002
	PHQ-9		.216	<.001		.190	<.001		.155	.002
	HRSD-24[d]		.199	<.001		−0.043	.41		.051	.33
	APOI[e]		−0.124	.008		.005	.92		.082	.10
	$\text{Support}_{\text{Email}}$		−0.004	.96		.066	.45		.002	.98
Step 2		0.215			0.015		.04	0.006		.31
	PHQ-9		−0.040	.50		.192	.007		.146	.04
	HRSD-24		.045	.32		−0.024	.66		.068	.22
	APOI		−0.129	.002		.001	.98		.080	.11
	$\text{Support}_{\text{Email}}$		−0.068	.36		.073	.41		.003	.97
	C2[f]-dummy		−0.346	<.001		−0.030	.69		−0.033	.66
	C3[f]-dummy		.342	<.001		−0.132	.01		−0.082	.13
Total R^2		.34			.05			.03		
N		409			409			409		

[a] Predictors included Patient Health Questionnaire-9 (PHQ-9) at screening, 24-item Hamilton Rating Scale for Depression (HRSD-24) at screening, attitudes toward Web-based intervention, email support, and early change patterns.

[b] PHQ-9: Patient Health Questionnaire-9.

[c] No. of assessments = No. of PHQ-9 assessments after week 4.

[d] HRSD-24: 24-item Hamilton Rating Scare for Depression.

[e] APOI: attitudes toward psychological online interventions.

[f] C2 and C3 are dummy coded class membership variables with Class 1: early response after registration, used as reference class.

Prediction of Early Change Based on Patient Intake Characteristics

Next, we investigated the relationships between class membership, initial impairment, participants' intake characteristics and attitudes toward Web interventions via separate ANOVAs ($\text{APOI}_{\text{total}}$, $\text{SF-12}_{\text{Physical Health Scale}}$, $\text{SF-12}_{\text{Mental Health Scale}}$, and FEP-2). Using Bonferroni corrected *P* values, baseline scores on the FEP-2, $\text{SF-12}_{\text{Physical Health Scale}}$, and $\text{SF-12}_{\text{Mental Health Scale}}$ showed significant relationships with class membership.

With regard to the $\text{SF-12}_{\text{Physical Health Scale}}$, C3 participants showed significantly lower values than C1 (*d*=0.40) and C2 (*d*=0.43) participants, indicating a higher level of physical impairment in C3 participants. On the $\text{SF-12}_{\text{Mental Health Scale}}$, C2 participants reached significantly higher values than C1 (*d*=0.76) and C3 (*d*=0.49) participants, indicating that participants in C2 were less mentally impaired than participants in C1 and C3. Impairment measured by the FEP-2 differed significantly between C2 and C1 (*d*=0.42) as well as C2 between and C3 (*d*=0.44), with C2 showing the lowest values, indicating the lowest level of impairment.

When adding these significant variables and email support to multinomial logistic regressions, depressive symptoms measured by the PHQ-9 (χ^2_2=75.4; *P*<.001) and HRSD-24 (χ^2_2=34.8; *P*<.001) as well as physical health ($\text{SF-12}_{\text{Physical Health Scale}}$; χ^2_2=6.6; *P*=.04) demonstrated specific predictive power for class membership. Results of multinomial logistic regression analyses with patient characteristics as predictors of class membership are presented in Table 4.

PHQ-9 intake scores significantly discriminated between classes. Higher scores were associated with a lower probability of belonging to C3 or C2 compared with C1 and a higher probability of belonging to C3 compared with C2. In addition, HRSD-24 intake scores also discriminated between classes with higher scores associated with a lower probability of belonging to C2 compared with C1 and a higher probability of belonging to C3 compared with C2. Higher $\text{SF-12}_{\text{Physical Health Scale}}$ intake scores, indicating lower impairment, were associated with a lower probability of membership in C3 compared with C1.

Table 4. Prediction of class membership by patient intake characteristics via multinomial logistic regression analyses. R^2=.51 (Cox & Snell) and 0.588 (Nagelkerke). Model χ^2_8=284.3. For each comparison, the class mentioned first is used as the reference class in the multinomial logistic regression.

Variables			95% CI for odds ratio		
	regression coefficient B (standard error)	*P*	Lower	Odds ratio	Upper
Class 1[a] versus class 2[b]					
Intercept	14.84 (2.96)	<.001			
PHQ-9[c]	−1.31 (0.19)	<.001	0.18	0.27	0.39
HRSD-24[d]	−0.12 (0.03)	<.001	0.84	0.89	0.94
FEP-2[e]	0.19 (0.67)	.78	0.32	1.20	4.46
SF-12[f] $_{\text{Physical Health}}$	−0.01 (0.02)	.73	0.95	0.99	1.03
SF-12$_{\text{Mental Health}}$	0.02 (0.03)	.44	0.97	1.02	1.08
Email support	0.07 (0.65)	.91	0.30	1.08	3.83
Class 1 versus class 3[g]					
Intercept	3.65 (2.72)	.18			
PHQ-9	−0.26 (0.13)	.04	0.60	0.77	0.98
HRSD-24	0.04 (0.02)	.08	0.99	1.04	1.09
FEP-2	0.37 (0.60)	.54	0.45	1.45	4.68
SF-12$_{\text{Physical Health}}$	−0.05 (0.02)	.01	0.92	0.96	0.99
SF-12$_{\text{Mental Health}}$	0.01 (0.03)	.70	0.96	1.01	1.06
Email support	−0.77 (0.68)	.26	0.12	0.47	1.75
Class 2 versus class 3					
Intercept	−11.19 (3.12)	<.001			
PHQ-9	1.05 (0.20)	<.001	2.55	2.87	4.21
HRSD-24	0.16 (0.03)	<.001	1.07	1.18	1.25
FEP-2	0.19 (0.73)	.80	0.29	1.21	5.07
SF-12$_{\text{Physical Health}}$	−0.04 (0.02)	.07	0.92	0.96	1.00
SF-12$_{\text{Mental Health}}$	−0.01 (0.03)	.70	0.93	0.99	1.05
Email support	−0.84 (0.65)	.20	0.12	0.43	1.56

[a] Class 1= early response after registration.

[b] Class 2= early response after screening.

[c] PHQ-9: Patient Health Questionnaire.

[d] HRSD-24 = 24-item Hamilton Rating Scale for Depression.

[e] FEP-2: Questionnaire for the Evaluation of Psychotherapeutic Progress-2.

[f] SF-12: 12-item short form health survey.

[g] Class 3: early deterioration.

Discussion

Principal Findings

This study examined patterns of early change during the first 4 weeks of a 12-week CBT-oriented Web-based intervention for depression by applying a PGMM analysis. We were able to identify 3 early change patterns: The first was characterized by early improvement after screening, the second by early improvement after registration, and the third by early deterioration. Furthermore, latent classes differed with regard to outcome and adherence measured by the number of assessments (number of completed PHQ-9s) and number of modules used (for a duration of at least ten minutes), but not with regard to the overall amount of time spent using the system. Class membership improved outcome prediction by 21.5% over impairment at intake (PHQ-9 at screening, HRSD-24) and attitudes toward online interventions (APOI). In addition, initial impairment on the PHQ-9 and class membership significantly predicted the number of assessments. Furthermore, group membership of patients was significantly predicted by initial impairment on the PHQ-9 and HRSD-24 as well as by impairment on the SF-12 scale physical health.

The early response and deterioration patterns identified in this study of a CBT-oriented Web-based intervention have also been found in studies of individual face-to-face therapy [23]. In this study, a more differentiated investigation of early response was made possible by including the phase from screening to registration in the analysis. The identification of a subgroup that improved before treatment started may be indicative of a regression to the mean or “spontaneous remission” effect for some patients. Yet spontaneous remission may only explain part of the effect in this group. The decision to start treatment and the knowledge of being screened and accepted for the Web intervention may have already created a positive effect by inducing hope and positive treatment expectations, therefore leading to continuous positive changes in outcome, reaching to the end of treatment. Interestingly, participants in this class had significantly less email support than the other 2 classes. Clearly, more studies, which consider the pretreatment phase in the study of early change patterns, are necessary.

In contrast, early response after registration may correspond to a response pattern, which has recently been described as the pliant response pattern [48]. For this patient group, the impact of the specific treatment is essential: response to treatment is excellent, if the treatment provided is excellent and poor and if the treatment provided is poor. In line with the results of previous studies [24,23], participants with early positive change were likely to be improved (reliably) at the end of the treatment and, on average, showed a higher mean effect size than other participants.

In our study, the rate of participants showing early deterioration (16.1%, 66/409) was somewhat higher than in other studies (4.6%, [23]; 2.4%, [43]), yet more studies are required before conclusions regarding the risk of deterioration during Web interventions can be drawn. Participants who deteriorate early may be facing crisis and be in need of more immediate help than can be provided by an Web intervention. They may especially benefit from treatment selection and the combination of face-to-face and Web interventions [49]. In any case, email support was not lower in this group than in the early response after registration group.

One possible explanation of the mixed findings regarding the frequency of early negative response patterns could be the varying settings, with different early response rates in face-to-face, medication, and Web interventions. However, further studies must also investigate the influence of varying (outcome?) instruments and definitions of early response within this area of investigation [23].

Although email support did not predict outcome, initial PHQ-9 and HRSD-24 scores as well as attitudes toward Web interventions remained significant predictors of outcome after controlling for class membership. The finding of an association between attitudes toward Web interventions and outcome fits well with findings concerning the contribution of treatment expectation to treatment outcome [50]. In Web interventions, one of the first aims should therefore be to promote positive attitudes and motivation with regard to the intervention, making improvement more likely, while preventing dropout.

With regard to adherence, it could be shown that participants with early symptom deterioration completed fewer modules of the intervention and fewer assessments than the early response after registration group. However, class membership predicted number of assessments only. Participants who experience improvement may feel more inclined to track their progress and to make maximum use of the limited time available (12 weeks), whereas participants who show less early improvement may be discouraged from using the intervention more intensively. While early response is often associated with shorter treatment length [51], it has also already been reported that in time-limited treatment protocols, early response participants tend to complete the protocol and are less likely to drop out of treatment [23].

Somewhat surprisingly, there was no difference between early response groups with regard to usage time. A possible explanation could be that the Web intervention was tailored to patients, resulting in individual patients taking different paths within the Web intervention. These paths varied in length, presenting participants with critical problems with more content and longer paths, which took more time.

Physical health was associated with a higher probability of belonging to the early response after registration group compared with the early deterioration group, indicating that physical health may be an important factor not only in face-to-face treatment but also in Web interventions. In addition, Web interventions may be needed, which take poor physical health into account, for example, by providing psychoeducation and/or special coping strategies for patients with symptoms of pain. This could increase adherence by addressing a possibly important concern of some participants who may otherwise feel like the intervention is not adequately targeting their problems [52]. Furthermore, poor physical health may decrease motivation and increase negative expectations such as “nothing is going to change” or “I can’t do this” leading to dropout or lack of improvement. In this case, the initiation of motivation and hope may be especially crucial. Similar to face-to-face settings, early change patterns during Web interventions may have important implications for treatment selection, the continuation and adaptation of treatment, as well as the development of new Web or blended interventions. Early response monitoring may support the decision-making process with regard to the addition of special content (eg, coping with physical impairment and enhancing positive treatment expectations) or the necessity of higher intensity treatments. Furthermore, physical health and attitudes toward Web interventions may be important factors that influence early response or early deterioration and may be useful indicators when deciding whether a specific Web intervention should be applied. Although some interventions target multiple problems [53], it is still unclear whether such interventions can raise the early positive response rate. Also, given that the participants in the early response after screening and early response after registration groups showed improvement, it may be that varying factors contribute to early response.

Conclusions

Clearly, hope and positive expectations have an impact on early response; however, we don’t yet know much about specific personal characteristics such as self-efficacy. It would be interesting to investigate whether participants that improve or show an early positive change differ with regard to self-efficacy and whether high or low self-efficacy influences outcome in the long-term.

To summarize, more research is still necessary to understand which factors contribute to early response, which factors are indicate risk of early deterioration or dropout, and how clinicians or developers of Web interventions can best adapt interventions, particularly in routine care settings [53,54].

In summary, identifying patterns of early change can have implications for treatment outcome and treatment completion rates. Session-by-session monitoring and feedback of this information may increase awareness of these early change patterns and be applied as part of a stepped-care treatment approach [49].

Limitations

The following limitations of this study should be considered when interpreting the results. Unfortunately, number of modules of the Web intervention and usage time could not be assessed on a weekly basis, limiting what can be said about the progress of adherence in relation to the progress of symptoms. Also, due to economic considerations, the PHQ-9 was used as the sole outcome measure over the course of treatment. In future studies, a broader range of outcome measures (eg, an anxiety measure) and usage variables could be regularly monitored, improving the estimation and investigation of outcome and adherence. In future studies, additional predictors of adherence should also to be studied.

In addition, only participants with at least one assessment during the intervention were included in the analyses. We addressed this issue by testing for differences between included and excluded participants. Although we did not find any differences, these results should not be generalized to participants who, for whichever reason, did not complete any assessments during the first weeks of treatment.

In addition, it must be mentioned that the application of GMM and the associated selection of an optimal number of groups is not without disadvantages [55,56]. One disadvantage is the possibility of specification errors, which can result in the overextraction of trajectory classes through GMM [57]. For this reason, after examining 2 common fit indices (BIC and BLRT), we decided to take the 3-class model into account only. When interpreting early change patterns extracted using GMM, it should not be forgotten that the result is a simplification of a more complex reality, which warrants caution [55]. GMM remains just one possibility to identify early change patterns and other identification possibilities should be considered.

Despite these limitations, this study underlines the potential of early change patterns as predictors of treatment outcome as well as adherence, which, in the future, may guide treatment decisions regarding the content and continuation of Web interventions.

Acknowledgments

Funding was provided by German Federal Ministry of Health, II A 5 - 2512 FSB 052. The funding body played no role in study design, data collection, analysis, or interpretation of the data. Additional funding was provided by GAIA AG (Hamburg, Germany), which provided all participants in the trial with licenses for the Web intervention (Deprexis) as well as technical support. GAIA AG played no role in data collection, data analysis, or the decision to publish the results.

We would like to thank the EVIDENT study team: Berlin: Matthias Rose (core team member), Sandra Nolte, Anna Paulitschek, Leonie Gmöhling, and Leonie Schickedanz; Bern: Thomas Berger (core team member); Bielefeld: Wolfgang Greiner (core team member) and Viola Gräfe; Hamburg: Björn Meyer (core team member), Steffen Moritz (core team member), Johanna Schröder (core team member), Mirja Behrens, Cecile Hoermann, Anna J. Katharina Jahns, Thies Lüdtke, and Eik Vettorazzi; Lübeck: Fritz Hohagen (principal investigator), Philipp Klein (core team member), Christina Späth (core team member), and Antje Roniger; Trier: Wolfgang Lutz (core team member), Alice Arndt, Julian Rubel, Liv Glindemann, David Rosenbaum, and Kathinka Wolter; and Tübingen: Martin Hautzinger (core team member), Flora Bach, Elisabeth Beck, Kristina Fuhr, Katharina Krisch, and Melanie Wahl. External advisors: Franz Caspar, Bern; Bernd Löwe, Hamburg; and Gerhard Andersson, Linköping.

Conflicts of Interest

Jan Philipp Klein received payment for presentations, workshops, and books on psychotherapy for chronic depression and psychiatric emergencies. Björn Meyer is employed as research director at GAIA AG, the company that developed, owns, and operates the Web intervention investigated in this trial. All other authors and members of the EVIDENT study team report no financial or other relationship relevant to this article.

References

1. Barak A, Grohol JM. Current and future trends in internet-supported mental health interventions. J Technol Hum Serv 2011 Jul;29(3):155-196. [doi: 10.1080/15228835.2011.616939]
2. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-870. [doi: 10.1586/ern.12.63]

3. Moritz S, Schröder J, Meyer B, Hauschildt M. The more it is needed, the less it is wanted: attitudes toward face-to-face intervention among depressed patients undergoing online treatment. Depress Anxiety 2013 Feb;30(2):157-167. [doi: 10.1002/da.21988] [Medline: 22930656]
4. Ryan ML, Shochet IM, Stallman HM. Universal online interventions might engage psychologically distressed university students who are unlikely to seek formal help. Adv Mental Health 2014 Dec 17;9(1):73-83. [doi: 10.5172/jamh.9.1.73]
5. Titov N. Internet-delivered psychotherapy for depression in adults. Curr Opin Psychiatry 2011 Jan;24(1):18-23. [doi: 10.1097/YCO.0b013e32833ed18f] [Medline: 20827199]
6. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
7. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014 May;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
8. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015;351:h5627 [FREE Full text] [Medline: 26559241]
9. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
10. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
11. Andersson G, Hedman E. Effectiveness of guided internet-based cognitive behavior therapy in regular clinical settings. Verhaltenstherapie 2013;23(3):140-148. [doi: 10.1159/000354779]
12. Renton T, Tang H, Ennis N, Cusimano MD, Bhalerao S, Schweizer TA, et al. Web-based intervention programs for depression: a scoping review and evaluation. J Med Internet Res 2014 Sep 23;16(9):e209 [FREE Full text] [doi: 10.2196/jmir.3147] [Medline: 25249003]
13. Leykin Y, Muñoz RF, Contreras O, Latham MD. Results from a trial of an unsupported internet intervention for depressive symptoms. Internet Interv 2014 Oct 1;1(4):175-181 [FREE Full text] [doi: 10.1016/j.invent.2014.09.002] [Medline: 25485233]
14. Richards D, Timulak L, Doherty G, Sharry J, Colla A, Joyce C, et al. Internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression: protocol for a randomized controlled trial. BMC Psychiatry 2014 May 21;14:147 [FREE Full text] [doi: 10.1186/1471-244X-14-147] [Medline: 24886179]
15. Laurenceau J, Hayes AM, Feldman GC. Some methodological and statistical issues in the study of change processes in psychotherapy. Clin Psychol Rev 2007 Jul;27(6):682-695 [FREE Full text] [doi: 10.1016/j.cpr.2007.01.007] [Medline: 17328996]
16. Cuijpers P, van Lier PA, van SA, Donker M. Examining differential effects of psychological treatment of depressive disorder: an application of trajectory analyses. J Affect Disord 2005 Dec;89(1-3):137-146. [doi: 10.1016/j.jad.2005.09.001] [Medline: 16274750]
17. Lutz W, Stulz N, Köck K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. J Affect Disord 2009 Nov;118(1-3):60-68. [doi: 10.1016/j.jad.2009.01.019] [Medline: 19217669]
18. Bradford A, Cully J, Rhoades H, Kunik M, Kraus-Schuman C, Wilson N, et al. Early response to psychotherapy and long-term change in worry symptoms in older adults with generalized anxiety disorder. Am J Geriatr Psychiatry 2011 Apr;19(4):347-356 [FREE Full text] [doi: 10.1097/JGP.0b013e3181f18061] [Medline: 21427643]
19. Lewis CC, Simons AD, Kim HK. The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. J Consult Clin Psychol 2012 Aug;80(4):525-534. [doi: 10.1037/a0029131] [Medline: 22730951]
20. Crits-Christoph P, Connolly MB, Gallop R, Barber JP, Tu X, Gladis M, et al. Early improvement during manual-guided cognitive and dynamic psychotherapies predicts 16-week remission status. J Psychother Pract Res 2001;10(3):145-154. [Medline: 11402077]
21. Gunlicks-Stoessel M, Mufson L. Early patterns of symptom change signal remission with interpersonal psychotherapy for depressed adolescents. Depress Anxiety 2011 Jun 30;28(7):525-531. [doi: 10.1002/da.20849]
22. Hunter AM, Muthén BO, Cook IA, Leuchter AF. Antidepressant response trajectories and quantitative electroencephalography (QEEG) biomarkers in major depressive disorder. J Psychiatr Res 2010 Jan;44(2):90-98. [doi: 10.1016/j.jpsychires.2009.06.006]
23. Lutz W, Hofmann SG, Rubel J, Boswell JF, Shear MK, Gorman JM, et al. Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. J Consult Clin Psychol 2014;82(2):287-297. [doi: 10.1037/a0035535]
24. Delgadillo J, McMillan D, Lucock M, Leach C, Ali S, Gilbody S. Early changes, attrition, and dose-response in low intensity psychological interventions. Br J Clin Psychol 2013 Sep 30;53(1):114-130. [doi: 10.1111/bjc.12031]

25. Krause MS, Howard KI, Lutz W. Exploring individual change. J Consult Clin Psychol 1998;66(5):838-845. [doi: 10.1037/0022-006X.66.5.838]
26. Lutz W. Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. Psychother Res 2002 Sep;12(3):251-272. [doi: 10.1080/713664389]
27. National Institute for Health and Clinical Excellence. NICE. 2011. Common mental health disorders: identification and pathways to care URL: https://www.nice.org.uk/guidance/cg123 [accessed 2017-05-22] [WebCite Cache ID 6qeO5Prfl]
28. van SA, Hill J, Richards DA, Cuijpers P. Stepped care treatment delivery for depression: a systematic review and meta-analysis. Psychol Med 2015 Jan;45(2):231-246. [doi: 10.1017/S0033291714000701] [Medline: 25065653]
29. Nordberg SS, Castonguay LG, Fisher AJ, Boswell JF, Kraus D. Validating the rapid responder construct within a practice research network. J Clin Psychol 2014 Sep;70(9):886-903. [doi: 10.1002/jclp.22077] [Medline: 24652786]
30. Schibbye P, Ghaderi A, Ljótsson B, Hedman E, Lindefors N, Rück C, et al. Using early change to predict outcome in cognitive behaviour therapy: exploring timeframe, calculation method, and differences of disorder-specific versus general measures. PLoS One 2014 Jun;9(6):e100614 [FREE Full text] [doi: 10.1371/journal.pone.0100614] [Medline: 24959666]
31. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
32. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. Internet Interventions 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
33. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. The EVIDENT-trial: protocol and rationale of a multicenter randomized controlled trial testing the effectiveness of an online-based psychological intervention. BMC Psychiatry 2013 Sep 28;13:239 [FREE Full text] [doi: 10.1186/1471-244X-13-239] [Medline: 24074299]
34. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. Effects of a psychological internet intervention in the treatment of mild to moderate depressive symptoms: results of the EVIDENT Study, a randomized controlled trial. Psychother Psychosom 2016;85(4):218-228. [doi: 10.1159/000445355] [Medline: 27230863]
35. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
36. Rossi A, Alberio R, Porta A, Sandri M, Tansella M, Amaddeo F. The reliability of the Mini-International neuropsychiatric interview-Italian version. J Clin Psychopharmacol 2004;24(5):561-563. [doi: 10.1097/01.jcp.0000139758.03834.ad]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
38. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12]
39. Burdine JN, Felix MR, Abel AL, Wiltraut CJ, Musselman YJ. The SF-12 as a population health measure: an exploratory examination of potential for application. Health Serv Res 2000 Oct;35(4):885-904 [FREE Full text] [Medline: 11055454]
40. King JT, Horowitz MB, Kassam AB, Yonas H, Roberts MS. The short form-12 and the measurement of health status in patients with cerebral aneurysms: performance, validity, and reliability. J Neurosurg 2005 Mar;102(3):489-494. [doi: 10.3171/jns.2005.102.3.0489] [Medline: 15796384]
41. Lutz W, Schürch E, Stulz N, Böhnke JR, Schöttke H, Rogner J, et al. Entwicklung und psychometrische Kennwerte des Fragebogens zur Evaluation von Psychotherapieverläufen (FEP). Diagnostica 2009 Apr;55(2):106-116. [doi: 10.1026/0012-1924.55.2.106]
42. Schröder J, Sautier L, Kriston L, Berger T, Meyer B, Späth C, et al. Development of a questionnaire measuring attitudes towards psychological online interventions-the APOI. J Affect Disord 2015 Nov 15;187:136-141. [doi: 10.1016/j.jad.2015.08.044] [Medline: 26331687]
43. Rubel J, Lutz W, Kopta SM, Köck K, Minami T, Zimmermann D, et al. Defining early positive response to psychotherapy: an empirical comparison between clinically significant change criteria and growth mixture modeling. Psychol Assess 2015 Jun;27(2):478-488. [doi: 10.1037/pas0000060] [Medline: 25496087]
44. Wang M, Bodner TE. Growth mixture modeling. Organizational Research Methods 2007 Oct;10(4):635-656. [doi: 10.1177/1094428106289397]
45. Schwarz G. Estimating the dimension of a model. Ann. Statist 1978 Mar;6(2):461-464. [doi: 10.1214/aos/1176344136]
46. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Modeling 2007 Oct 23;14(4):535-569. [doi: 10.1080/10705510701575396]
47. Uher R, Muthén B, Souery D, Mors O, Jaracz J, Placentino A, et al. Trajectories of change in depression severity during treatment with antidepressants. Psychol Med 2010 Aug;40(8):1367-1377. [doi: 10.1017/S0033291709991528] [Medline: 19863842]

48. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The personalized advantage index: translating research on prediction into individualized treatment recommendations. A demonstration. PLoS One 2014;9(1):e83875 [FREE Full text] [doi: 10.1371/journal.pone.0083875] [Medline: 24416178]
49. Lutz W, De JK, Rubel J. Patient-focused and feedback research in psychotherapy: where are we and where do we want to go? Psychother Res 2015;25(6):625-632. [doi: 10.1080/10503307.2015.1079661] [Medline: 26376225]
50. Greenberg RP, Constantino MJ, Bruce N. Are patient expectations still relevant for psychotherapy process and outcome? Clin Psychol Rev 2006 Oct;26(6):657-678. [doi: 10.1016/j.cpr.2005.03.002] [Medline: 15908088]
51. Stulz N, Lutz W, Leach C, Lucock M, Barkham M. Shapes of early change in psychotherapy under routine outpatient conditions. J Consult Clin Psychol 2007 Dec;75(6):864-874. [doi: 10.1037/0022-006X.75.6.864] [Medline: 18085904]
52. Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P. Negative effects of internet interventions: a qualitative content analysis of patients' experiences with treatments delivered online. Cogn Behav Ther 2015;44(3):223-236. [doi: 10.1080/16506073.2015.1008033] [Medline: 25705924]
53. Deady M, Mills KL, Teesson M, Kay-Lambkin F. An online intervention for co-occurring depression and problematic alcohol use in young people: primary outcomes from a randomized controlled trial. J Med Internet Res 2016 Mar 23;18(3):e71 [FREE Full text] [doi: 10.2196/jmir.5178] [Medline: 27009465]
54. Drozd F, Vaskinn L, Bergsund HB, Haga SM, Slinning K, Bjørkli CA. The implementation of internet interventions for depression: a scoping review. J Med Internet Res 2016 Sep 08;18(9):e236 [FREE Full text] [doi: 10.2196/jmir.5670] [Medline: 27608548]
55. Bauer DJ. Evaluating individual differences in psychological processes. Curr Dir Psychol Sci 2011 Apr;20(2):115-118. [doi: 10.1177/0963721411402670]
56. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 2010 Mar;6:109-138. [doi: 10.1146/annurev.clinpsy.121208.131413] [Medline: 20192788]
57. Bauer DJ. Observations on the use of growth mixture models in psychological research. Multivariate Behav Res 2007 Dec 28;42(4):757-786. [doi: 10.1080/00273170701710338]

Abbreviations

AIC: Akaike information criterion
ANOVA: analysis of variances
APOI: attitudes toward psychological online interventions
BIC: Bayesian information criterion
BLRT: bootstrapped likelihood ratio test
C1: Class 1: Early Response after registration
C2: Class 2: Early Response after screening
C3: Class 3: Early deterioration
CBT: cognitive behavioral therapy
CG: control group
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
ES: effect size
FEP-2: Questionnaire for the Evaluation of Psychotherapeutic Progress-2
GMM: growth mixture modeling
HDRS-24: Hamilton Depression Rating Scale
IAPT: Improving Access to Psychological Therapies
ICD-10: International Classification of Diseases, Tenth Edition
IG: intervention group
M.I.N.I.: Mini International Neuropsychiatric Interview
PHQ-9: Patient Health Questionnaire-9
PGMM: piecewise growth mixture modeling
RC: reliable change
RCI: reliable change index
SABIC: sample size adjusted BIC.
SD: standard deviation
SF-12: Short-Form Health Survey-12

Edited by G Eysenbach; submitted 20.01.17; peer-reviewed by J Delgadillo, T Kaiser, K Schladweiler; comments to author 02.03.17; revised version received 07.04.17; accepted 19.04.17; published 09.06.17.

Please cite as:
Lutz W, Arndt A, Rubel J, Berger T, Schröder J, Späth C, Meyer B, Greiner W, Gräfe V, Hautzinger M, Fuhr K, Rose M, Nolte S, Löwe B, Hohagen F, Klein JP, Moritz S
Defining and Predicting Patterns of Early Response in a Web-Based Intervention for Depression
J Med Internet Res 2017;19(6):e206
URL: http://www.jmir.org/2017/6/e206/
doi:10.2196/jmir.7367
PMID:28600278

Original Paper

Using Feedback to Promote Physical Activity: The Role of the Feedback Sign

Jan-Niklas Kramer[1], MSc; Tobias Kowatsch[1], PhD

[1] Center for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St.Gallen, Switzerland

Corresponding Author:
Tobias Kowatsch, PhD
Center for Digital Health Interventions
Institute of Technology Management
University of St. Gallen
Dufourstrasse 40a
St.Gallen,
Switzerland
Phone: 41 0 71 224 ext 7244
Fax: 41 0 71 224 7301
Email: tobias.kowatsch@unisg.ch

Abstract

Background: Providing feedback is a technique to promote health behavior that is emphasized by behavior change theories. However, these theories make contradicting predictions regarding the effect of the feedback sign—that is, whether the feedback signals success or failure. Thus, it is unclear whether positive or negative feedback leads to more favorable behavior change in a health behavior intervention.

Objective: The aim of this study was to examine the effect of the feedback sign in a health behavior change intervention.

Methods: Data from participants (N=1623) of a 6-month physical activity intervention was used. Participants received a feedback email at the beginning of each month. Feedback was either positive or negative depending on the participants' physical activity in the previous month. In an exploratory analysis, change in monthly step count averages was used to evaluate the feedback effect.

Results: The feedback sign did not predict the change in monthly step count averages over the course of the intervention ($b=-84.28$, $P=.28$). Descriptive differences between positive and negative feedback can be explained by regression to the mean.

Conclusions: The feedback sign might not influence the effect of monthly feedback emails sent out to participants of a large-scale physical activity intervention. However, randomized studies are needed to further support this conclusion. Limitations as well as opportunities for future research are discussed.

(J Med Internet Res 2017;19(6):e192) doi:10.2196/jmir.7012

KEYWORDS

feedback; internet; physical activity; health behavior; activity trackers

Introduction

Background

In 2012, noncommunicable diseases (NCDs) such as diabetes, cardiovascular diseases, chronic respiratory diseases, or cancer were responsible for 68% of deaths worldwide [1].

Physical activity plays a crucial role in the prevention and management of NCDs, as it has been found to affect the incidence and course of NCDs such as diabetes [2], asthma [3], and cancer [4], as well as associated risk factors such as hypertension [5], overweight [6], or high blood sugar [7]. To reduce the burden of NCDs on the worlds' health and health care systems, researchers have focused on the development of effective physical activity interventions among others.

Physical activity interventions often use feedback as a method to change behavior [8-11]. For example, in a review of technology-enabled health interventions [11], feedback was employed in 55 out of approximately 110 (50.0%) reviewed interventions targeting physical activity and was identified as the second-most used behavior change technique. Abraham and Michie [12] defined feedback as "providing data about recorded behavior or evaluating performance in relation to a set standard or others' performance" (p.382). Despite its widespread use in physical activity interventions however, detailed analysis of feedback and its characteristics has so far been limited to behavioral domains other than health behavior, such as learning [13], professional care practice [14], or employee performance [15]. Research in these areas has indicated positive, yet highly variable effects of feedback on behavior [13-18]. For example, in their meta-analysis, Kluger and DeNisi [18] found a general positive effect of feedback on performance of mostly cognitive and motor tasks (d=.41), but in over one-third of all considered studies, feedback decreased performance. In a Cochrane review examining the effect of feedback on the compliance of health care professionals with desired practice, Jamtvedt et al [17] found a median increase in compliance of 5%. However, results varied from −16% to 70%. These results suggest the existence of further variables that may mediate or moderate the effect of feedback on behavior.

Behavioral theories provide a detailed specification of causal processes that lead to behavior change and can thus help to understand how feedback affects behavior [19]. Two different theories, control theory (CT) [20] and social-cognitive theory (SCT) [21], advocate the use of feedback, but both define different underlying processes. A better understanding of these processes can facilitate the design of feedback in physical activity (and other behavioral) interventions and can help to explain the variability of results of past research.

Feedback According to Control Theory

CT provides a model of self-regulation for intentional (or goal-directed) behavior (eg, walking 10,000 steps a day). Self-regulation is vital for physical activity promotion as it constitutes the basis for self-directed change [21], and physical activity interventions based on CT strategies have been found to be more effective than other physical activity interventions [22]. According to CT, people regulate their behavior by periodically comparing the perceived qualities of their own behavior with a salient reference value (eg, a goal) [23]. Whenever a discrepancy between one's performance and a goal is recognized, a behavior is triggered in order to reduce the discrepancy (negative feedback loop). Thus, feedback affects behavior via the comparison with a set goal or standard, which has important implications for feedback design: performance is likely to be increased after negative feedback, as negative feedback informs the recipient that a goal or standard has not been met. In turn, performance is likely to be maintained or decreased after positive feedback, as positive feedback signals that the recipient exceeded the goal, which can thus be reached with less effort as well. Figure 1 illustrates the underlying causal processes of the effect of feedback according to CT.

Feedback According to Social Cognitive Theory

In contrast to CT, SCT assumes that the mere perception of behavior and standards is insufficient to regulate behavior. It rather posits that cognitions such as self-efficacy beliefs are central factors that impact goal pursuit and self-regulation [24]. Self-efficacy beliefs are personal beliefs about the capability to exercise control over one's actions and constitute the foundation of human motivation and action [24]. More precisely, unless one believes that a desired health outcome (eg, walking 10,000 steps a day) can be produced by one's own actions, he or she has no incentive to act in the first place. Furthermore, self-efficacy beliefs influence the regulation of behaviors in a variety of ways. According to SCT, self-efficacy directly influences how goals are set and how strong one feels committed to achieving them, what outcomes one expects to realize, and how impediments are viewed [24]. It further affects how failures are attributed and how one reacts to setbacks [25]. People holding strong self-efficacy beliefs tend to be more motivated to reach set goals as they are convinced of their own capabilities, expect to realize favorable outcomes with their actions, and view impediments as surmountable.

Understandably, a major source of self-efficacy includes personal experiences of success and failure [26]. Since feedback usually contains an evaluation of one's performance with regard to a set goal, feedback conveys success if it is positive, that is, the recipient has achieved or exceeded the goal. In turn, feedback conveys failure if it is negative, that is, the recipient has not met the goal. Following the reasoning of SCT, performance is likely to increase after positive feedback because self-efficacy (and in turn performance) is increased. Consequently performance is decreased after negative feedback, which can undermine one's confidence in the ability to reach a set standard. Figure 2 illustrates the underlying causal processes of the effect of feedback according to SCT.

Figure 1. Effect of feedback according to control theory.

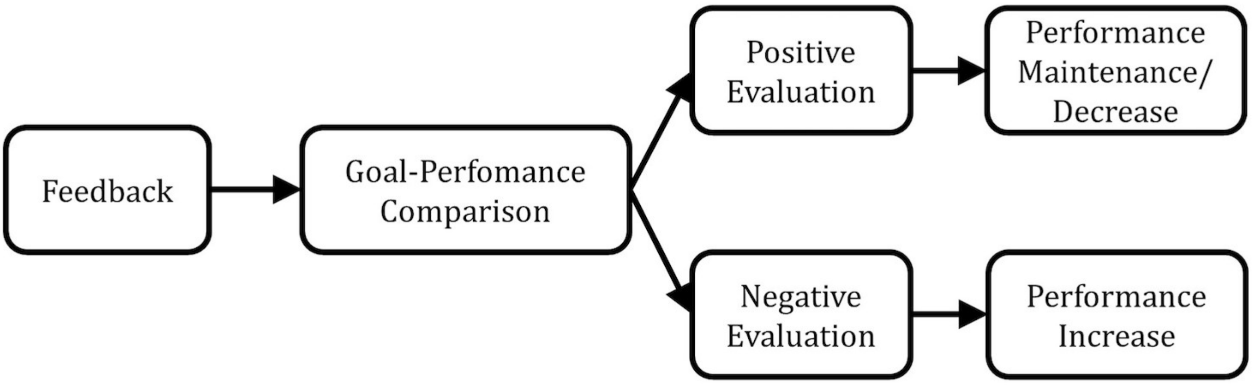
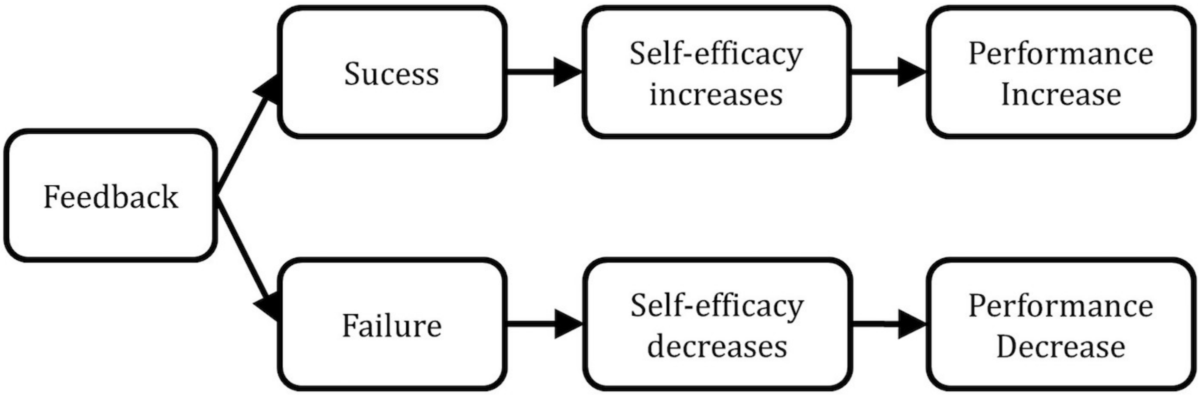


Figure 2. Effect of feedback according to social cognitive theory.



Research Question

Both CT and SCT contradict each other in their implications for feedback design. Whereas CT predicts that negative evaluation of performance leads to favorable behavior change, SCT predicts the same for positive evaluation of performance. Whether the feedback contains a positive or a negative evaluation of performance is often referred to as the sign of the feedback message [18]. Past research has so far produced inconclusive results regarding the effect of the feedback sign. In their meta-analysis, Kluger and DeNisi found no significant effect of the feedback sign, thus favoring neither of the two theoretical positions [18]. Both authors suggested later that the effect might be moderated by the regulatory focus of participants [27,28]. A more recent meta-analysis looking at the effect of feedback [29] could not evaluate the effect of the feedback sign, since too few studies directly compared positive and negative feedback. Consequently, it is unclear whether positive or negative feedback is more likely to change behavior when incorporated in physical activity interventions. Summarizing the line of reasoning above, we therefore pose the following research question: Does positive or negative feedback lead to more favorable behavior change in a physical activity intervention?

Methods

Procedure

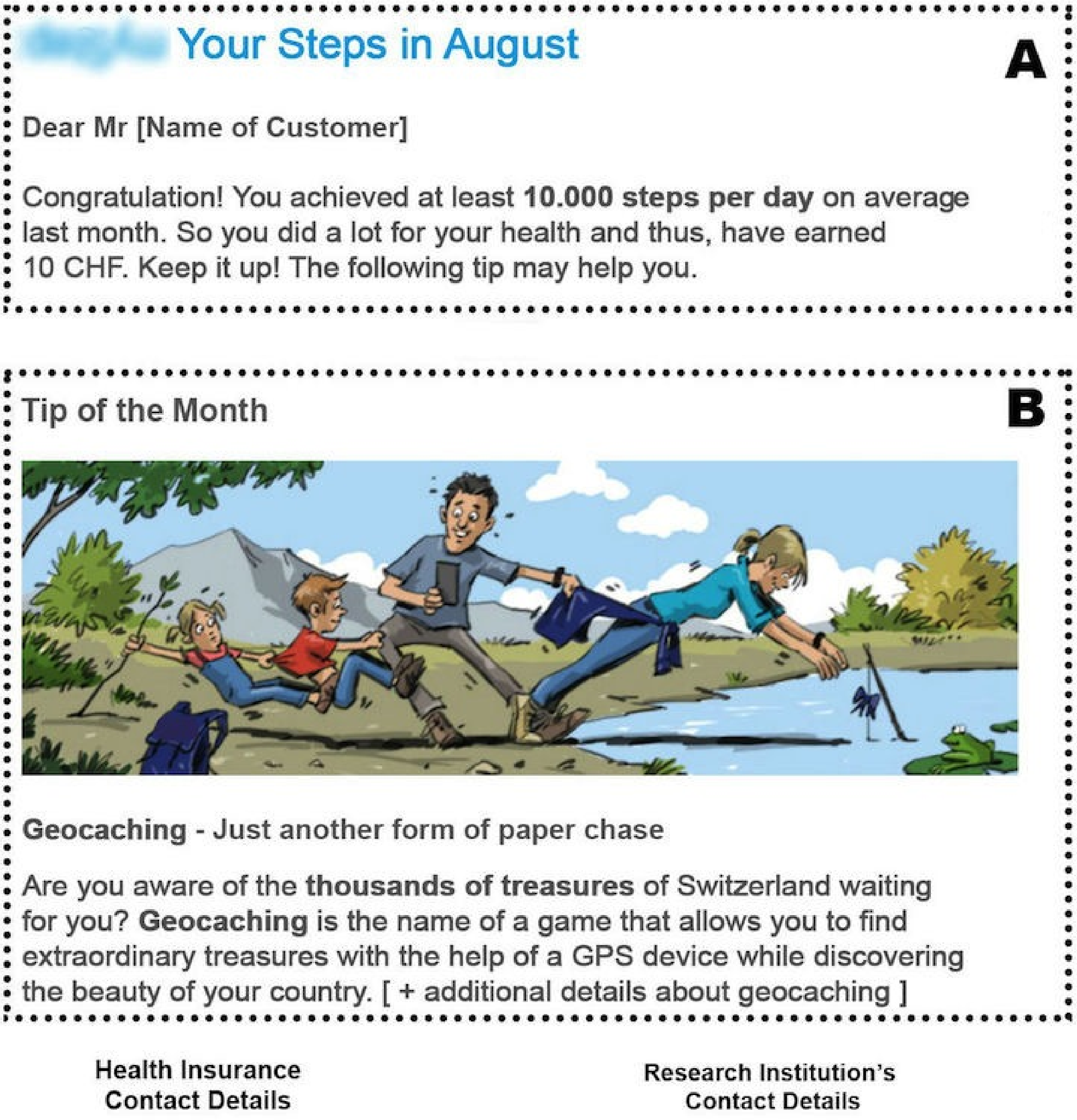
In order examine the effects of positive and negative feedback, we exploratory analyzed data from a cluster-randomized trial that primarily focused on the effects of different incentives on the acceptance of a digital physical activity intervention [30]. This paper specifically examines the effect of monthly feedback emails that were sent out over the course of this intervention. The study was approved by the ethics committee of the authors' university.

A total of 26,773 customers of a large Swiss health insurance company were invited through email, along with eligible family members, to participate in a physical activity intervention that was conducted from July 2015 to December 2015. In order to participate, customers had to be at least 18 years old, be registered in a complementary insurance program, accept participation conditions and privacy terms, and declare to be free of any medical condition that does not permit increased physical activity.

Before invitation, potential participants were clustered based on their state of residence and clusters were then randomly allocated to one out of three incentive conditions: In the financial incentive condition, participants received CHF10 (US $10) for each month they walked >10,000 steps a day on average. To prevent frustration, participants received CHF5 when their monthly step count average was below 10,000 but over 7500 steps, which matches the approximate minimum recommendation for daily physical activity [31]. The charitable incentive condition was equal to the financial incentive condition, except that participants could choose to donate an amount of choice of their earned money to a charitable organization. The control group received no incentives. To ensure fair treatment of all participants, the control group was entitled to financial incentives of CHF20 (for monthly averages above 10,000 steps) and CHF10 (for monthly averages below 10,000 and above 7500 steps) for the second half of the intervention. Participants completed a Web-based questionnaire at the beginning (T_1) and at the end of the intervention (T_2) to measure demographic and control variables relevant for the primary purpose of the study. A detailed description of the intervention as well as of all variables measured is available in Kowatsch et al [30].

In line with recommendations for health promotion, participants were advised to perform at least 150 min of moderate-intensity activity a week, which on average translates to a goal of 10,000 steps a day [31]. Via the Web-based customer account, participants could at any time gain insight into their physical activity data as well as their degree of achievement with regard to their goal of 10,000 steps a day.

Figure 3. Exemplary feedback email (authors' translation). A: Feedback message with positive or negative feedback depending on the performance of the participant. B: Season-based tip on how to increase physical activity (here: recommendation to participate in a geocaching activity).



Your Steps in August

A

Dear Mr [Name of Customer]

Congratulation! You achieved at least **10.000 steps per day** on average last month. So you did a lot for your health and thus, have earned 10 CHF. Keep it up! The following tip may help you.

Tip of the Month

B

Geocaching - Just another form of paper chase

Are you aware of the **thousands of treasures** of Switzerland waiting for you? **Geocaching** is the name of a game that allows you to find extraordinary treasures with the help of a GPS device while discovering the beauty of your country. [+ additional details about geocaching]

Health Insurance
Contact Details

Research Institution's
Contact Details

Feedback Emails

Starting after the first month, every participant received a feedback message by email at the beginning of the month that contained information on goal achievement of the last month. Consequently, every participant received 5 feedbacks over the course of the intervention. If participants failed to reach an average step count of at least 7500 steps a day, a negative feedback was provided (eg, "Unfortunately you did not reach the goal of 7500 steps a day on average last month"). In all other cases the feedback was positive (eg, "Well done, you have achieved at least 7500 steps a day on average over the last month and did a lot for your health"). In the financial and charitable incentive condition, feedback emails also contained information about the amount of money earned in the past month. Moreover, and in line with theory [18], all feedback emails contained a season-based tip on how to increase physical activity over the next month (eg, winter specific activities like snow shoe hiking or visiting Christmas markets were recommended in winter months, and outdoor activities like hiking or geocaching were recommended in summer and autumn months). Feedback emails were sent out by the insurance company. Figure 3 provides an example of a feedback email that was sent out in September.

Analyses

Descriptive Statistics

Data from the baseline questionnaire was used to describe the sample of this study. We calculated means and standard deviations (SDs) for continuous variables and absolute and relative frequencies for categorical variables. Data on age and gender of the participants was provided by the insurance company. Statistics on monthly average step counts were obtained by calculating the mean of all participants' mean step counts for each month.

Effect of the Feedback Sign

Because feedback on participants' physical activity referred to monthly average step counts, we used the change in monthly step averages as the outcome variable to compare the effect of positive and negative feedback. Specifically, we calculated the difference between monthly step count averages before and after dispatch of the feedback email. Since each participant received 5 feedback emails, we consequently obtained 5 difference measures per participant. This difference indicates whether a participant increased or decreased his or her average monthly step count in the month after receiving feedback. Differences with an absolute value of more than 10,000 steps are likely to be the result of irregular recorded step counts (eg, very few and very low recorded step counts in 1 month) and were regarded as outliers and excluded from analysis. Exclusion of outliers resulted in the removal of 26/8115 (0.03%) observations and did not affect the results of the analyses.

To determine what analyses should best be used to examine the effect of the feedback sign, a 2-level hierarchical linear model with measurements as level-1 unit of analysis and participants as level-2 unit of analysis was fit to the data. However, the comparison of an intercept-only and a random-intercept model revealed that modeling the nested data structure did not significantly improve the model fit (χ^2_1=0.0, P=.99). Hence, measurements were treated as independent, and a linear regression model was fit to the data in order to examine the effect of the feedback sign. Because the feedback emails informed participants in the incentive groups also about the reception of a financial incentive, the main effect of the experimental group was included in the model. To further account for confounding effects of seasonality and regression to the mean, the model was adjusted for the effects of time and participants' baseline physical activity [32]. In line with Barnett et al [32], baseline physical activity was defined as the average step count of the month preceding the dispatch of the feedback mail.

Results

Descriptive Statistics

In total, 1410 directly invited customers and 213 family members participated in the program resulting in a sample of N=1623 participants. On average, participants were 42.66 (SD=13.06) years old and slightly more men (848/1623, 52.25%) than women (770/1623, 47.44%) participated in the prevention program. Five participants (0.31%, 5/1623) did not disclose their gender. Compared with the Swiss population [33], people aged 20-39 years were overrepresented in the sample (46.40% [753/1623] vs 33.45% [2,225,129/6,651,623]), whereas people aged 64 years and above were underrepresented (7.64% [124/1623] vs 22.48% [1,495,052/6,651,623]). Moreover, program participants who filled out the Web-based questionnaire at T_1 (n=1220) tend to have higher net income (between CHF 5000 and 10,000: 61.4% [593/966] vs 33.5%, more than CHF 10,000: 13.2% [127/966] vs 3.5%; no absolute values were available for the different income categories) and a lower chance of not obtaining a high school degree (2.30% [27/1172] vs 12.09% [564,889/4,671,164]; education data from the Federal Bureau of Statistics was available only for persons aged between 25 and 64 years and participants' characteristics were adjusted accordingly). Additional participant characteristics obtained via the baseline questionnaire are outlined in Table 1.

Over the course of the intervention, participants walked on average 10,410 steps a day (Table 2), which can be considered high [31] but comparable with the results of other studies conducted with Swiss samples [34]. Monthly average step counts showed a slight negative trend over time. Naturally, step counts were higher for participants receiving a positive feedback mail because a positive feedback was sent out only if the average step count of the previous month exceeded 7500 steps. The attrition rate (23.35%, 379/1623) was comparable with previous pedometer-based interventions [35].

Effect of the Feedback Sign

Descriptively, a clear pattern emerged from the data as it is apparent from Table 3. For every feedback mail, step count averages decreased after a positive feedback and increased after a negative feedback. In absolute terms, the average increase after negative feedback (450 steps, SD=2032) was comparable with the average decrease after positive feedback (−425 steps, SD=1858). For emails 2 and 3, the negative change after a positive feedback was less pronounced whereas for those emails the positive change after a negative feedback was almost twice the amount of steps when compared with emails 1, 4, and 5. However, mean changes in step counts were accompanied by large SDs, calling into question whether positive and negative feedback indeed led to significantly different changes in step counts. Moreover, the observed pattern of results can possibly be explained by a regression to the mean due to the dependency of the feedback sign on the participants' physical activity.

Table 1. Participant characteristics.

Characteristics	T_1 questionnaire, n (%) (n=1220)
Education	
University	548 (44.92)
Professional school	208 (17.05)
High school	389 (31.89)
Secondary school	23 (1.89)
Primary school	4 (0.33)
Not declared	48 (3.93)
Place of residence	
Town	142 (11.64)
Outskirts of town	300 (24.59)
Village	598 (49.02)
Countryside	180 (14.75)
Income in CHF	
<2500	62 (5.08)
2501-5000	184 (15.08)
5001-7500	383 (31.39)
7501-10,000	210 (17.21)
>10,000	127 (10.41)
Not declared	254 (20.82)
Nationality	
Swiss	1098 (90.00)
German	55 (4.51)
Other	53 (4.34)
Not declared	14 (1.15)
Pedometer brand	
Fitbit	782 (64.10)
Fitbit app	249 (20.41)
Garmin	130 (10.66)
Jawbone	59 (4.84)
Pedometer bought for participation	
Yes	673 (55.16)
No	511 (41.89)
Not declared	36 (2.95)

Table 2. Descriptive statistics of monthly average step counts by feedback sign.

	Monthly step count average (SD)			Dropout, n (%)
	Positive feedback[a]	Negative feedback[a]	Total[a]	
Month 1	-	-	10967.02 (3744.64)	53 (03.27)
Month 2	11581.79 (3273.48)	6293.99 (2668.99)	10710.19 (3732.68)	68 (04.19)
Month 3	11639.40 (3145.38)	6470.87 (2298.96)	10714.99 (3597.49)	65 (04.00)
Month 4	11533.57 (3326.82)	6450.72 (2269.40)	10657.20 (3717.10)	53 (03.27)
Month 5	11216.21 (3395.22)	6046.97 (1945.79)	10366.47 (3742.52)	91 (05.61)
Month 6	11308.43 (4283.81)	5968.97 (2144.36)	10299.51 (4479.57)	49 (03.02)
Total	11462.60 (3489.37)	6252.79 (2291.78)	10409.96 (3427.29)	379 (23.35)

[a] Values represent monthly average step counts depending on the feedback at the beginning of the month.

Table 3. Descriptive statistics of change in monthly step count averages by feedback sign.

	Mean difference in step counts (SD)		
	Positive feedback[a]	Negative feedback[a]	Total[a]
Mail 1	−482.71 (2035.95)	363.30 (2114.51)	−349.37 (2070.88)
Mail 2	−328.77 (1858.73)	717.76 (2153.88)	−148.82 (1952.29)
Mail 3	−399.11 (1869.87)	587.71 (2012.82)	−234.52 (1929.12)
Mail 4	−497.77 (1854.33)	215.26 (1654.28)	−382.55 (1842.09)
Mail 5	−422.96 (1607.90)	329.19 (2112.57)	−285.12 (1735.20)
Total	−425.34 (1858.38)	450.29 (2032.06)	−279.51 (1916.24)

[a] Values represent the mean change in monthly average step counts after dispatch of the feedback mail.

Table 4. Summary of multiple regression results predicting change in average step counts.

Model parameter	*b*	Standard error (*b*)	Beta	*P*
Intercept	−139.97	98.39	-	.16
Time	−24.98	15.84	−0.018	.12
Baseline physical activity[a]	−0.13	0.01	−0.239	<.001
Group: financial incentives[b]	3.71	80.30	.002	.96
Group: charitable incentives[b]	5.42	81.85	.001	.95
Feedback sign[c]	−84.28	78.74	−0.062	.28

[a] Baseline physical activity was centered before entering the model.

[b] Group membership was represented as 2 dummy variables with the control group serving as the reference group.

[c] Feedback sign was represented as 1 dummy variable with negative feedback serving as the reference group.

Results of the multiple regression analysis of change in monthly step counts on time, baseline physical activity, group, and feedback sign are presented in Table 4. Overall, the linear model significantly predicted the changes in monthly average step counts ($F_{5,6836}$=90.84, P<.001) with R^2=.062. Baseline physical activity emerged as the only significant predictor of change in step counts. The feedback sign did not significantly predict the change in monthly average step counts when controlling for the effects of time, group, and baseline physical activity (Figure 4, right plot), indicating that the descriptive patterns in Table 3 were caused by regression to the mean.

Figure 4. Left: Scatterplot of changes in monthly step counts against baseline physical activity for the first feedback email; the dark solid line represents perfect agreement (no change) and the colored lines are regression lines for positive and negative feedback. Right: Overall difference between negative and positive feedback emails adjusted for baseline physical activity and other covariates.

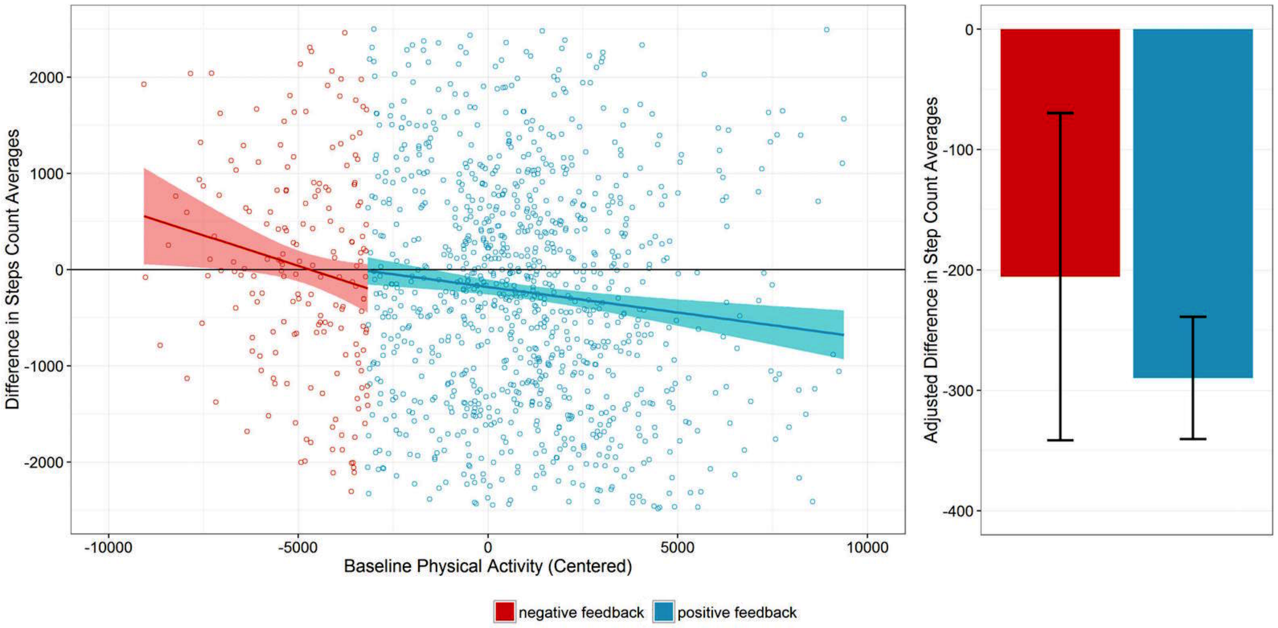


Figure 4 helps to further disentangle the effects of the feedback sign and baseline physical activity. The scatterplot in Figure 4 plots the dependent variable against participants' baseline physical activity to illustrate the influence of regression to the mean using data from the first feedback email [32]. The colored lines represent separate regression lines for participants that received positive and negative feedback. The slopes of the regression lines indicate that the change in physical activity following the feedback mail is dependent on physical activity in the previous month for both positive and negative feedback. More specifically, changes in physical activity occurred predominantly for individuals with extreme baseline values. In line with the assumption of regression to the mean being the cause of the observed change, participants with very low baseline activity levels increased their activity in the following month, whereas participants with very high baseline levels showed a decrease in the following month. The descriptively observed difference between positive and negative feedback vanishes when controlling for baseline physical activity (Figure 4, right plot).

Discussion

Principal Findings

In this paper we analyzed the effect of positive and negative feedback emails on physical activity of participants of a large-scale physical activity intervention. Using a quasi-experimental approach, we found no difference between the effect of positive and negative feedback emails. Substantial differences found on a descriptive level could be explained by regression to the mean. Contrary to the theory outlined in this paper, our results might suggest that the feedback sign does not influence the effect of feedback. Similar results have been found by previous research in other behavioral domains [18]. However, due to the quasi-experimental setting and possible alternative explanations for our results, we refrain from drawing firm conclusions regarding our research question.

Both frequency and relevance of the feedback could have limited the general effect of feedback on participants' physical activity levels in our study, thereby, explaining the missing effect of the feedback sign. Feedback in our study was only provided once at the beginning of each month, which might have not been frequent enough to affect monthly physical activity levels. Indeed, meta-analyses of feedback interventions in the area of health care demonstrate that feedback is more effective if it is delivered more frequently, for example, weekly [14]. Moreover, although the feedback messages contained novel and potentially interesting information (such as the final confirmation of earning a reward or tips on how to increase physical activity levels), the relevance of the feedback messages may have been compromised due to the permanent opportunity for participants to receive information on their step counts via their Web-based customer account or their activity tracker. Because irrelevant feedback is less likely to be processed by the recipient [36], this as well could have limited the effectiveness of the feedback emails. Including further personal information that is not yet provided by the activity tracker (eg, social comparisons) could enhance relevance of the feedback messages and might reveal differences between positive and negative feedback.

Limitations

Some methodological issues arise as part of the practical setting of our study. Participants were not randomly allocated to a negative feedback and a positive feedback condition. Rather, positive and negative feedback was dependent on participants' physical activity. As a consequence, we must not infer causality as inherent group differences beyond the included control variables may have affected our results. Furthermore, the internal validity of our results is limited as we were not able to check whether participants actually read the feedback emails.

If a substantial proportion of participants ignored the feedback, we might be able to observe an effect of the feedback sign only within a subgroup of our sample that read the feedback emails. Finally, we were not able to include a true control group that did not receive any feedback emails. However, providing evidence for the general effectiveness of feedback was not the primary focus of this paper as this has been investigated and confirmed in other studies [8,14,18]. Feedback represents a widely used technique of physical activity interventions and is likely to be continuously used in the future due to the central role of self-regulation for both theoretical models of behavior change [20,21,37] and the practical effectiveness of physical activity interventions [22,38]. We therefore chose to focus on the comparison between positive and negative feedback to explore the way to best incorporate feedback into physical activity and health behavior change interventions.

Future Research

Research regarding the effect of feedback in health behavior change interventions is in its infancy. Thus, we urgently call for the conduction of randomized controlled trials examining the effects of feedback on health behavior as well as related mediators and moderators. Results of those studies can help researchers and practitioners to decide how to best incorporate feedback in their health behavior interventions and thereby ensure a positive effect of feedback. In this context, digital technology might be a promising resource to maximize the effect of feedback [29]. Digital technology not only allows the delivery of automated and personalized feedback [39], sensors incorporated in mobile phones and wearables also facilitate so called just-in-time adaptive interventions (JITAIs) [40]. JITAIs are digital interventions that provide support to the user considering his or her state of vulnerability, opportunity, and receptivity. In other words, JITAIs provide support only to those persons who need it at the exact moment when they need it and can most benefit from it. Using the mobile phone, for example, feedback messages could be sent out only when the recipient is likely to actively process the feedback message (eg, when she is at home and no activities are scheduled in her calendar). Identifying the right conditions and context for feedback to be most effective holds great potential for health behavior change interventions.

Conclusions

There is no difference between the effect of positive and negative feedback emails that were sent out on a monthly basis in a large-scale physical activity intervention. Framing of the feedback in terms of success and failure may not be crucial when the feedback is given infrequently and in situations when individuals are likely to be aware of their levels of behavior. However, feedback characteristics, including the feedback sign, should be carefully considered when designing feedback to change health behaviors.

Acknowledgments

We would like to thank the institutional review board of the University of St. Gallen for their valuable feedback and support. The study protocol was approved by the Ethics Committee of the University of St. Gallen, Switzerland (reference number: HSG-EC-2015-04-22-A; date of approval February, 17th, 2016). Informed consent to participate was obtained from all participants of the study.

The study was funded in part by the CSS insurance, Switzerland. The funder had no role in reviewing or approving the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Mendis S, Armstrong T, Bettcher D, Branca F, Lauer J, Mace C, et al. Global status report on noncommunicable diseases. Geneva, Switzerland: World Health Organization; 2014.
2. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and diabetes study. Diabetes Care 1997 Apr;20(4):537-544. [Medline: 9096977]
3. Carson KV, Chandratilleke MG, Picot J, Brinn MP, Esterman AJ, Smith BJ. Physical training for asthma. Cochrane Database Syst Rev 2013 Sep 30(9):CD001116. [doi: 10.1002/14651858.CD001116.pub4] [Medline: 24085631]
4. Ibrahim EM, Al-Homaidh A. Physical activity and survival after breast cancer diagnosis: meta-analysis of published studies. Med Oncol 2011 Sep;28(3):753-765. [doi: 10.1007/s12032-010-9536-x] [Medline: 20411366]
5. Siebenhofer A, Jeitler K, Horvath K, Berghold A, Posch N, Meschik J, et al. Long-term effects of weight-reducing drugs in people with hypertension. Cochrane Database Syst Rev 2016 Mar 02;3:CD007654. [doi: 10.1002/14651858.CD007654.pub4] [Medline: 26934640]
6. Shaw K, Gennat H, O'Rourke P, Del MC. Exercise for overweight or obesity. Cochrane Database Syst Rev 2006(4):CD003817. [doi: 10.1002/14651858.CD003817.pub3] [Medline: 17054187]
7. Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. Diabetes Care 2001 Mar;24(3):561-587. [Medline: 11289485]

8. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? a review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112 [FREE Full text] [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
9. O'Brien N, McDonald S, Araújo-Soares V, Lara J, Errington L, Godfrey A, et al. The features of interventions associated with long-term effectiveness of physical activity interventions in adults aged 55-70 years: a systematic review and meta-analysis. Health Psychol Rev 2015 Jul;9(4):417-433. [doi: 10.1080/17437199.2015.1012177] [Medline: 25689096]
10. Olander EK, Fletcher H, Williams S, Atkinson L, Turner A, French DP. What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2013;10:29 [FREE Full text] [doi: 10.1186/1479-5868-10-29] [Medline: 23452345]
11. Winter SJ, Sheats JL, King AC. The use of behavior change techniques and theory in technologies for cardiovascular disease prevention and treatment in adults: a comprehensive review. Prog Cardiovasc Dis 2016;58(6):605-612. [doi: 10.1016/j.pcad.2016.02.005] [Medline: 26902519]
12. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
13. Hattie J, Timperley H. The power of feedback. Rev Educ Res 2007 Mar 01;77(1):81-112. [doi: 10.3102/003465430298487]
14. Ivers NM, Grimshaw JM, Jamtvedt G, Flottorp S, O'Brien MA, French SD, et al. Growing literature, stagnant science? systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. J Gen Intern Med 2014 Nov;29(11):1534-1541 [FREE Full text] [doi: 10.1007/s11606-014-2913-y] [Medline: 24965281]
15. Alvero AM, Bucklin BR, Austin J. An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). J Organ Behav Manage 2001 Feb;21(1):3-29. [doi: 10.1300/J075v21n01_02]
16. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012;6:CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
17. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2006 Apr 19(2):CD000259. [doi: 10.1002/14651858.CD000259.pub2] [Medline: 16625533]
18. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 1996;119(2):254-284. [doi: 10.1037/0033-2909.119.2.254]
19. Michie S, Abraham C. Interventions to change health behaviours: evidence-based or evidence-inspired? Psychol Health 2004 Feb;19(1):29-49. [doi: 10.1080/0887044031000141199]
20. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]
21. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
22. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
23. Carver CS, Scheier MF. Origins and functions of positive and negative affect: a control-process view. Psychol Rev 1990;97(1):19-35. [doi: 10.1037/0033-295X.97.1.19]
24. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
25. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
26. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
27. Van Dijk D, Kluger AN. Task type as a moderator of positive/negative feedback effects on motivation and performance: a regulatory focus perspective. J Organiz Behav 2010 Sep 07;32(8):1084-1105. [doi: 10.1002/job.725]
28. Van-Dijk D, Kluger AN. Feedback sign effect on motivation: is it moderated by regulatory focus? Appl Psychol 2004 Jan;53(1):113-135. [doi: 10.1111/j.1464-0597.2004.00163.x]
29. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Human Behav 2016 Apr;57:61-74. [doi: 10.1016/j.chb.2015.12.023]
30. Kowatsch T, Kramer J, Kehr F, Wahle F, Elser N, Fleisch E. Effects of charitable versus monetary incentives on the acceptance of and adherence to a pedometer-based health intervention: study protocol and baseline characteristics of a cluster-randomized controlled trial. JMIR Res Protoc 2016 Sep 13;5(3):e181 [FREE Full text] [doi: 10.2196/resprot.6089] [Medline: 27624645]
31. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De CK, Giles-Corti B, et al. How many steps/day are enough? for adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
32. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005 Feb;34(1):215-220. [doi: 10.1093/ije/dyh299] [Medline: 15333621]

33. Bundesamt für Statistik. 2017. STAT-TAB - Interaktive Tabellen URL: https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/forschung/stat-tab-online-datenrecherche.html [WebCite Cache ID 6opta985l]
34. Sequeira MM, Rickenbach M, Wietlisbach V, Tullen B, Schutz Y. Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey. Am J Epidemiol 1995 Nov 01;142(9):989-999. [Medline: 7572981]
35. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
36. Petty R, Cacioppo J. The elaboration likelihood model of persuasion. Adv Exp Soc Psychol 1986;19:123-205. [doi: 10.1016/S0065-2601(08)60214-2]
37. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors. Eur Psychol 2008 Jan;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]
38. Conn VS, Hafdahl AR, Brown SA, Brown LM. Meta-analysis of patient education interventions to increase physical activity among chronically ill adults. Patient Educ Couns 2008 Feb;70(2):157-172 [FREE Full text] [doi: 10.1016/j.pec.2007.10.004] [Medline: 18023128]
39. Yardley L. The potential of Internet-delivered behaviour change interventions. The European Health Psychologist 2011;13(3):40-43.
40. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34 Suppl:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]

Abbreviations

CT: control theory
JITAI: Just-in-Time adaptive Intervention
NCD: Non-communicable Disease
SCT: social-cognitive theory
SD: standard deviation

Edited by G Eysenbach; submitted 21.11.16; peer-reviewed by S Hermsen, RJ Renes, J Whiteley, K Ng; comments to author 05.02.17; revised version received 09.03.17; accepted 08.04.17; published 02.06.17.

Please cite as:
Kramer JN, Kowatsch T
Using Feedback to Promote Physical Activity: The Role of the Feedback Sign
J Med Internet Res 2017;19(6):e192
URL: http://www.jmir.org/2017/6/e192/
doi:10.2196/jmir.7012
PMID:28576757

Original Paper

Young People, Adult Worries: Randomized Controlled Trial and Feasibility Study of the Internet-Based Self-Support Method “Feel the ViBe” for Adolescents and Young Adults Exposed to Family Violence

Karin van Rosmalen-Nooijens[1], MD; Sylvie Lo Fo Wong[1], MD, PhD; Judith Prins[2], PhD; Toine Lagro-Janssen[1], MD

[1]Gender & Women’s Health, Department of Primary and Community Care, Radboud university medical center, Nijmegen, Netherlands

[2]Department of Medical Psychology, Radboud university medical center, Nijmegen, Netherlands

Corresponding Author:
Karin van Rosmalen-Nooijens, MD
Gender & Women’s Health
Department of Primary and Community Care
Radboud university medical center
Internal post 117
P.O. Box 9101
Nijmegen, 6500HB
Netherlands
Phone: 31 0 243613110
Fax: 31 0 243541862
Email: karin.vanrosmalen-nooijens@radboudumc.nl

Abstract

Background: Adolescents and young adults (AYAs) are of special interest in a group of children exposed to family violence (FV). Past-year prevalence of exposure to FV is known to be highest in AYAs and has severe consequences. Peer support is an effective approach to behavior change and the Internet is considered suitable as a mode of delivery.

Objective: The study aimed to evaluate both effectiveness and feasibility of a randomized controlled trial (RCT) and feasibility study of the Internet-based self-support method “Feel the ViBe” (FtV) using mixed-methods approach to fully understand the strengths and weaknesses of a new intervention.

Methods: AYAs aged 12-25 years and exposed to FV were randomized in an intervention group (access to FtV + usual care) and a control group (minimally enhanced usual care) after they self-registered themselves. From June 2012 to July 2014, participants completed the Impact of Event Scale (IES) and Depression (DEP) and Anxiety (ANX) subscales of the Symptom CheckList-90-R (SCL-90) every 6 weeks. The Web Evaluation Questionnaire was completed after 12 weeks. Quantitative usage data were collected using Google analytics and content management system (CMS) logs and data files. A univariate analysis of variance (UNIANOVA) and mixed model analysis (intention-to-treat [ITT], complete case) were used to compare groups. Pre-post *t* tests were used to find within-group effects. Feasibility measures structurally address the findings. The CONsolidated Standards Of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth (CONSORT-EHEALTH) checklist was closely followed.

Results: In total, 31 out of 46 participants in the intervention group and 26 out of 47 participants in the control group started FtV. Seventeen participants (intervention: n=8, control: n=9) completed all questionnaires. Mixed model analysis showed significant differences between groups on the SCL-90 DEP (*P*=.04) and ANX (*P*=.049) subscales between 6 and 12 weeks after participation started. UNIANOVA showed no significant differences. Pre-post paired sample *t* tests showed significant improvements after 12 weeks for the SCL-90 DEP (*P*=.03) and ANX (*P*=.046) subscales. Reported mean Web-based time per week was 2.83 with a session time of 36 min. FtV was rated a mean 7.47 (1-10 Likert scale) with a helpfulness score of 3.16 (1-5 Likert scale). All participants felt safe. Two-thirds of the intervention participants started regular health care.

Conclusions: No changes on the IES were found. SCL-90 DEP and ANX showed promising results; however, the calculated sample size was not reached (n=18). FtV functions best as a first step for adolescents and young adults in an early stage of change. FtV can be easily implemented without extensive resources and fits best in the field of public health care or national governmental care.

Trial Registration: Netherlands National Trial Register (NTR): NTR3692; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3692 (Archived by WebCite at http://www.webcitation.org/6qIeKyjA4)

(J Med Internet Res 2017;19(6):e204) doi:10.2196/jmir.6004

KEYWORDS

domestic violence; child abuse; exposure to violence; adolescent; young adult; telemedicine; peer group; peer influence; Internet; feasibility studies; randomized controlled trial; delivery of health care

Introduction

Family violence (FV) mostly affects women and children—about 30% of all women in a relationship reported to have experienced some form of violence in their relationship [1,2]; and in approximately 60% of the cases, children are living in these violent households [3,4]. Nowadays, it is widely acknowledged that children in violent homes are almost always exposed to this violence. The most common form of exposure to FV is exposure to intimate partner violence, including assaults by parents on siblings or among siblings. There are many possible forms of exposure that vary from direct exposure (seeing or hearing the violence) to indirect exposure (having to deal with the consequences of the violence for daily life) [5]. Recent studies show that 8-12% of all children were exposed to some form of FV in the preceding year [4,6,7]. In the last two decades, numerous studies were published on the effects of exposure to FV on children. Exposure to violence is associated with emotional, behavioral, and adjustment problems as internalizing and externalizing behavior, educational dropout, and mental health disorders such as post-traumatic stress disorder (PTSD), affective and depressive disorders, and suicidal attempts. It is associated with adolescent dating violence, high-risk sexual behavior, teenage pregnancy, and intergenerational transmission—becoming a victim or perpetrator of FV in adult life [8-20]. These consequences closely resemble those of direct victims of physical abuse. In the Netherlands, exposure to FV is therefore considered a form of child abuse [21].

Adolescents and young adults (AYAs) are a group of special interest in a group of children exposed to FV. Hamby (2011) showed that the past-year prevalence of any exposure to FV was highest in 14-17 year olds (13.8%) [6]. Adolescence is an important and life-altering period in human life with tremendous physical and psychological changes [22,23]. In this period, peers are highly important and are considered, more than family, significant others when facing problems.

The transtheoretical stages-of-change model by Prochaska and DiClemente [24] describes the process of intentional behavior change and can be used to describe the situation of AYAs exposed to FV. We hypothesize that AYAs exposed to FV to be in a precontemplation phase. In this phase, they are not yet ready to take action to change their situation. AYAs in a precontemplation phase may be unaware of the abnormality of their living situation and therefore unable to change it. In this phase, education focusing on what is and what is not normal in families could help move AYAs from precontemplation to a contemplation phase. This could make AYAs aware of the abnormality of their home situation and thus help them move from precontemplation to contemplation.

In the contemplation phase, AYAs exposed to FV are aware of the abnormality of this violence but are not committed to change this: they might be hesitant to share experiences with peers or significant others out of shame or fear that their home situation may harm their status (peer pressure) [25,26]. Besides, knowing that AYAs often support their mother by trying to protect her instead of being protected, they are not likely to seek help for themselves [27-29]. To induce change to a preparation phase, in which someone is ready to initiate change, it is needed to educate a person in a general and nonprovocative manner. Therefore, it is important to identify and reach AYAs exposed to FV as soon as possible.

Several reviews have addressed adolescent help-seeking behavior for mental health problems, including being a victim of violence. Barriers include attitude toward health care; confidentiality and trust issues; fear or stress about help-seeking; and lack of knowledge, accessibility, or recognition by others [30-32]. Having detected them, it is challenging to provide health care that is appropriate to both the age group and their specific problems [32-34]. Traditional programs are often not ready for participants in a precontemplation phase—directed at immediate change, they are not suitable for participants who are still weighing the pros and cons of changing their situation. Besides, most interventions need parent involvement, are regional, or do not offer specialized care.

In need of alternative ways to provide low threshold care to AYAs exposed to FV, the Internet could be an effective way to deliver care. In 2013, 100% of the adolescents between 12-25 years old in the Netherlands had access to the Internet [35]. The Internet is used by AYAs to retrieve health information and to self-disclose problems to peers on the Web, rather than asking for help elsewhere [36-38]. eHealth is a rapidly developing and upcoming mode of therapy. Internet-based intervention have been shown to be cost-effective in different health problems [39-43], including mental health in an adolescent population [44-46]. Provision of information, sharing experiences, and peer modeling are known to increase self-management [47-49] and peer- and social support are effective methods to change behavior, both offline and on the Web [49-55]. Furthermore, social support has proven to be effective in adults exposed to violence and is associated with good mental and physical health outcomes [56-58].

It is likely that AYAs exposed to FV will search for information on the Internet too. Therefore, an Internet-based self-help intervention targeted at this group could be a way to reduce barriers to help-seeking and help AYAs exposed to FV to become ready for change. To the best of our knowledge, there are no eHealth interventions for AYAs exposed to FV. Therefore in 2011, “Feel the ViBe” (feel the violence beaten) was developed as an alternative way to reach and support AYAs exposed to FV.

This study first describes a randomized controlled trial (RCT) and then, a feasibility study performed with the following research question: “Is Feel the ViBe an effective and feasible way of reaching and delivering support to adolescents and young adults exposed to family violence?”

Methods

Intervention

Feel the ViBe (FtV) is a freely-available, Internet-based self-support method for AYAs exposed to FV (self-assessed) [59,60] with three main goals: (1) providing information, (2) offering (peer) support, and (3) lowering the threshold to regular health care services by supporting participants to move to a higher level of change and to find health care fitting their needs. The intervention comprises a variety of elements, being among others a forum, a chat function, informational pages, and an “ask the expert” function (Table 1). A community manager (CM) moderates the intervention, answers questions, assesses safety, and supports participants when needed, both on demand when asked for, and actively when they judge a participant could use support of additional information. The CM is a semiprofessional with a background in health care and additional training on FV.

eHealth interventions often suffer from the law of attrition: the phenomenon of participants stopping usage or being lost to follow-up [61,62]. To facilitate exposure, we used expert literature and interviews with AYAs exposed to FV to compose the intervention and information on the Web, performed a search-engine optimization (SEO), and included general information about FV on the website of FtV to facilitate exposure for first-time visitors.

Participants could access FtV from any computer; they needed only their login name and password. FtV is to be used ad libitum without endpoint to the intervention; however, to facilitate exposure for returning visitors we included structured events and reminders [63-65]. FtV is described in detail in the study protocol of FtV [60]. The CONSORT-EHEALTH checklist was closely followed [66].

Participants

Participants included were AYAs aged 12-25 years, exposed to FV as defined previously, and who registered themselves at the homepage of FtV (feel-the-vibe.nl). Participants were included from June 2012 to January 2014. We did not actively recruit participants; all participants found FtV on Google or through other websites. In this early phase, inquiring participants about the type and severity of the violence would enlarge the threshold for participation. Therefore, we chose to consider every potential participant eligible as target group. All participants who started to use FtV were exposed to FV, including 2 participants who were excluded after randomization because of their age. Excluded were participants not in command of the Dutch language, as FtV is in Dutch.

If participants read the patient information letter available on the homepage and had no further questions, they were eligible to participate and were randomized in 2 parallel groups with a 1:1 allocation ratio: an intervention group, having access to FtV + usual care (UC), and a control group, having access to minimally enhanced usual care (mEUC), meaning that they were placed on a waiting list for 12 weeks with access to 24 h emergency care. Both groups have been extensively described in the study protocol of FtV (Figure 1) [60].

On their first login, participants and parents of participants aged 12 to 16 years gave informed consent electronically. Multiple identities were prevented by making the registration a manual process, including an Internet protocol (IP) address check and email contact, and the informed consent process.

A safety protocol ensured participant safety: FtV is based on a secured server and participants had to use a nickname. All personal data on the Web was removed by a CM, who also monitored the intervention. Participants could contact the CM in case of emergency, independent of their randomization group and electronic consent. All participants were obligated to give contact details of any adult they trust. In case of severe danger, the CM could contact this person with consent of the participant; or, if a participant was below 16 years, the CM could also contact without consent. Participants could not be blinded due to the nature of the study. Participants, recruitment, randomization, consent, and intervention are all described in detail in the study protocol of FtV [60].

Table 1. Overview of "Feel the ViBe" elements.

Element	Extra information	Restrictions
General information on exposure to family violence	Information by age (under 12 years, 12-17 years, 18-25 years, and parents) and by subject.	Public
Research information and disclaimer	Information for participants and parents about research, safety, and privacy.	Public
Information on sponsoring	Homepage, bottom left.	Public
Contact page	Option to register or ask questions to the community manager or researchers.	Public
News page	Twitter newsfeed included. The news page states important information for participants such as major bug fixes, changes in content, and scheduled maintenance.	Public
Emergency exit	A button on every page directing participants to a search engine without option to go back in the browser.	Public
Electronic consent for participants	Consent is necessary to get access to other elements behind login	Available after first login.
Electronic consent for parents	Consent is necessary for participants under 16 years old to get access to other elements behind login	Accessible by email with a code.
Questionnaires	Questionnaires will be activated in the personal menu. Questionnaires can be filled out one-by-one. Whenever possible, adaptive questioning is being used to make the burden as low as possible. There is a maximum of 15 questions per page. All items need to be filled out to submit a questionnaire. Participants cannot review their answers.	Available after first login, and every 6 weeks.
Personal menu	Menu for the participants with overview to all the available elements, access to the participants profile, digital testament, research information, and contact information.	Login needed
User profile	The profile contains information on the participant, being: full name, nickname, avatar, sex, age, contact details, and contact person. Only the nickname is available for other participants. The participant can choose a theme for the layout.	Login needed
Digital testament	The digital testament is required to fill out and lets participants choose how their data must be handled if they stop their participation.	Login needed
Ask the expert	Option to ask questions by email to several experts, including a general practitioner, a sexologist, a psychologist, and an expert in the field of family violence. Participants can also contact the community manager for general questions and questions regarding regular health care services. Response is given within 72 h.	Login needed
Forum	The forum is meant to stimulate peer support. The community manager moderates the forum and stimulates contact.	Login needed
Chat	Every 2 weeks we will offer a chat session for the participants with a specific theme and supported by an expert and the community manager. Every other week there will be an unguided chat.	Login needed
Information	Depending on the age in the profile, participants have access to tailored information about partner violence, sexual health, reproductive health, relations, and health care.	Login needed
Facts and figures	In a 12-week cycle, participants receive a 1-sentence fact of figure about family violence, sexual health, or reproductive health every day on their mobile or by email.	Consent needed

Primary Outcomes

The Impact of Event Scale (IES) was chosen to measure PTSD symptoms. The IES is a short set of 15 questions measuring the impact of events and the amount of distress associated with events. It comprises the subscales Intrusion (8 items, mean alpha=.86) and Avoidance (7 items, mean alpha=.82) [67-70].

The Depression and Anxiety subscales of the Symptom CheckList-90-R (SCL-90-R DEP and ANX) were chosen to measure an improvement in symptoms of depression and anxiety. The SCL-90-R DEP and ANX subscales measure symptoms of depression and anxiety during the previous week on a 5-point Likert scale. Both subscales showed good convergent and divergent validity and high internal consistencies. The SCL-90-R is validated for participants aged 12 years and older. The Depression subscale comprises 16 items (alpha=.90), and the Anxiety subscale comprises 10 items (alpha=.88) [71].

Feasibility and General Measures

The General Questionnaire (GQ) collected data on demographics and other (health) care and support received.

The Web Evaluation Questionnaire (WEQ) contained questions about content, layout, and the perceived effectiveness and usefulness of the website. The WEQ was meant to identify issues for further improvement of FtV, collect possible facilitators and barriers for implementation, and evaluate if the website met the expectations of the target group.

"Use" was measured by the collection of quantitative data (login frequency and duration, visited pages, and visitor numbers) and qualitative data (forum and chat entries and questions asked to the experts) and was monitored on a continuous base [63,72].

Qualitative data were collected from open-ended questions in questionnaires and from CM reports, including their daily activities and actions [73-75].

Data Collection

Participants in both groups were asked to complete the IES and SCL-90 DEP and ANX at baseline (T0) and every 6 weeks until 24 weeks after inclusion (T1-T2-T3-T4). The GQ was completed at baseline and after 12 (T2) and 24 (T4) weeks. The WEQ was completed 12 weeks after receiving full access to FtV, T2 for the intervention group and T4 for the control group. According to protocol, participants received up to 2 reminders if they did not complete a questionnaire within 1 week. All outcomes were self-assessed through Web-based questionnaires. All other data were collected on a continuous basis from Google analytics, content management system (CMS) logs, and data files (Figure 1).

Sample Size

As described in the study protocol, we calculated the sample size from studies on eHealth also using the IES as primary outcome measure [44,45,76,77]. From these studies we estimated that we needed 9 participants in each group after 12 weeks. Considering the relatively high effect sizes found in these studies and the expected loss to attrition, we aimed to include 50 participants for each group. We managed to include 93 participants within the inclusion period, of which 40 participants completed their baseline questionnaires and 17 participants (8 intervention, 9 control) completed at least their T1 and T2 questionnaires. This means that we did not reach our estimated sample size of 9 participants in each group.

Changes to Protocol and Unexpected Events

No changes were made to the intervention. During the study period, every month between 1:00 AM and 6:00 AM there were brief (5-15 min) security updates. There were 5 bug fixes, three for the chat function, one for the login procedure, and one for the firewall. Bug fixes were planned in the night and lasted a maximum of 30 min. No content was changed. The procedure book for the intervention had minor changes 2 times during the intervention, none of which had visible impact for participants.

In the study protocol, we stated that all questionnaires were obligatory. In practice, however, due to ethical concerns, we did not exclude participants if they didn't complete all baseline questionnaires. We were not able to report results on routine outcome measurements (ROM); participants often used the emergency exit instead of logging out, making results on ROM unusable. The adapted version of the "Seks onder je 25e" (Sex-under-25) questionnaire was hardly completed by participants because they did not feel comfortable with the subject. Therefore, we could not use the questionnaire and excluded it from analysis.

Data Analysis

Effectiveness

Participant characteristics were collected from the contact forms and GQ. To maximize feelings of safety and anonymity, data on FV exposure was not actively collected. Therefore, data on FV exposure was collected from qualitative data. Characteristics of the intervention and the control group were compared to check whether randomization resulted in similar groups.

Initially, we planned an intention-to-treat (ITT) analysis, in which the intervention group, irrespective of Web-based activity and adherence to protocol, was compared with the control group at T0, T1, and T2. However, imputing missing data, which is essential to perform a full ITT analysis, is an important problem we expected to happen in an Internet-based self-support intervention for AYAs because of loss to attrition. After consultation of a medical statistician, we decided not to use imputation techniques for missing data, as this would lead to a too large level of uncertainty in the relatively small group.

Therefore we primarily performed a complete case analysis using univariate analysis of variance (UNIANOVA) and mixed model analysis in statistical package for the social sciences (SPSS version 22; IBM Corp) to analyze any effects of using FtV in the first 12 weeks, based on the randomization in the initial group (ITT).

Second, we performed a pre-post *t* test analysis for all participants being an active user during 12 weeks, independent of their original randomization group. *P* values <.05 were considered statistically significant. Whenever relevant, age-related differences were included.

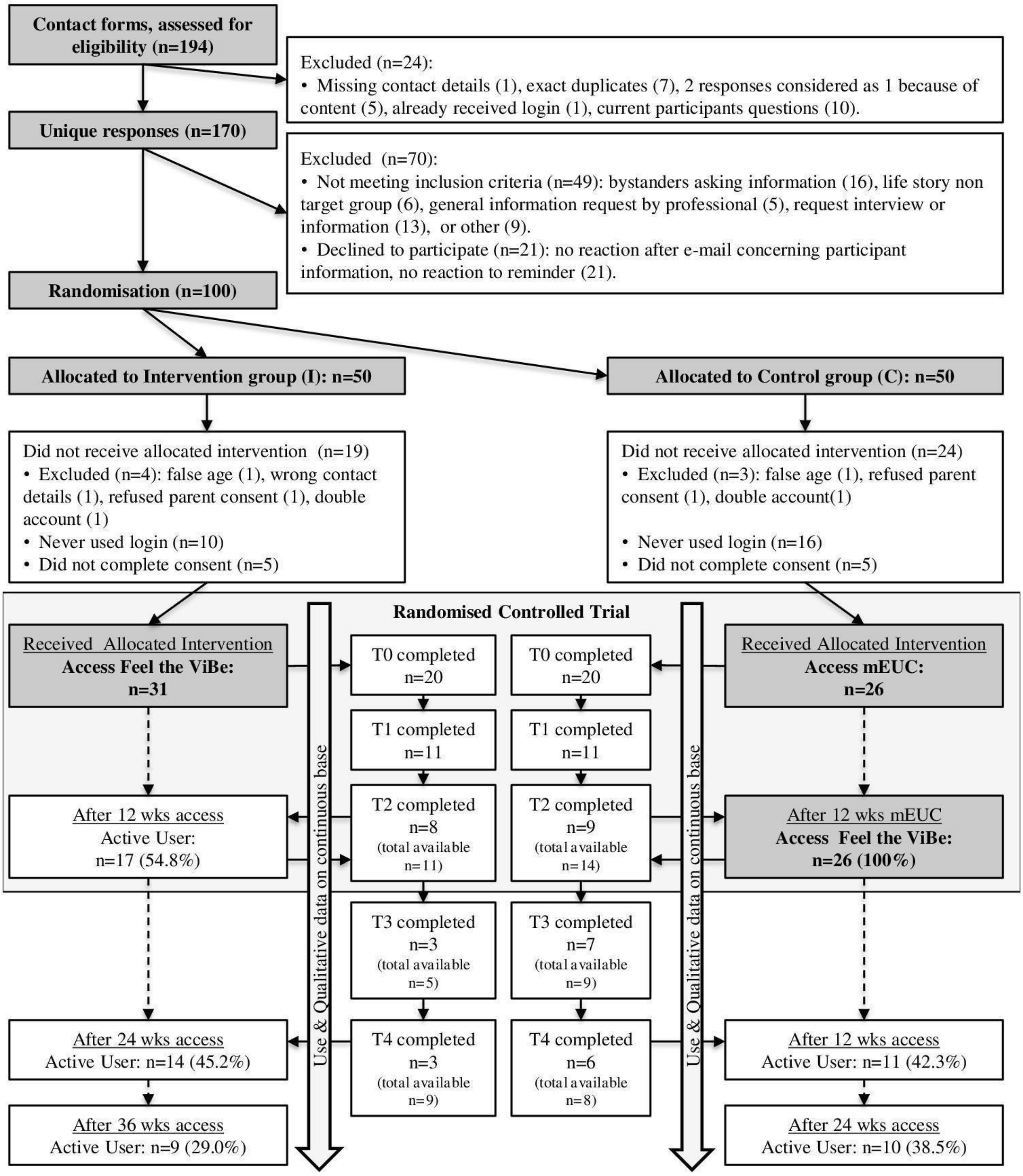
Feasibility

Feasibility studies aim to study several areas of focus to be able to fully understand the strengths and weaknesses of a new intervention: the potential for success when implementing it in the real-world. Feasibility was analyzed using several areas of focus, as suggested by Bowen et al (2009) [78,79]. Participant characteristics and Google analytics were used to assess expressed interest. Quantitative usage data were used to assess demand. First we focused on intention to use, followed by an analysis of actual use to assess continued use. Both individual elements as total usage time were included. Quantitative data were supported by self-reported qualitative data on use.

Quantitative and qualitative satisfaction assessed acceptability. Appropriateness was evaluated comparing user's wishes and needs, including safety, with expected goals as reported in the GQ. CM reports and a costs analysis were used to investigate possibilities for implementation in an uncontrolled design.

All qualitative data were analyzed using a thematic coding approach. Two researchers (KRN and SLFW) read and coded all qualitative log files. Consensus was reached in mutual discussion. Coded fragments were grouped in themes using feasibility measures as overall guideline. Any disagreements in coding were discussed in the supervising research team. Quotes are given a quote number corresponding with the number in Table 6.

Figure 1. Flowchart.



T=0 (baseline)	**T=1 (6 wks)**	**T=2 (12 wks)**	**T=3 (18 wks)**	**T=4 (24 wks)**	**Use**	**Qualitative data**
IES SCL-90 DEP SCL-90 ANX GQ	IES SCL-90 DEP SCL-90 ANX	IES SCL-90 DEP SCL-90 ANX GQ WEQ (for I)	IES SCL-90 DEP SCL-90 ANX	IES SCL-90 DEP SCL-90 ANX GQ WEQ (for C)	Google analytics CMS Logs CMS Data	Open-ended questions CM reports

Table 2. Participant characteristics as measured in the General Questionnaire (n=40).

Participant characteristics		Intervention (N=20)	Control (N=20)	*P* value
Age (years)				
	Mean age	18.40 (SD 3.62)	18.20 (SD 3.02)	.85
	12-17	8	10	.54
	18-25	12	10	
Sex				.32
	Male	1	0	
	Female	19	20	
Country of birth				.04
	Netherlands	20	16	
	Belgium	0	4	
Country of birth mother				.24
	Netherlands	17	12	
	Belgium	0	5	
	Other	3	3	
Country of birth father				.55
	Netherlands	17	13	
	Belgium	0	3	
	Other	3	4	
Religion				.29
	Christianity	11	9	
	Islam	1	1	
	No religion	5	8	
	Other	3	2	
Importance of religion				.06
	Not important	11	16	
	A bit important	7	4	
	Very important	2	0	
Employment				.54
	Full time education	10	11	
	Employed	3	3	
	Both studying and job	6	5	
	Disabled	1	1	
Education				.65
	Lower education	3	3	
	Middle education	9	5	
	Higher education	8	12	

Participant characteristics		Intervention (N=20)	Control (N=20)	*P* value
Current relationship				.38
	Boyfriend	5	7	
	Girlfriend	0	1	
	Dating	1	1	
	None	14	11	
Living situation				.71
	At home with parents	14	15	
	With partner	1	0	
	Alone	3	4	
	Sheltered housing	2	1	
Alcohol use				.90
	Daily	1	0	
	>1 time/week	3	5	
	<1 time/week	6	6	
	Never	10	9	
Smoking				.67
	Yes	1	2	
	Before	3	1	
	No	16	17	
Use of drugs				.35
	>1 time/week	1	0	
	<1 time/week	1	1	
	Never	18	19	

Ethics

This study was conducted in the Netherlands and was registered in The Netherlands National Trial Register (NTR3692).

The Committee on Research Involving Human Subjects of the Radboud University Nijmegen Medical Centre (Dutch initials: CMO) has assessed this study and judged that the study does not fall within the remit of the Medical Research Involving Human Subjects Act (WMO). Therefore, the study could be carried out (in the Netherlands) without approval by an accredited research ethics committee (2011/053. NL nr 35813.091.11. March, 16th, 2012).

Results

Effectiveness

100 participants, of which 9 male and 91 female participants with a mean age of 18.55 (SD 4.23) were included. After randomization, 7 participants were excluded. Fifteen participants in the intervention group and 21 participants in the control group did not give electronic consent, meaning that 31 out of 46 participants in the intervention group and 26 out of 47 participants in the control group started to use FtV. Of these participants, 20 participants in each group completed their baseline questionnaires and were included in the analysis (Figure 1). All these participants were exposed to FV. Eight participants in the intervention group, compared with 7 in the control group, were not only exposed but also a direct victim of FV. Overall participant characteristics are in Table 2.

Of the 40 participants who completed their baseline participants, 17 participants completed all questionnaires. There were no significant differences in age, sex, and type of violence between the participants who completed T0 (n=40, mean age 18.38 [SD 3.23]) and the participants who did not (n=17, mean age 16.94 [SD 3.49]) (*P*=.14). There were no significant differences in patient characteristics between the intervention and the control group at T0 (Table 2), except for country of birth and baseline measurements: The intervention group scored significantly lower scores on the IES with a mean score of 33.95, interpreting as “powerful impact event,” whereas the control group scored 45.60, interpreting as “severe impact event” (*P*=.01). Dropouts, participants who completed T0 but not T2 (n=15), were compared with completers (n=25); there were no significant differences at T0 for the intervention group, but in the control group, dropouts scored higher on the IES avoidance subscale (not significant, ns) and significant lower on the IES intrusion subscale (16.67 [SD 9.69] vs 24.64 [SD 5.40], *P*=.03).

We performed a UNIANOVA to correct for the differences on baseline scores. This showed no overall significant differences between T0 and T2 (Table 3).

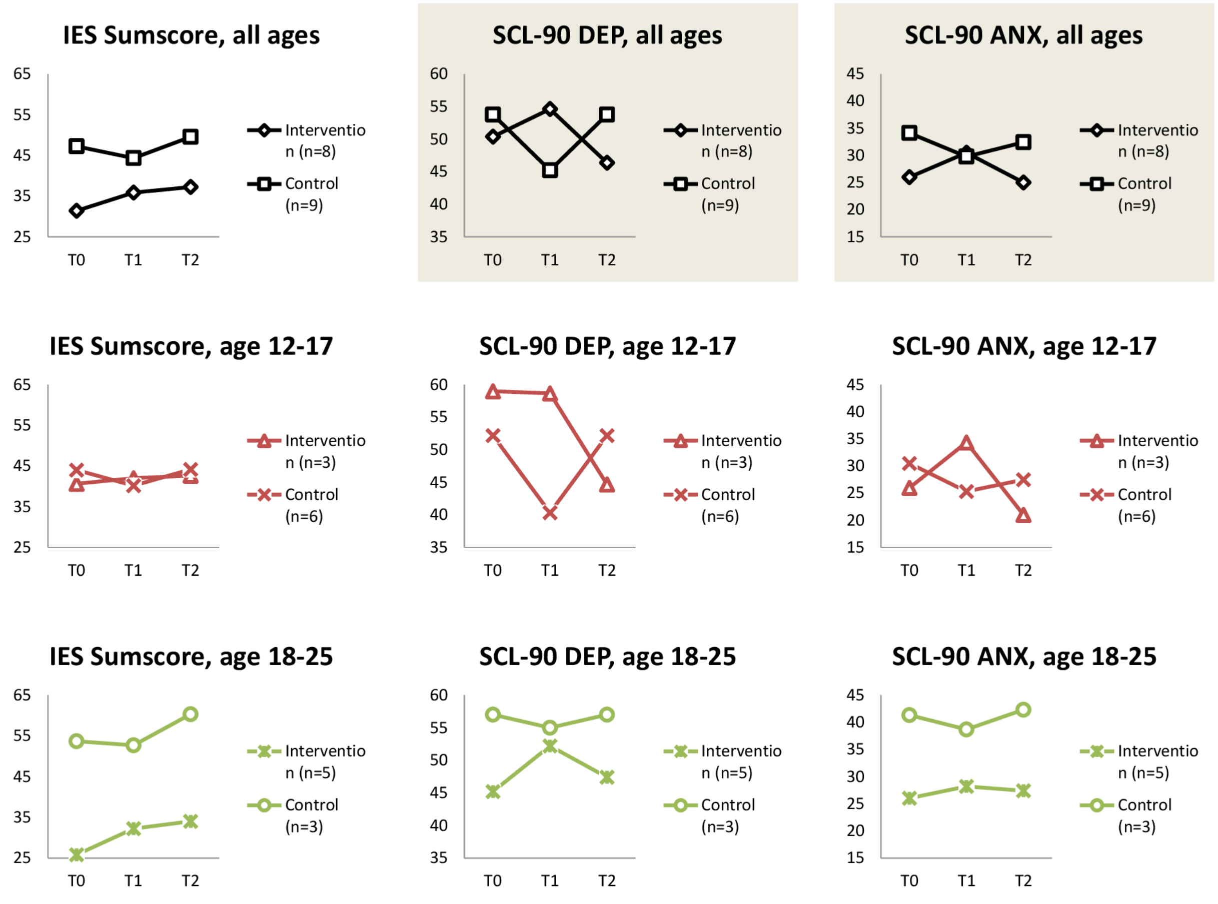
A mixed model analysis showed that the course of intervention participants is significant different for control participants on the SCL-DEP and ANX subscores. Repeating the analysis for separate age groups, results follow the same course, although results are no longer significant (Figure 2).

According to the protocol, the control group received full access to FtV after 12 weeks (Figure 1). Considering their T2-T4 measurements as T0-T2, we performed pre-post test measurements to further investigate the aforementioned findings. Considering them as two separate intervention groups, we saw that both groups improve on the SCL-90 DEP and ANX subscales between T0 and T2. For group 2, who had been on a waiting list for 12 weeks before receiving access, these differences were significant for all measurements (Table 4).

Qualitative Effectiveness

Of the participants who completed the WEQ, 58% (11/19) answered to an open-ended question that they were doing okay-to-fine at that moment. To investigate subjective efficacy, we asked participants whether FtV was helpful for them: on a 1-5 Likert scale, the helpfulness score was 3.16, with 42% (8/19) of participants saying that FtV was helpful or very helpful and 26% (5/19) saying that FtV was partly helpful. Additionally, 5% (1/19) said that FtV was not helpful at all. Being asked to motivate, participants said that “meeting others who experience violence” (58%, 11/19), “recognizing other’s stories” (58%, 11/19), and “being able to ask questions about violence” (58%, 11/19) was helpful. “Talking about their personal situation” (53%, 10/19) and “finding out if something was normal or not” (53% 10/19) were also indicated as helpful. Forty-two percent (8/19) of all participants said that giving support to others was also helpful for themselves.

Figure 2. Course in time for intervention compared with control group participants, gray area meaning that course difference is significant (*P*<.05).



Feasibility

Demand and Use

Google analytics data showed that 18,534 visitors visited the public website of FtV from June 1, 2012 to January 1, 2014. About 65.00% (12,047/18,534) of the users visited FtV only once with a bounce rate (percentage of users that leaves feel-the-vibe.nl after the first page they visit) of 58.00% (6987/12,047), which is considered good. The remaining 35.00% (6487/18,534) were recurrent visitors, visiting an average of 7.5 pages in 7 min per visit in the public part of the website. Most visitors accessed FtV either directly (32.00%, 5930/18,534) or by Google (47.00%, 8711/18534). A minority accessed FtV trough links on websites giving general information on child abuse or FV. Furthermore, 84% (16/19) of all participants who completed the WEQ agreed that FtV was easy to find on Google. A total of 194 visitors submitted a contact form. From these, 100 concerned eligible participants. Seven participants were excluded after randomization. Additionally, 31 out of 46 participants in the intervention group and 26 out of 47 participants in the control group started to use FtV (Figure 1). Participants under 16 years (needing parental consent), did not give less consent compared with older participants (65% [13/20] compared to 62% [44/71], respectively). Young adults (18-25 years) tend to give less often consent (55%, 27/49) than adolescents (12-17) (71%, 30/42), however, not significant (*P*=.13). After giving consent, 29 participants (among which all male participants) did not use the intervention. Four users used the intervention for 12-24 weeks. The remaining 24 users used FtV for 24 weeks or longer. There were no significant differences between adolescents (12-17) and young adults (18-25) in login count or session time, but adolescents (12-17) used the chat significant more than young adults in their first 12 weeks of access. Overall participant activity can be found in Table 5. Table 5 shows usage data of participants during their first and second 12 weeks on the Web.

Table 3. Effect of “Feel the ViBe” after 6 weeks (T1) and 12 weeks (T2) controlling for T0 (intention-to-treat, complete case). UNIANOVA calculating mean score, controlling for T0 to correct for differences between groups.

Outcome	Time	Intervention (n=8) Mean (95% CI)	Control (n=9) Mean (95% CI)	B	*P*
Impact of Event Scale					
	T1[a]	42.34 (34.71-49.97)	38.59 (31.44-45.73)	3.75	.48
	T2[a]	42.89 (34.38-51.40)	44.54 (36.57-52.51)	−1.65	.78
Symptom CheckList-90-R, Depression subscale					
	T1[b]	56.44 (48.60-64.29)	43.61 (36.21-51.00)	12.84	.02
	T2[b]	47.70 (38.73-56.67)	52.60 (44.15-61.05)	−4.90	.41
Symptom CheckList-90-R, Anxiety subscale					
	T1[c]	34.67 (28.16-41.18) 26.07 (19.96-32.18)	8.60	.07	
	T2[c]	29.18 (25.64-32.71)	28.73 (25.41-32.05)	.45	.85

[a] Mean score at T1 and T2 corrected for the overall mean score at T0 = 39.76 (n=17).

[b] Mean score at T1 and T2 corrected for the overall mean score at T0 = 52.18 (n=17).

[c] Mean score at T1 and T2 corrected for the overall mean score at T0 = 30.29 (n=17).

Table 4. Pre-post test analysis (n=14).

Outcome	Group 1 Mean (SD) (n=8)[a]			Group 2 Mean (SD) (n=6)[b]			All Mean (SD) (n=14)		
	T0	T2	*P*	T0	T2	*P*	T0	T2	*P*
IES[c] sum	31.38 (18.25)	37.25 (15.12)	.07	52.83 (9.77)	36.33 (10.78)	.02	40.57 (18.37)	36.86 (12.94)	.36
SCL-90 DEP[d]	50.38 (17.70)	46.38 (15.93)	.47	57.17 (11.99)	42.50 (10.71)	.01	53.29 (15.36)	44.71 (13.59)	.03
SCL-90 ANX[e]	26.00 (9.18)	25.00 (8.91)	.54	34.00 (9.82)	27.67 (8.34)	.045	29.43 (9.97)	26.14 (8.45)	.046

[a] Originally randomized to intervention group.

[b] Originally randomized to control group, considering T2-T4 measurements (after receiving full access) as T0-T2.

[c] IES: Impact of Event Scale.

[d] SCL-90 DEP: Symptom Checklist-90 Depression subscale.

[e] SCL-90 AUX: Symptom Checklist-90 Anxiety subscale.

Actual use measured on the Web corresponded well with self-reported data by participants. The 19 participants who completed the WEQ reported a mean Web-based time of 2.83 (2-3 times a week), with a mean Web-based session time of 36 min. Participants said they visited the forum and chat most and information on sex and relations least often, which corresponds with the quantitative usage data.

Table 5. Participant activity for all active users ≥24 weeks (n=24). It shows usage data of participants during their first and second 12 weeks on the Web.

Activity		0-12 weeks access	12-24 weeks access	*P* value
		Mean (SD)	Mean (SD)	
Total sessions (n)				
	All ages (n=24) 50 (45,57)	16 (15.63)	<.001	
	Age 12-17 years (n=11)	43 (45.56)	19 (15.29)	.05
	Age 18-25 years (n=13)	55 (46.72)	13 (15.89)	.001
	P	.54	.32	
Session duration minutes				
	All ages (n=24)	27 (14.12)	23 (14.37)	.35
	Age 12-17 years (n=11)	31 (13.46)	24 (14.14)	.32
	Age 18-25 years (n=13)	24 (14.36)	21 (15.00)	.71
	P	.23	.59	
Chat per session				
	All ages (n=24)	0.54 (0.28)	0.61 (0.34)	.16
	Age 12-17 years (n=11)	0.73 (0.21)	0.69 (0.26)	.31
	Age 18-25 years (n=13)	0.37 (0.22)	0.55 (0.40)	.04
	P	<.001	.34	
Forum per session				
	All ages (n=24)	0.69 (0.32)	0.65 (0.38)	.56
	Age 12-17 years (n=11)	0.59 (0.35)	0.49 (0.45)	.09
	Age 18-25 years (n=13)	0.78 (0.27)	0.79 (0.24)	.86
	P	.15	.06	
Chat and/or Forum per session				
	All ages (n=24)	0.91 (0.12)	0.91 (0.15)	.99
	Age 12-17 years (n=11)	0.95 (0.06)	0.91 (0.09)	.18
	Age 18-25 years (n=13)	0.89 (0.15)	0.91 (0.19)	.66
	P	.21	.98	
Information per session				
	All ages (n=24)	0.10 (0.09)	0.05 (0.07)	.02
	Age 12-17 years (n=11)	0.11 (0.11)	0.05 (0.07)	.09
	Age 18-25 years (n=13)	0.09 (0.07)	0.05 (0.08)	.12
	P	.64	.78	
Ask-the-Expert per session				
	All ages (n=24)	0.05 (0.05)	0.05 (0.11)	.91
	Age 12-17 years (n=11)	0.06 (0.06)	0.04 (0.07)	.27
	Age 18-25 years (n=13)	0.04 (0.04)	0.06 (0.14)	.60
	P	.33	.65	

Acceptability

Results from the baseline GQ and WEQ show that first impressions of FtV were positive—participants were enthusiastic (47%, 9/19) or felt that FtV was made especially for people in their own situation (32%, 6/19). At T0, "contact with fellow sufferers" was said to be both the most important wish and the most important need for their participation in FtV, followed by "someone listening without taking action right away." "Gathering information" and "receiving support or advice" were other frequently named wishes, also expressed as need. "Receiving help from Web-based health care providers" was indicated third most expressed wish for FtV but was not named as a need. Asked to goals, both in open and multiple-choice questions, the same 5 categories were named as most important (Table 6, Q1).

Table 6. Qualitative quotes by participants and community managers.

Question number	Participant number	Age	Source	Quote
Q1	290	19	GQ[a]	Giving people information and helping or supporting. And chatting with people who have gone through the same en by doing this helping each other.
Q2	209	20	WEQ[b]	Keep up the good work. I wish there were more people like you guys.
Q3	204	24	WEQ	I think the website is super, good initiative. You can find good and clear information. For me though I am not feeling a real connection or click with the others, which I think is because of the age difference. I pity that.
Q4	241	17	WEQ	The professional and the other participants answer your questions directly and help you immediately, and it feels like a relieve when you had a conversation like that.
Q5	207	21	WEQ	I don't think this is relevant, sex has, to my opinion, not always something to do with domestic violence. Sometimes it seems that, if it concerns adolescents, there always has to be a part about sexual education
Q6	228	17	WEQ	I feel safe because there is an emergency exit and your contact details are being stored, so when you really need help they can help you and they answer your questions personally.
Q7	209	20	WEQ	The Community Manager is very committed and gives you a warm feeling. I am not scared at all that she will tell anybody or forces me to do anything (which I know from my own experience).
Q8	CM[c]1	51	CM report	A strength of FtV is the time for participants to build a trusting relation and I fear this is not possible with a student.
Q9	CM3	26	CM report	I feel that FtV works, because it is not seen as healthcare by the participants, being not linked to any kind of organisation (...) thus feeling safe.

[a] GQ: general questionnaire.

[b] WEQ: Web evaluation questionnaire.

[c] CM: community manager.

Whereas support and information were very important, direct action was less asked for: Only one in five participants wished the violence at home would stop because of their participation and no participants identified it as a need. "Stopping the violence" (21%, 4/19) and "getting someone to receive help from a health care provider" (16%, 3/19) were chosen least often as goals of FtV.

After being on the Web for 12 weeks, participants expressed mostly joy about the existence of FtV (58%, 11/19) but only 16% (3/19) felt FtV had already helped them solve their problems. FtV was rated a mean score of 7.47 (range 6-9) on a 1-10 Likert scale (Q2, Q3). Overall, content, language, structure, user interface, and layout were rated good. Guided chat (42%, 8/19) and forum (37%, 7/19) were chosen the best parts of FtV because of the possibility to share stories and ask questions (Q4). All theme chats were valued positively, including the professional contact options and information about FV. Least valued were the unguided chat and informational pages, especially on sex and relations (Q5).

Safety

Safety was named second most frequently as need for FtV, direct after "contact with fellow sufferers." All participants said they felt safe, because of the relative anonymity (26%, 5/19); the emergency exit (21%, 4/19); acknowledgement and recognition by peers and professionals (21%, 4/19); the enjoyable, pleasant, and cozy environment (16%, 3/19); the approach and availability of the CMs (11%, 2/19); reliability of FtV (5%, 1/19); the technical security protocols (5%, 1/19); and/or having a safety protocol with contact details in case of participant danger (5%, 1/19) (Q6, Q7).

Seventy-four percent (14/19) of the participants thought that a personal message with the results of their questionnaires would make them feel even safer and 16% (3/19) thought that a better explanation of the safety protocols or 1 phone call with the CM could improve safety. In general, however, participants would not want more of their details to be known. Parent consent was perceived negatively.

Implementation and Practicality

CMs spent an average of 14 h a week on FtV. A guided chat lasted a mean 100 min, an individual chat, held only in case of danger or request for immediate help, 70 min. Around 70% (136/194) of all contact forms were answered within 36 h, the remaining 30% (58/194) within the maximum 72 h. All emergency messages were answered within 24 h, mainly within 6 h. CMs were asked to name essentials for success (Table 7). If these cannot be met, risks indicated concerned mainly harm for the participant and suboptimal care, but also a risk of burn-out for the CM, when time investment and commitment are too high.

Table 7. Essentials for successful implementation of “Feel the ViBe.”

General essentials[a]	Community manager characteristics[b]
Unrestricted access to the Internet for community managers	Computer and Internet skills
Unrestricted access to the internet for participants	Trained in giving Web-based support and help
Safety protocols to handle the variety of problems and participants	Trained in assessing safety during Web-based communication
24/7 availability in case of emergency (ICE) from pool of community managers	Flexible and able to adapt quickly in time, nature, and language of help provided
Colleagues to discuss participants’ situation	-

[a] Community managers were asked to name elements essential for FtV.

[b] Community managers were asked to identify personal characteristics of community managers essential for FtV.

An estimation of costs of development, execution, and future implementation was made calculating the costs from 2012 to 2014 (Table 8). Costs were calculated using actual costs for 2012, 2013, and 2014 (mean score is given when applicable).

Table 8. Estimation of costs for implementation of “Feel the ViBe.” Costs were calculated using actual costs for 2012, 2013, and 2014 (mean score is given when applicable).

Category	Necessity	Costs (US $)	Recurrent?
Intervention	Development only[a]	50,000	One-time only
Mobile app	Optional	10,000	One-time only
Hosting, security, and updates	Essential	12,000	Yearly
Internet, mobile and office resources, for example, computer, mobile phone, subscriptions.	Essential[b]	2000	Yearly
Salary costs: community manager (20h/week), coordinator (8h/week)	Essential[c,d]	48,000	Yearly
Professionals on consultation base	Essential	6000	Yearly

[a] Adaptation will cost about 10-25% of development costs, depending on need for translation.

[b] Costs based on minimally needed resources.

[c] Costs are calculated for Dutch salary norms, meaning that the actual costs can vary across countries depending on the salary norms.

[d] Costs could be lowered using trained volunteers or medical students.

CMs agreed that their work might be done by a student or volunteer with a background in health care, although they feared that the continuity of care would be endangered (Q8). Above that, they felt that the strain on personal life and the risk to become too personally involved were rather high.

To identify the position of FtV within the current field of health care, we asked participants in the baseline GQ to identify all possible sources of help when encountering FV. The mentor (73%, 29/40) and the school counselor (60%, 24/40) were the most named persons. The family physician was the most well-known health care provider—55% (22/40) knew how to get help there. One third of the participants did not know they could ask for help at the national emergency line and the police. At baseline, 18 participants received another form of health care, being mostly informal care from a mentor (15%, 6/40) or school counselor (13%, 5/40), or formal care from a psychologist (10%, 4/40). After the first 12 weeks of the intervention, 24 participants filled out the follow-up GQ. Two-thirds of them (15/24) started regular health care: mental health care (10/15), general practice (1/15), counselor or mentor (4/15), youth care (2/15), other Web-based help (2/15).

CMs agreed that FtV should be integrated within public health care or as part of the national services for FV and child abuse instead of existing health care institutions: they feared limited availability and accessibility, lack of anonymity, legal rules and/or regulations, and insurance requirements. CMs thought this would lower feelings of safety and enlarge the threshold for participation (Q9).

Discussion

Principal Findings

AYAs exposed to FV need health care in an early phase to deal with the consequences of this exposure. FtV was developed as a low-threshold, Internet-based, self-support intervention to provide AYAs exposed to FV with (peer) support and information, to lower the threshold to regular health care for those who need this. This study investigated first, the efficacy of FtV and second, the feasibility of FtV using a mixed methods approach. To our knowledge, we are the first to use this approach for the evaluation of an eHealth method in the field of FV. No strict conclusions on efficacy could be drawn as the participant rate was rather low. We conclude that the acceptability for FtV, including satisfaction and safety, was good. In the following paragraphs we would like to highlight some of the most important findings.

Effectiveness

Overall we feel that FtV is a suitable and satisfying intervention for the target group. However, in a small population with large differences in characteristics, it is difficult to find meaningful effects. Besides, we did not reach the sample size (17 instead of 18 participants). In our study sample, mean scores at baseline were very high, indicating a potential post-traumatic stress disorder in almost all participants. Intervention participants worsened in their scores before improving to levels above their start level. Control group participants showed the other way around. Further research exploring possible explanations is needed: One could hypothesize that control group participants improve because they are happy to have sought help, stepping up to a higher level of change, whereas intervention participants become aware of their situation and the abnormality of it. Depressive thoughts could increase as well as anxiety out of fear for the consequences of seeking help for their situation.

Overall, it is well known from literature that AYAs exposed to FV suffer from mental health problems, comparable with being a victim of child abuse themselves [6,8,25,80-82]. Treating this may take much longer than 6 or 12 weeks. Considering that, results indicating that SCL-90 ANX and DEP scores improve significantly within the second 6 weeks of access are promising. The subjective effect is high: participants feel helped by their participation and two-thirds of all participants started other health care while being a participant of FtV. Therefore, we think that FtV is a promising intervention, although future research should study prolonged effects over time.

Stages of Change

Characteristics of participants registering for FtV are diverse. All participants were exposed to FV. Fifteen participants were not only exposed but also a victim themselves, which is in accordance with literature [4]. Wishes, needs, and goals, however, are quite uniform and mainly directed at support, information, and safety, which is in line with other studies [28,30,31,33,83-86].

As we described in the introduction, we used the transtheoretical model of behavior change of Prochaska and Di Clemente [24] to categorize potential participants of FtV and hypothesized that most of them would be in a precontemplation phase, whereas regular health care is mostly directed at participants in an action phase. Results show us that the bounce rate is relatively low, meaning that many visitors visiting FtV want to know more about the subject. However, only a small part of the visitors to the public website of FtV send a contact form. This corresponds with a precontemplation phase; recognizing that their situation at home is different, participants in this phase might be looking for information only. This could be a too early phase to participate in FtV. Therefore, it is important to optimize access to FtV by extending both the public part of the intervention and Google optimization. Participants who send a contact form, use their login, and give consent might be one step further than what we hypothesized: in the preparation or planning phase. Participants in this phase are thinking about how to change their situation and are ready to take action in the near future. For them, is it important to lower the threshold to start their participation as far as possible; questionnaires and other obligated parts could hinder participants in this still early phase of change. Of the 57 participants who gave consent, only 40 participants completed their baseline questionnaires and 28 participants became active users. Therefore, interventions that mainly support participants in (precontemplation and preparation phases should focus first on safety and second on support and advice. A major pitfall in these phases is intervention requirements which enlarge the threshold for continued participation. In our target group for example, the safety protocol, including the contact person, could have enlarged the threshold for participation.

Attrition

eHealth interventions have to cope with the law of attrition: high levels of nonusage and dropout, which can be as high as 80-90% [61]. Considering these percentages, drop out for FtV is relatively low, with one-third of the participants still active after 36 weeks. Attrition can partly be explained by the stages of change model, but it is also important to identify possible other factors influencing nonusage and dropout since this influences the feasibility of the intervention. The threshold for participation in FtV was made as low as possible without compromising safety. However, a low threshold to participation and easy enrolment could lead to high dropout, as users fully realize the consequences of their participation only after they start participating. More strict information, personal contact (telephone, face-to-face), and making participants pay for their participation, all decrease nonusage and dropout attrition but also increase the threshold, which does not fit the early stage of change of most participants.

Another important factor in this study could be expectation management. Only after randomization and consent participants could see the intervention content and layout. Looking at the wishes and needs indicated by the participants, encountering others is very important. Finding out that there are not many participants in your first visit could lead to nonusage and dropout, meaning that new participants encounter the same problem again. We tried to solve this problem by giving the CMs more time to be on the Web and participate in discussions and chats in the first months of FtV.

Personal contact lowers dropout attrition rate. As FtV focuses on (peer) support from CMs and participants, this could have lowered the dropout attrition rate.

eHealth and Violence

Especially in this target group, that values safety besides peer support as most important need for participation, the advantages of an eHealth intervention are numerous. Due to its nature, it is easier to conserve anonymity and safety, lowering the threshold for participation. It is flexible and accessible. The participants lived in a large geographical range, and more often, they had difficulties visiting other health care institutions because they had to use public transport and needed money to do that. Being able to include participants from a large geographical range means that it is easier to collect enough people to give adequate support and to be economically profitable. Besides, eHealth means not having to explain absence due to therapy or other treatment requirements. To our opinion, this makes eHealth an ideal method to start health care for AYAs exposed to FV.

Limitations

eHealth studies have to deal with a wide range of challenges and traditional designs may be less suitable for an Internet-based self-support intervention [87]. Therefore, we chose not only a traditional RCT design but also an innovative feasibility design using mixed methods. However, there are limitations that should be considered when interpreting the findings of this study. First, it is only possible to apply the results to female participants, as no male participants completed their participation. Only 9 males applied; none of which completed their questionnaires. The set-up of FtV, explicitly stating that FtV gives (peer) support and information, could be more attractive to women. Males may be less hesitating to find regular health care, having less fear to harm family and surroundings. Besides, they are also more at risk for externalizing behavioral problems, which might need a different approach.

Second, in our study the control group and the intervention group differed on their baseline measurements. There could be several explanations for this finding. Due to the nature of this study, participants knew their group when starting to fill out the questionnaires. Participants in the control group may have felt let down or could have wanted to show their need for participation. It could also mean that less severe affected participants don’t start the waiting list because their problems are not severe enough. The groups did not differ on depression and anxiety at T0.

A limitation is the fact that, due to safety protocols and according to protocol, control group participants received mEUC, meaning that they could send an emergency message when they felt they needed this. Although in none of the cases there was actual danger and all messages were handled according to protocol, we still feel that participants may experience these short contacts as support, thus influencing results.

Implications and Recommendations

FtV is a promising intervention. Future research should focus on larger samples and investigate the optimal intervention duration.

We found that the role of the CM is very important. The CM supports, informs, and motivates, thus functioning as a coach around the clock. However, in the design of FtV, the CM was only intended to monitor, support, and inform from the background. Therefore, further research should investigate the position of the CM in a qualitative manner.

As we concluded that eHealth seems an adequate method to provide AYAs exposed to FV with care, we recommend further research to study eHealth in other groups coping with violence. One could expect results to be similar, although Internet literacy and access could be limiting potential effects.

FtV can be easily implemented without extensive resources. Nevertheless, implementing FtV within an existing health care organization, could lead to an enlargement of the threshold, or a situation in which participants who need more care are not always referred to the best health care option because of incomplete knowledge or strategy of the implementing organization. Therefore, we feel that FtV, when implemented, should be in the field of public health care or national governmental care, to provide the lowest possible threshold and long-term sustainability. FtV functions, to our opinion, best as a first step for AYAs in an early stage of change to get them ready for action and to fill the gap between exposure to FV and starting regular health care to stop violence and treat the consequences.

Conclusions

Based on the available data, we conclude that preset goals for FtV, that is, peer support, giving information, and support in finding regular health care, have been met, making FtV a promising intervention. Reaching back to the stages of change model, we feel that participants who are in a preparation stage benefit best from FtV in gathering information and receiving support, maximizing safety, and minimizing external control. FtV may help them to move on to the action stage, get ready to start regular health care treatment, or discover that they do not need more help as FtV provides them with sufficient support.

Acknowledgments

The authors would like to thank Hans Bor, statistician at the Radboud University Medical Center, Nijmegen, the Netherlands, for his help in the efficacy analysis of this study. The project is funded by Fonds Slachtofferhulp (victim support fund, the Netherlands). The sponsor had no influence on any part of the study.

Authors' Contributions

SLFW, JP, ALJ, and KRN are responsible for the design of the study. KRN performed the analysis and wrote the first draft of the manuscript. SLFW, JP, and ALJ revised the manuscript critically. All authors read and approved the final manuscript.

Conflicts of Interest

Dr Lo Fo Wong reported receiving governmental and nonprofit grants for a research project on mentor mother help for interpersonal violence, working as a trainer on recognizing partner violence, and receiving travel expenses to a World Health Organization (WHO) consultation group meeting on violence against women. Prof Lagro-Janssen reported receiving governmental and nonprofit grants to study partner violence. Prof Prins reported receiving grants to study eHealth interventions for patients with cancer.

Multimedia Appendix 1

Demonstration Powerpoint slides "Feel the ViBe.".

[PPSX File, 4MB - jmir_v19i6e204_app1.ppsx]

Multimedia Appendix 2

CONSORT EHealth Checklist.

[PDF File (Adobe PDF File), 501KB - jmir_v19i6e204_app2.pdf]

References

1. Devries KM, Mak JY, Bacchus LJ, Child JC, Falder G, Petzold M, et al. Intimate partner violence and incident depressive symptoms and suicide attempts: a systematic review of longitudinal studies. PLoS Med 2013;10(5):e1001439 [FREE Full text] [doi: 10.1371/journal.pmed.1001439] [Medline: 23671407]
2. Devries KM, Mak JY, García-Moreno C, Petzold M, Child JC, Falder G, et al. Global health. The global prevalence of intimate partner violence against women. Science 2013 Jun 28;340(6140):1527-1528. [doi: 10.1126/science.1240937] [Medline: 23788730]
3. Fantuzzo J, Boruch R, Beriama A, Atkins M, Marcus S. Domestic violence and children: prevalence and risk in five major U.S. cities. J Am Acad Child Adolesc Psychiatry 1997 Jan;36(1):116-122. [doi: 10.1097/00004583-199701000-00025] [Medline: 9000789]
4. McDonald R, Jouriles EN, Ramisetty-Mikler S, Caetano R, Green CE. Estimating the number of American children living in partner-violent families. J Fam Psychol 2006 Mar;20(1):137-142. [doi: 10.1037/0893-3200.20.1.137] [Medline: 16569098]
5. Holden GW. Children exposed to domestic violence and child abuse: terminology and taxonomy. Clin Child Fam Psychol Rev 2003 Sep;6(3):151-160. [Medline: 14620576]
6. Hamby S, Finkelhor D, Turner H, Ormrod R. NCJRS. 2011. Children's exposure to intimate partner violence and other family violence URL: https://www.ncjrs.gov/pdffiles1/ojjdp/232272.pdf [accessed 2016-04-19] [WebCite Cache ID 6gt02RIVb]
7. Finkelhor D, Turner HA, Shattuck A, Hamby SL. Violence, crime, and abuse exposure in a national sample of children and youth: an update. JAMA Pediatr 2013 Jul;167(7):614-621. [doi: 10.1001/jamapediatrics.2013.42] [Medline: 23700186]
8. Kitzmann KM, Gaylord NK, Holt AR, Kenny ED. Child witnesses to domestic violence: a meta-analytic review. J Consult Clin Psychol 2003 Apr;71(2):339-352. [Medline: 12699028]
9. Evans S, Davies C, DiLillo D. Exposure to domestic violence: a meta-analysis of child and adolescent outcomes. Aggress Violent Behav 2008 Mar;13(2):131-140. [doi: 10.1016/j.avb.2008.02.005]
10. Wolfe DA, Crooks CV, Lee V, McIntyre-Smith A, Jaffe PG. The effects of children's exposure to domestic violence: a meta-analysis and critique. Clin Child Fam Psychol Rev 2003 Sep;6(3):171-187. [Medline: 14620578]
11. Van NC, Pachter LM, Wade R, Felitti VJ, Stein MT. Adverse events in children: predictors of adult physical and mental conditions. J Dev Behav Pediatr 2014 Oct;35(8):549-551. [doi: 10.1097/DBP.0000000000000102] [Medline: 25225793]
12. Sternberg K, Baradaran L, Abbott C, Lamb M, Guterman E. Type of violence, age, and gender differences in the effects of family violence on children’s behavior problems: a mega-analysis☆. Dev Rev 2006 Mar;26(1):89-112. [doi: 10.1016/j.dr.2005.12.001]
13. Chan Y, Yeung J. Children living with violence within the family and its sequel: a meta-analysis from 1995–2006. Aggress Violent Behav 2009 Sep;14(5):313-322. [doi: 10.1016/j.avb.2009.04.001]
14. Bair-Merritt MH, Blackstone M, Feudtner C. Physical health outcomes of childhood exposure to intimate partner violence: a systematic review. Pediatrics 2006 Feb;117(2):e278-e290. [doi: 10.1542/peds.2005-1473] [Medline: 16452335]
15. Dube SR, Felitti VJ, Dong M, Giles WH, Anda RF. The impact of adverse childhood experiences on health problems: evidence from four birth cohorts dating back to 1900. Prev Med 2003 Sep;37(3):268-277. [Medline: 12914833]
16. Holt S, Buckley H, Whelan S. The impact of exposure to domestic violence on children and young people: a review of the literature. Child Abuse Negl 2008 Aug;32(8):797-810. [doi: 10.1016/j.chiabu.2008.02.004] [Medline: 18752848]

17. Ehrensaft MK, Cohen P, Brown J, Smailes E, Chen H, Johnson JG. Intergenerational transmission of partner violence: a 20-year prospective study. J Consult Clin Psychol 2003 Aug;71(4):741-753. [Medline: 12924679]
18. Widom CS, Czaja SJ, Dutton MA. Childhood victimization and lifetime revictimization. Child Abuse Negl 2008 Aug;32(8):785-796 [FREE Full text] [doi: 10.1016/j.chiabu.2007.12.006] [Medline: 18760474]
19. Eriksson L, Mazerolle P. A cycle of violence? examining family-of-origin violence, attitudes, and intimate partner violence perpetration. J Interpers Violence 2015 Mar;30(6):945-964. [doi: 10.1177/0886260514539759] [Medline: 24997102]
20. Black DS, Sussman S, Unger JB. A further look at the intergenerational transmission of violence: witnessing interparental violence in emerging adulthood. J Interpers Violence 2010 Jun;25(6):1022-1042 [FREE Full text] [doi: 10.1177/0886260509340539] [Medline: 19801446]
21. Netherlands Youth Institute. Nederlands Jeugdinstituut (Netherlands Youth Institute). 2016. Kinderen die getuige zijn van huiselijk geweld (children exposed to domestic violence) URL: http://www.nji.nl/Vormen-Kinderen-die-getuige-zijn-van-huiselijk-geweld [accessed 2017-05-16] [WebCite Cache ID 6qUwHcy0T]
22. DiClemente R, Hansen W, Ponton L. Handbook of adolescent health risk behavior. New York: Plenum Press; 1996.
23. Lerner R, Steinberg L. Handbook of adolescent psychology. Hoboken, NJ: John Wiley & Sons; 2009.
24. Prochaska J, DiClemente C. Transtheoretical therapy: Toward a more integrative model of change. Psychotherapy: theory, research & practice 1982;19(3):276.
25. Cunningham A, Baker L. The adolescent's experience of intimate partner violence and implications for intervention. In: Graham-Bermann SA, editor. How Intimate Partner Violence Affects Children: Developmental Research, Case Studies, and Evidence-based Intervention. Washington DC: American Psychological Association (APA); 2011:247-272.
26. Sox R. Integrative review of recent child witness to violence research. Clin Excell Nurse Pract 2004;8(2):68-78.
27. Kelley ML, French A, Bountress K, Keefe HA, Schroeder V, Steer K, et al. Parentification and family responsibility in the family of origin of adult children of alcoholics. Addict Behav 2007 Apr;32(4):675-685. [doi: 10.1016/j.addbeh.2006.06.010] [Medline: 16839693]
28. Freake H, Barley V, Kent G. Adolescents' views of helping professionals: a review of the literature. J Adolesc 2007 Aug;30(4):639-653. [doi: 10.1016/j.adolescence.2006.06.001] [Medline: 16867305]
29. Fitzgerald MM, Schneider RA, Salstrom S, Zinzow HM, Jackson J, Fossel RV. Child sexual abuse, early family risk, and childhood parentification: pathways to current psychosocial adjustment. J Fam Psychol 2008 Apr;22(2):320-324. [doi: 10.1037/0893-3200.22.2.320] [Medline: 18410219]
30. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
31. Barker G, Olukoya A, Aggleton P. Young people, social support and help-seeking. Int J Adolesc Med Health 2005;17(4):315-335. [Medline: 16445071]
32. Cossar J, Brandon M, Jordan P. IOE. London: Office of the Children's Commissioner; 2011. 'Don't make assumptions': Children's and young people's views of the child protection system and messages for change URL: http://dera.ioe.ac.uk/2690/1/force_download.php%3Ffp%3D%252Fclient_assets%252Fcp%252Fpublication%252F486%252FChildrens_and_young_peoples_views_of_the_child_protection_system_.pdf [accessed 2016-04-19] [WebCite Cache ID 6gsyiZim3]
33. Buckley H, Carr N, Whelan S. 'Like walking on eggshells' service user views and expectations of the child protection system. Child Fam Soc Work 2011 Feb;16(1):101-110. [doi: 10.1111/j.1365-2206.2010.00718.x] [Medline: 000285754400011]
34. Whitehall-Smith M. Counselling & support services for young people aged 12-16 who have experienced sexual abuse: a study of the provision in Italy, The Netherlands & the United Kingdom. Edited by Mary Baginsky, NSPCC, London, 2001. 140pp. ISBN 1-84228-020-1 (Pbk). In: Child Abuse Review. London: National Society for the Prevention of Cruelty to Children; Sep 2002:333-334.
35. Centraal Bureau voor de Statistiek. 2013. ICT gebruik van personen naar persoonskenmerken (internet usage of persons divided to personal characteristics) database on the internet URL: http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71098NED&D1=0-14,33-133&D2=3,7-8&D3=7-8&HD=140213-1522&HDR=G1,G2&STB=T [accessed 2016-04-19] [WebCite Cache ID 6gt0z0go1]
36. Rowe SL, French RS, Henderson C, Ougrin D, Slade M, Moran P. Help-seeking behaviour and adolescent self-harm: a systematic review. Aust N Z J Psychiatry 2014 Dec;48(12):1083-1095. [doi: 10.1177/0004867414555718] [Medline: 25335872]
37. Best P, Manktelow R, Taylor B. Social work and social media: online help-seeking and the mental well-being of adolescent males. Br J Soc Work 2014 Nov 11;46(1):257-276. [doi: 10.1093/bjsw/bcu130]
38. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
39. Riper H, Kramer J, Keuken M, Smit F, Schippers G, Cuijpers P. Predicting successful treatment outcome of web-based self-help for problem drinkers: secondary analysis from a randomized controlled trial. J Med Internet Res 2008;10(4):e46 [FREE Full text] [doi: 10.2196/jmir.1102] [Medline: 19033150]

40. van SA, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
41. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Jan;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
42. Ritterband LM, Andersson G, Christensen HM, Carlbring P, Cuijpers P. Directions for the International Society for Research on Internet Interventions (ISRII). J Med Internet Res 2006;8(3):e23 [FREE Full text] [doi: 10.2196/jmir.8.3.e23] [Medline: 17032639]
43. McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:55-62 [FREE Full text] [Medline: 15231556]
44. Wagner B, Knaevelsrud C, Maercker A. Internet-based cognitive-behavioral therapy for complicated grief: a randomized controlled trial. Death Stud 2006 Jun;30(5):429-453. [doi: 10.1080/07481180600614385] [Medline: 16610157]
45. Lange A, Ruwaard J. Ethical dilemmas in online research and treatment of sexually abused adolescents. J Med Internet Res 2010 Dec 19;12(5):e58 [FREE Full text] [doi: 10.2196/jmir.1455] [Medline: 21169170]
46. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen Mathijs F G. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012;344:e2598 [FREE Full text] [Medline: 22517917]
47. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
48. Backett-Milburn K, Wilson S. Understanding peer education: insights from a process evaluation. Health Educ Res 2000 Feb;15(1):85-96 [FREE Full text] [Medline: 10788205]
49. Melling B, Houguet-Pincham T. Online peer support for individuals with depression: a summary of current research and future considerations. Psychiatr Rehabil J 2011;34(3):252-254. [doi: 10.2975/34.3.2011.252.254] [Medline: 21208865]
50. Parry M, Watt-Watson J. Peer support intervention trials for individuals with heart disease: a systematic review. Eur J Cardiovasc Nurs 2010 Mar;9(1):57-67. [doi: 10.1016/j.ejcnurse.2009.10.002] [Medline: 19926339]
51. Ancker JS, Carpenter KM, Greene P, Hoffman R, Kukafka R, Marlow LA, et al. Peer-to-peer communication, cancer prevention, and the internet. J Health Commun 2009;14 Suppl 1:38-46 [FREE Full text] [doi: 10.1080/10810730902806760] [Medline: 19449267]
52. Høybye MT, Dalton SO, Deltour I, Bidstrup PE, Frederiksen K, Johansen C. Effect of internet peer-support groups on psychosocial adjustment to cancer: a randomised study. Br J Cancer 2010 Apr 27;102(9):1348-1354 [FREE Full text] [doi: 10.1038/sj.bjc.6605646] [Medline: 20424614]
53. Kaplan K, Salzer MS, Solomon P, Brusilovskiy E, Cousounis P. Internet peer support for individuals with psychiatric disabilities: a randomized controlled trial. Soc Sci Med 2011 Jan;72(1):54-62. [doi: 10.1016/j.socscimed.2010.09.037] [Medline: 21112682]
54. Salzer MS, Palmer SC, Kaplan K, Brusilovskiy E, Ten HT, Hampshire M, et al. A randomized, controlled study of Internet peer-to-peer interactions among women newly diagnosed with breast cancer. Psychooncology 2010 Apr;19(4):441-446. [doi: 10.1002/pon.1586] [Medline: 19484712]
55. Takahashi Y, Uchida C, Miyaki K, Sakai M, Shimbo T, Nakayama T. Potential benefits and harms of a peer support social network service on the internet for people with depressive tendencies: qualitative content analysis and social network analysis. J Med Internet Res 2009;11(3):e29 [FREE Full text] [doi: 10.2196/jmir.1142] [Medline: 19632979]
56. Coker AL, Smith PH, Thompson MP, McKeown RE, Bethea L, Davis KE. Social support protects against the negative effects of partner violence on mental health. J Womens Health Gend Based Med 2002 Jun;11(5):465-476. [doi: 10.1089/15246090260137644] [Medline: 12165164]
57. Levendosky AA, Huth-Bocks A, Semel MA. Adolescent peer relationships and mental health functioning in families with domestic violence. J Clin Child Adolesc Psychol 2002 Jun;31(2):206-218. [doi: 10.1207/S15374424JCCP3102_06] [Medline: 12056104]
58. Tajima EA, Herrenkohl TI, Moylan CA, Derr AS. Moderating the effects of childhood exposure to intimate partner violence: the roles of parenting characteristics and adolescent peer support. J Res Adolesc 2011 Jun;21(2):376-394 [FREE Full text] [doi: 10.1111/j.1532-7795.2010.00676.x] [Medline: 21765624]
59. van Rosmalen-Nooijens K, Lagro-Janssen A, Prins J, LoFoWong S. Feel the ViBe. Nijmegen: Radboud university medical center, Gender & Women's Health; 2012 Jun. URL: https://www.feel-the-vibe.nl/ [accessed 2016-04-18] [WebCite Cache ID 6gruYSpxE]
60. van Rosmalen-Nooijens KA, Prins JB, Vergeer M, Wong SH, Lagro-Janssen AL. “Young people, adult worries”: RCT of an internet-based self-support method “Feel the ViBe” for children, adolescents and young adults exposed to family violence, a study protocol. BMC Public Health 2013;13:226 [FREE Full text] [doi: 10.1186/1471-2458-13-226] [Medline: 23497359]

61. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
62. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
63. Crutzen R, de NJ, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. Health Educ Behav 2011 Feb;38(1):49-62. [doi: 10.1177/1090198110372878] [Medline: 21189422]
64. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
65. Brouwer W, Oenema A, Crutzen R, de NJ, de Vries Nanne K, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]
66. 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]
67. Brom D, Kleber R. De schokverwerkingslijst (The Dutch version of the Impact of Event Scale). Ned T Psychologie en haar Grensgebieden 1985;40(3):164-168.
68. Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: a measure of subjective stress. Psychosom Med 1979 May;41(3):209-218. [Medline: 472086]
69. Sundin EC, Horowitz MJ. Impact of event scale: psychometric properties. Br J Psychiatry 2002 Mar;180:205-209 [FREE Full text] [Medline: 11872511]
70. van der Ploeg E, Mooren TT, Kleber RJ, van der Velden PG, Brom D. Construct validation of the Dutch version of the impact of event scale. Psychol Assess 2004 Mar;16(1):16-26. [doi: 10.1037/1040-3590.16.1.16] [Medline: 15023089]
71. Arrindell W, Ettema J. SCL-90 Guide for a multidimensional psychopathology indicator in Dutch: handleiding bij een multidimensionale psychopathologie indicator. Lisse: Swets Tests Services; 2009.
72. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
73. Britten N. Qualitative interviews in medical research. BMJ 1995 Jul 22;311(6999):251-253 [FREE Full text] [Medline: 7627048]
74. Friese S. Qualitative Data Analysis with ATLAS.ti. Thousand Oaks: SAGE Publications; 2011.
75. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
76. Lange A, Rietdijk D, Hudcovicova M, van de Ven JP, Schrieken B, Emmelkamp PM. Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. J Consult Clin Psychol 2003 Oct;71(5):901-909. [doi: 10.1037/0022-006X.71.5.901] [Medline: 14516238]
77. Lange A, van de Ven J, Schrieken B, Emmelkamp PM. Interapy. Treatment of posttraumatic stress through the Internet: a controlled trial. J Behav Ther Exp Psychiatry 2001 Jun;32(2):73-90. [doi: 10.1016/s0005-7916(01)00023-4]
78. Visser A, van Laarhoven HW, Govaert PH, Schlooz MS, Jansen L, van DT, et al. Group medical consultations in the follow-up of breast cancer: a randomized feasibility study. J Cancer Surviv 2015 Sep;9(3):450-461. [doi: 10.1007/s11764-014-0421-z] [Medline: 25579623]
79. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
80. Carpenter G, Stacks A. Developmental effects of exposure to intimate partner violence in early childhood: a review of the literature. Child Youth Serv Rev 2009 Aug;31(8):831-839. [doi: 10.1016/j.childyouth.2009.03.005]
81. Edleson J. Children's witnessing of adult domestic violence. J Interpers Violence 1999 Aug;14(8):839-870. [doi: 10.1177/088626099014008004]
82. Edleson JL, Ellerton AL, Seagren EA, Kirchberg SL, Schmidt SO, Ambrose AT. Assessing child exposure to adult domestic violence. Child Youth Serv Rev 2007 Jul;29(7):961-971. [doi: 10.1016/j.childyouth.2006.12.009]
83. Crisma M, Bascelli E, Paci D, Romito P. Adolescents who experienced sexual abuse: fears, needs and impediments to disclosure. Child Abuse Negl 2004 Oct;28(10):1035-1048. [doi: 10.1016/j.chiabu.2004.03.015] [Medline: 15519434]
84. Jobe A, Gorin S. 'If kids don't feel safe they don't do anything': young people's views on seeking and receiving help from children's social care services in England. Child Fam Soc Work 2012;18(4):429-438. [doi: 10.1111/j.1365-2206.2012.00862.x]
85. O'Connor PJ, Martin B, Weeks CS, Ong L. Factors that influence young people's mental health help-seeking behaviour: a study based on the health belief model. J Adv Nurs 2014 Nov;70(11):2577-2587. [doi: 10.1111/jan.12423] [Medline: 24720449]

86. Sheffield JK, Fiorenza E, Sofronoff K. Adolescents' willingness to seek psychological help: promoting and preventing factors. J Youth Adolesc 2004 Dec;33(6):495-507. [doi: 10.1023/B:JOYO.0000048064.31128.c6]
87. Ahern DK, Patrick K, Phalen JM, Neiley JD. An introduction to methodological challenges in the evaluation of eHealth research: perspectives from the health e-technologies initiative. Eval Program Plann 2006 Nov;29(4):386-389. [doi: 10.1016/j.evalprogplan.2006.08.008] [Medline: 17950866]

Abbreviations

AYAs: adolescents and young adults
CM: community manager
FtV: Feel the ViBe
FV: family violence
GQ: General Questionnaire
IES: Impact of Event Scale
mEUC: minimally enhanced usual care
SCL-90 DEP: Symptom CheckList-90R, Depression subscale
SCL-90 ANX: Symptom CheckList-90R, Anxiety subscale
UC: usual care
WEQ: Web Evaluation Questionnaire

Edited by G Eysenbach; submitted 20.05.16; peer-reviewed by K Hegarty, N Trabold, D Levesque, TR Soron; comments to author 24.08.16; revised version received 04.11.16; accepted 14.03.17; published 12.06.17.

Please cite as:
van Rosmalen-Nooijens K, Lo Fo Wong S, Prins J, Lagro-Janssen T
Young People, Adult Worries: Randomized Controlled Trial and Feasibility Study of the Internet-Based Self-Support Method “Feel the ViBe” for Adolescents and Young Adults Exposed to Family Violence
J Med Internet Res 2017;19(6):e204
URL: http://www.jmir.org/2017/6/e204/
doi:10.2196/jmir.6004
PMID:28606893

Original Paper

Effectiveness of a Web-Based Intervention in Reducing Depression and Sickness Absence: Randomized Controlled Trial

Till Beiwinkel[1], MA; Tabea Eißing[1], MA (Sociology), MSc; Nils-Torge Telle[1], PhD; Elisabeth Siegmund-Schultze[2], MD; Wulf Rössler[1,3,4,5], MSc, MD

[1]Innovation Incubator, Competence Tandem Integrated Care, Leuphana University of Lüneburg, Lüneburg, Germany

[2]KKH Kaufmännische Krankenkasse, Abteilung Leistungs- und Versorgungsmanagement, Hannover, Germany

[3]Institute of Psychiatry, Laboratory of Neuroscience (LIM 27), University of Sao Paulo, Sao Paulo, Brazil

[4]Psychiatric University Hospital, Zürich University, Zürich, Switzerland

[5]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Berlin, Germany

Corresponding Author:
Till Beiwinkel, MA
Innovation Incubator
Competence Tandem Integrated Care
Leuphana University of Lüneburg
Scharnhorststr 1
Lüneburg, 21335
Germany
Phone: 49 4131677 ext 7826
Fax: 49 41316777601
Email: till.beiwinkel@leuphana.de

Abstract

Background: Depression is highly prevalent in the working population and is associated with significant loss of workdays; however, access to evidence-based treatment is limited.

Objective: This study evaluated the effectiveness of a Web-based intervention in reducing mild to moderate depression and sickness absence.

Methods: In an open-label randomized controlled trial, participants were recruited from a large-scale statutory health insurance and were assigned to two groups. The intervention group had access to a 12 week Web-based program consisting of structured interactive sessions and therapist support upon request. The wait-list control group had access to unguided Web-based psycho-education. Depressive symptoms were self-assessed at baseline, post-treatment, and follow-up (12 weeks after treatment) using the Patient Health Questionnaire (PHQ-9) and Beck Depression Inventory (BDI-II) as primary outcome measures. Data on sickness absence was retrieved from health insurance records. Intention-to-treat (ITT) analysis and per-protocol (PP) analysis were performed.

Results: Of the 180 participants who were randomized, 88 completed the post-assessment (retention rate: 48.8%, 88/180). ITT analysis showed a significant between-group difference in depressive symptoms during post-treatment in favor of the intervention group, corresponding to a moderate effect size (PHQ-9: d=0.55, 95% CI 0.25-0.85, P<.001, and BDI-II: d=0.41, CI 0.11-0.70, P=.004). PP analysis partially supported this result, but showed a non-significant effect on one primary outcome (PHQ-9: d=0.61, 95% CI 0.15-1.07, P=.04, and BDI-II: d=0.25 95% CI −0.18 to 0.65, P=.37). Analysis of clinical significance using reliable change index revealed that significantly more participants who used the Web-based intervention (63%, 63/100) responded to the treatment versus the control group (33%, 27/80; P<.001). The number needed to treat (NNT) was 4.08. Within both groups, there was a reduction in work absence frequency (IG: −67.23%, P<.001, CG: −82.61%, P<.001), but no statistical difference in sickness absence between groups was found (P=.07).

Conclusions: The Web-based intervention was effective in reducing depressive symptoms among adults with sickness absence. As this trial achieved a lower power than calculated, its results should be replicated in a larger sample. Further validation of health insurance records as an outcome measure for eHealth trials is needed.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 02446836; http://www.isrctn.com/ISRCTN02446836 (Archived by WebCite at http://www.webcitation.org/6jx4SObnw)

(J Med Internet Res 2017;19(6):e213) doi:10.2196/jmir.6546

KEYWORDS

Internet; depression; absenteeism; cognitive therapy; randomized controlled trial

Introduction

Depression is highly prevalent in the working population [1]. It is estimated that over the course of one year, up to 26.7% of adults experience depressive symptoms and about 8.9% fulfill all criteria for a depressive disorder [2]. The resulting impairment and functional disability poses a substantial burden for the affected individual as well as the economy. Depressed employees have higher health care costs than those without depression [3,4], which in Europe contribute to a total estimated cost of 118 billion Euros per year [5].

Depression is linked to a high loss of work days [6]. In Germany, depression is a major driver of sickness absence and produces higher durations of sickness absence than other diagnoses of mental disorders [7]. When employees return to work after a depressive episode, distress often remains and performance is reduced [8,9]. Therefore, maintaining work capacity should be an important goal of clinical interventions. However, health promotion interventions targeting occupational health in employees with depression have been developed with mixed results [10-13]. Access to treatment remains limited, and the existing personal and structural barriers prevent those affected by depression from seeking timely, evidence-based help [14-16].

Web-based interventions are a promising tool to overcome the treatment gap in depression [17]. While generally using similar techniques as face-to-face therapy, such interventions are commonly delivered through websites and allow participants to access content at any time and work through lessons at their own pace. Web-based interventions vary in the level of therapist support [18], from entirely self-help to guided formats including regular therapist contact (eg, feedback via email). The advantages of Web-based interventions are their accessibility, a low threshold for help-seeking, relative anonymity, the patients' active role in (guided) self-help, and their low costs. However, in studies comparing Web-based interventions to usual care, risks associated with the dissemination of Web-based interventions have been reported as well [19,20]. Among the working population, Web-based interventions could especially benefit those who do not want to seek regular treatment because of negative perceptions of mental ill-health at the workplace.

The effectiveness of Web-based interventions in reducing depressive symptoms has been demonstrated repeatedly, but effect sizes vary considerably across studies [21-23]. For example, in the 19 studies that were included in the meta-analysis by Richards and Richardson [21], depression improvement in comparison with a control group ranged from no effect (d=−0.03) to strong effects (d=1.43). This heterogeneity makes it necessary to evaluate the interventions separately. Methodologically, weak control groups (eg, wait-list control instead of active control groups) and failure to employ intention-to-treat principles lead to an overestimation of the treatment effect [24]. Web-based interventions for depression have been studied among different clinical populations in Germany [25-29] but, to the best of our knowledge, no studies have yet focused on a Web-based intervention among a population with sick leave due to depression.

Participant self-reports are the primary outcome measure of eHealth trials. However, the lack of independent outcome assessments and the sole reliance on self-report measures limits this evolving field. For example, a report on the methodological quality of randomized controlled trials of Web-based interventions concluded that an increased use of independent outcome measurements is needed to improve the validity of efficacy studies [24]. To date, few studies employ independent outcomes and such attempts are limited to observer ratings of symptoms and do not extend to objective behavioral measurement of work absenteeism [30-32]. The lack of objective sickness absence measurements in research on Web-based interventions is surprising because sickness absence is frequently used as an integrated measure of health in other fields [33].

This study examined the effectiveness of a guided Web-based intervention in reducing depression and sickness absence among a high-risk population using both self-assessed depression and sickness absence assessments from health insurance records. We hypothesized that the Web-based intervention would be more effective in reducing depressive symptoms and sickness absence than the control group.

Methods

Study Design

This was a two-armed open-label randomized controlled trial. Participants were randomly assigned to either the intervention group (IG), with access to the guided Web-based intervention, or the wait-list control group (CG), with access to unguided Web-based psycho-education.

We used a computerized block randomization procedure (allocation ratio 1:1, block size 10). The researcher conducting the randomization had no information about the participants apart from their 6-digit codes and did not participate in the enrollment and assignment of the participants to study groups, which was handled by two different researchers. Outcome variables were assessed at baseline (T0) and 12 weeks after randomization (post-treatment, T1). In addition, a follow-up measurement was assessed 24 weeks after randomization (12 weeks after treatment, T2). Sample size calculation was based on expected between-group differences at follow-up. G*Power was used for sample size calculation [34]. First, we assumed a power of 0.80, an alpha level of 0.05, and a small to medium effect size (d=0.3), which results in N=357 to perform a two-sided t test for differences between two independent means. Second, adding 20% attrition rate at inclusion, post-assessment, and follow-up, we calculated that N=608 participants needed to be enrolled.

The study was approved by the ethical review board at Leuphana University of Lüneburg. The study was registered retrospectively on February 1, 2013, under the International Standard Randomized Controlled Trial Number ISRCTN02446836; http://www.controlled-trials.com/ISRCTN02446836. Despite retrospective registration, no participants were enrolled before registration.

Recruitment

Participants were recruited from a large-scale German statutory health insurance between January 2013 and April 2014, with the first participant enrolled in February 2013. We recruited members from Kaufmännische Krankenkasse (KKH), a statutory health insurance company with about 1.8 million members nationwide. First, to identify participants who were at high risk for sick leave due to depression, insurance members were screened for previous diagnosis of depression (International Classification of Disease codes F32.0, F32.1, F33.0, F33.1, and F34.1), previous sickness absence due to depression, and current sickness absence. Second, the study team sent an invitation letter to all positively screened insurance members along with study information, the informed consent form (see Multimedia Appendix 1), and a 6-digit code to login into the platform. Adults with a previous episode of mild to moderate depression (International Classification of Disease codes F32.0, F32.1, F33.0, F33.1) or dysthymia (F34.1) were included to avoid giving less intensive treatment than necessary. Before registration on the platform, a screening for exclusion criteria was performed. Participants with a score of ≥20 on the Patient Health Questionnaire (PHQ-9), indicating severe depression, were excluded. A second exclusion criterion was suicidality as measured by one item on the presence of suicidal thoughts. All participants had unrestricted access to treatment as usual during the study period, including access to the treatments and services which are typically available for depression in the German health care system (eg, psychotherapy and medication).

Intervention

The Web-based intervention “HelpID” is a 12-week, Web-based program based on cognitive-behavioral therapy, awareness training, and systemic counseling. The program was structured into 12 weekly sessions. Each lasted 30 to 45 minutes and included interactive elements, videos, and audios that explained depression-related themes (eg, symptoms and course) as well as graphs, illustrations, exercises, and guidance for awareness and relaxation. Each session was available one week after completing the prior session. Participants received weekly reminder emails when a new session was available. The program had a guided format with therapist contact upon request, that is, psychologists (bachelor level or higher) trained in the intervention approach provided feedback via email or telephone. The intervention was developed by a team of clinical psychologists headed by Dr Despina Lion, a clinical psychologist and therapist with extensive experience in systemic counseling, cognitive-behavioral therapy, and neurological psychology. It is accessible online [35] (see Multimedia Appendix 2). Since July 2016, the copyright of “HelpID” is owned by IVPNetworks GmbH, a private integrated care company. The intervention is commercially available to single users and is included in health care plans of statutory health insurances.

The intervention’s psychological approach includes cognitive-behavioral therapy, mindfulness training, and systemic counseling. During the development process, current research evidence on the respective therapies was used as the basis, and special emphasis was placed on a “person-based” approach, focusing on the perspectives of the people who would use the intervention. Cognitive-behavioral therapy is the most extensively researched psychological treatment approach in Web-based interventions [36]. From cognitive-behavioral therapy, the intervention used elements of cognitive restructuring, with an emphasis on dealing with negative moods and automatic thoughts, as well as exercises for behavioral activation. Mindfulness training has been used increasingly in psychotherapy over the past years. It was shown to be effective for depressive symptoms and can be adapted to online formats [37,38]. The intervention module on mindfulness engages the user in exercises to observe the self and to practice mindfulness in daily situations. Systemic counseling is a therapeutic approach that highlights the social context surrounding the individual and its resources [39]. Specifically, systemic questioning techniques and instructions were employed to make use of the participants’ social support. Systemic principles were presented in specific weekly sessions, while homework exercises on systemic therapy encouraged the participants to adopt a systemic viewpoint and behavior change in their everyday interactions.

Control Group

The control group was a wait-list plus psycho-education condition. During the 12-week study period, participants had access to text-based information on the nature of depression and its symptoms and treatment. The psycho-education content was developed by a team of trained psychologists (bachelor degree or higher) and was based upon scientific literature on depression (eg, the German S3-Guideline) [40]. This type of control condition was chosen because more active control groups (ie, psycho-education) are considered to be more methodologically valid than passive control groups (ie, wait-list conditions) [24]. There is evidence that psycho-education can reduce depressive symptoms and serve as an initial treatment in primary care [41]. The control group did not have access to therapist guidance. Participants were eligible to access the intervention after study completion, if they requested access.

Outcomes

The primary outcomes were self-assessed depressive symptoms with the Patient Health Questionnaire (PHQ-9) and the Beck Depression Inventory (BDI II). The PHQ-9 measured the severity of depressive symptoms over the preceding 2 weeks, resulting in a score between 0 to 27 points with higher values indicating more severe depression [42,43]. The PHQ-9 was shown to have good reliability and construct validity [42]. The BDI II uses 21 items to measure depression severity [44,45]. The BDI II showed good psychometric properties in German-speaking samples in regard to internal consistency, retest-reliability, and construct validity [46]. As a secondary outcome, quality of life was assessed using the Manchester Short Assessment of Quality of Life (MANSA) [47]—a 12-item scale rating the participants' satisfaction with different life domains. The MANSA has been validated in a Swedish sample and showed satisfactory internal consistency and construct validity [48]. User satisfaction was measured at post-assessment using the item "Overall, how satisfied are you with the program?" with four answer options: 1=very good, 2=good, 3=satisfactory, and 4=poor.

Information on work absenteeism was retrieved from health insurance records. In the German health care system, such standardized health data is collected routinely. Its primary purpose is cost reimbursement and quality assurance, but it can be made available for secondary analysis. Due to the routine data collection, health insurance records are assumed to have high ecological validity. We matched health insurance records from KKH health insurance with participants' data using a 6-digit participant code as identifier. The code was generated for each positively screened insurance member and was also used for registration on the study platform. We analyzed sickness absence data that covered the 90 days before randomization (baseline) and 90 days after intervention (post-assessment). Three sickness absence measures were constructed according to Hensing et al [33]. First, the number of persons who were absent at least once, second, absence frequency as the number of times a person was absent during the 90 day period irrelevant of duration, and third, absence duration as the total number of absence days during the 90 day period. Sickness absence data was not diagnosis-specific.

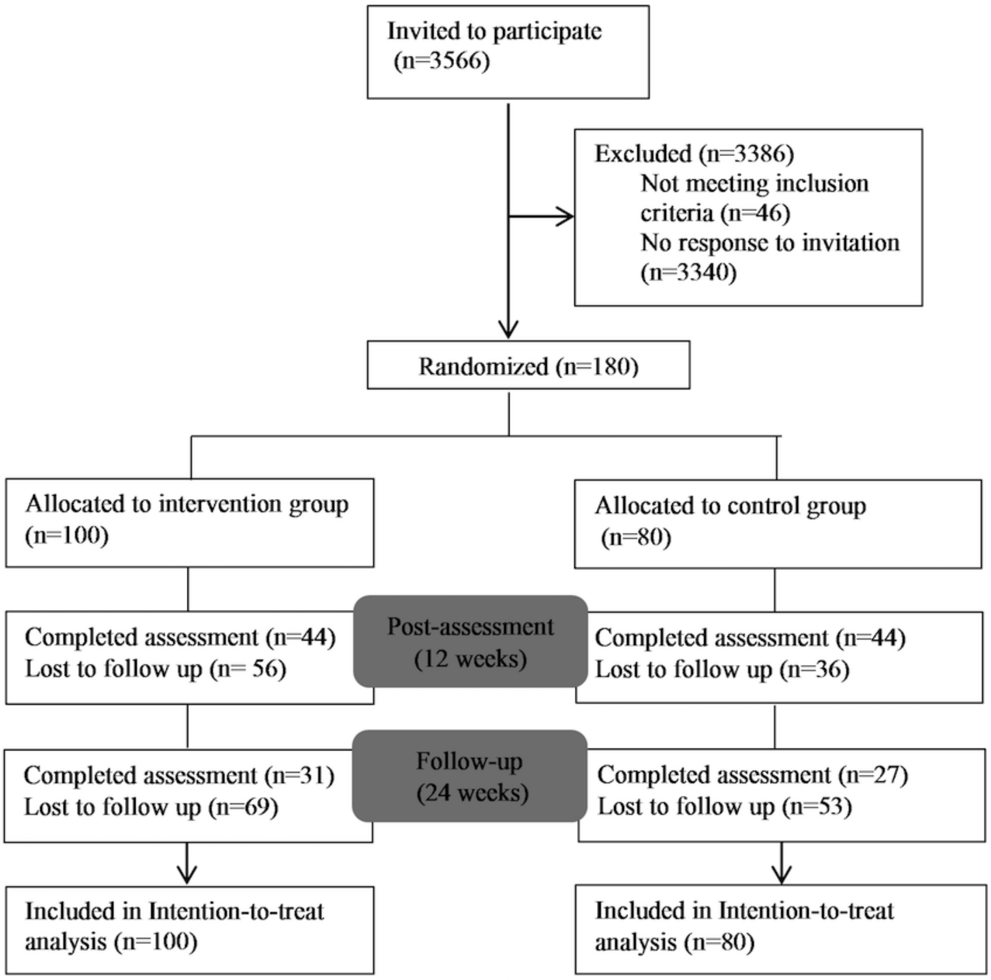
Statistical Analysis

Statistical analysis was performed on an intention-to-treat (ITT) basis according to the recommendations in the CONSORT statement [49] and its adaption for eHealth trials [50] (see Multimedia Appendix 3). Missing data at post-treatment was imputed using the Markov Chain Monte Carlo multiple imputation (missing data module in IBM SPSS version 22), where 10 estimations per missing value were specified and, besides the outcome variables, group assignment was included as an additional variable. Under the assumption that data is missing at random, multiple imputation was considered suitable to produce more precise estimates of the true intervention effect than other imputation methods, that is, last observation carried forward [51]. In addition, per-protocol (PP) analysis was performed to examine the robustness and sensitivity of the findings when including only participants who completed the post-assessment. *t* tests were used to determine differences in baseline characteristics and for within-group differences. The difference in the intervention outcomes between the intervention group and the control group at post-treatment was estimated using analysis of covariance (ANCOVA) with baseline scores as the covariate. Cohen *d* was calculated as a measure of the effect size, using pooled standard deviations [52]. For the between-group effect sizes, Cohen *d* was computed from the mean differences.

To assess clinical significance on an individual level, the reliable change index (RCI) was computed for the PHQ-9 [53]. Cronbach alpha=.89 from Kroenke et al [42] was used as an estimate of the reliability of the PHQ-9, along with pre-treatment standard deviations from the current study. Participants were classified as "responders" if they displayed a reliable positive change, or as "deteriorated" if they displayed a negative change on the RCI. A reliable positive change corresponds to less than −1.96 on the RCI and a change of PHQ-9 points to greater than −4.05. A reliable negative change corresponds to less than −1.96 on the RCI and a change in PHQ-9 points to greater than 4.05. Finally, the number needed to treat (NNT) [54] was computed.

All analysis was performed using Stata 13. The reported *P* values are two sided and in the 95% CI.

Figure 1. Study flowchart.



Results

Participant Characteristics

Figure 1 shows the flow of participants through the study. The complete pool of insurance members was screened, which resulted in 3929 positively screened insurance members who were subsequently invited to participate. Of those, 180 responded, met the inclusion criteria, provided informed consent, and were randomized. Of the 180 participants, 88 completed the post-assessment after 12 weeks (retention rate: 48.8%, 88/180), and 58 completed the follow-up assessment after 24 weeks (retention rate: 32.2%, 58/180). To estimate achieved power, a post-hoc power analysis was conducted. This revealed that with the sample of 180 participants, the achieved power to detect an effect of *d*=0.3 was 0.51.

Baseline characteristics between participants who completed the post-assessments and those who were lost to follow-up were tested for differences. Older participants (PHQ at T1: *P*=.02, BDI at T2: *P*=.03) and participants with higher education (PHQ at T1: *P*=.03, BDI at T2: *P*=.04) were more likely to complete the post-assessment on the primary outcome and the follow-up assessment. Participants who were not in psychotherapy during study enrollment were more likely to complete post-assessment on one of the primary outcomes (BDI at T1: *P*=.04) and the follow-up assessment (BDI at T2: *P*=.04) as compared with participants who were on a waiting list or in psychotherapy at enrollment. No other relevant differences between those who completed the post-assessments and those who were lost to follow-up were found.

Table 1 shows the participant characteristics at baseline. Participants had an average age of 48 years and were predominantly female (68%). The majority were married or had a partner (56%) and had completed secondary education or higher (85%). About half of the sample was working full-time (51%) and another 27% were working part time, while 20% of those working had an executive position. About one-fifth (21%) were not working. The majority had experienced a previous depressive episode (51%) or reported to have chronic depression (30%). During study enrollment, half of the participants were prescribed with depression medication (50%), 26% were in psychotherapy, and 13% were on a psychotherapy waiting list. Of the total number of enrolled participants, 43% had a sick leave certificate at the time. The mean baseline depressive symptoms in the sample were 11.10 points on the PHQ-9 (SD=4.45), indicating moderate depression. No clinically relevant differences in terms of any baseline characteristics were found, and we concluded that randomization was successful.

Table 1. Participant characteristics at baseline.

Characteristic	Intervention group n=100[a]	Control group n=80	Total sample n=180	*P* value
PHQ-9 Score, mean (SD)	11.53 (4.35)	10.56 (4.53)	11.10 (4.45)	.15
Age, mean (SD)	47.01 (10.36)	48.66 (11.59)	47.74 (10.92)	.31
Gender, n (%)				
Male	34 (34.0)	23 (28.7)	57 (31.7)	.45
Female	66 (66.0)	57 (71.3)	123 (68.3)	
Relationship, n (%)				
Single	24 (24.0)	18 (22.5)	42 (23.3)	.93
Married/Partner	56 (56.0)	46 (57.5)	102 (56.7)	
Divorced/Separated	15 (15.0)	14 (17.5)	29 (16.1)	
Widowed	5 (5.0)	2 (2.5)	7 (3.9)	
Education, n (%)				
Low	13 (13.3)	13 (16.2)	26 (14.6)	.38
Middle	63 (64.3)	42 (52.5)	105 (59.0)	
High	22 (22.4)	25 (31.3)	47 (26.4)	
Employment, n (%)				
Full-time	51 (52.0)	39 (50.6)	90 (51.4)	.41
Part-time	30 (30.6)	18 (23.4)	48 (27.4)	
Not working	17 (17.3)	20 (26.0)	37 (21.2)	
Executive position, n (%)				
Yes	20 (22.0)	13 (17.6)	33 (20.0)	.48
No	71 (78.0)	61 (82.4)	132 (80.0)	
Previous depression[b] , n (%)				
None	18 (18.0)	14 (17.7)	32 (17.9)	.53
Episodic	54 (54.0)	38 (48.1)	92 (51.4)	
Chronic	28 (28.0)	27 (34.2)	55 (30.7)	
Depression medication, n (%)				
Yes	53 (53.0)	36 (45.6)	89 (49.7)	.33
No	47 (47.0)	43 (54.4)	90 (50.3)	
In psychotherapy, n (%)				
Yes	30 (30.0)	16 (20.3)	46 (25.7)	.15
No	57 (57.0)	52 (65.8)	109 (60.9)	
Waiting List	13 (13.0)	11 (13.9)	24 (13.4)	

[a] All values (except for *P* values) are mean (SD) or n (%).

[b] Original item: “Did you have these symptoms for the first time?” Answer options: 1. “Yes,” 2. “No; I had one or multiple episodes,” 3. “No; the symptoms last for several years.”

Intervention Effectiveness

Table 2 shows the mean scores, standard deviations for the intervention outcomes at baseline and at post-assessment, effect sizes, and statistical significance, based on the intention-to-treat sample (imputed data). A significant between-group difference in favor of the intervention group was found for the PHQ-9 ($F_{1,179}$=15.06, *P*<.001), which corresponds to a medium effect size (*d*=0.55, CI 0.25-0.85). For those in BDI-II, a significant between-group difference at post-treatment was found ($F_{1,179}$=8.69, *P*=.004), which corresponds to a moderate effect size (*d*=0.40, CI 0.10-0.70).

Table 2. Means, standard deviations (SD), effect sizes, and statistical significance for intervention outcomes based on intention-to-treat sample (imputed data).

Outcome	Mean (SD)			Effect Size[a]	Statistical Significance[b]	
	Baseline (T0)	Post-assessment (T1)	Difference (T0−T1)	Cohen *d* (95% CI)	*F* Value	*P* value
PHQ-9						
Intervention	11.53 (4.35)	6.51 (2.87)	−5.02 (3.62)	0.55 (0.25-0.85)	15.06	<.001
Control	10.56 (4.53)	7.76 (3.63)	−2.80 (4.42)			
BDI-II						
Intervention	20.07 (7.99)	13.55 (6.46)	−6.17 (6.39)	0.41 (0.11-0.70)	8.69	.004
Control	18.78 (9.84)	15.52 (8.62)	−3.56 (6.68)			
MANSA						
Intervention	3.27 (0.72)	3.50 (0.67)	0.14 (0.71)	0,12 (−0.17 to 0.42)	0.72	.39
Control	3.30 (0.85)	3.44 (0.70)	0.22 (0.55)			

[a] Between-group effect size from mean differences.

[b] Based on ANCOVA controlling for baseline scores (T0).

In addition, both the intervention and the control group showed reductions in depressive symptoms as measured by within-group changes from baseline to post-assessment. In the intervention group, a mean reduction of 5 points on the PHQ-9 was found (t_{99}=14.28, *P*<.001), which corresponds to a large within-group effect size (*d*=1.42, CI 1.14-1.71). In the control group, a mean reduction of 2.79 points was found (t_{79}=5.82, *P*<.001), corresponding to a moderate effect size (*d*=0.65, CI 0.41-.89).

In the per-protocol analysis, we tested for differences in PHQ-9 scores between intervention completers and noncompleters at post-assessment. No significant difference was found (t_{178}=−.28, *P*=.78). Table 3 presents the results of the per-protocol analysis.

For the PHQ-9, a significant between-group difference in favor of the intervention group was found among completers (PHQ-9: $F_{1,77}$=8.98, *P*=.04), corresponding to a moderate effect size (*d*=0.61, CI 0.15-1.07). The mean PHQ-9 scores among completers were reduced by 5.70 points in the intervention group and by 2.24 points in the control group—this corresponds to a large effect size (*d*=1.72, CI 1.23-2.22) and a moderate effect size (*d*=0.49, CI 0.14-0.82) for within-group changes, respectively. For the BDI-II, the between-group effect failed to reach statistical significance in the per-protocol analysis ($F_{1,77}$=0.81, *P*=.37, *d*=0.25, CI −0.18 to 0.65). BDI-II within-group changes were significant in the intervention group (t_{43}=3.68, *P*<.001) and the control group (t_{42}=4.70, *P*<.001).

Table 3. Means, standard deviations (SD), effect sizes, and statistical significance for intervention outcomes based on per-protocol sample (nonimputed data).

Outcome	Mean (SD)			Effect Size[a]	Statistical Significance[b]	
	Baseline (T0)	Post-assessment (T1)	Difference (T0−T1)	Cohen *d* (95% CI)	*F* value	*P* value
PHQ-9						
Intervention	11.53 (4.35)	6.50 (3.85)	−5.03 (3.29)	0.61 (0.15-1.07)	8.98	.04
Control	10.56 (4.53)	7.95 (4.62)	−2.61 (4.61)			
BDI-II						
Intervention	20.07 (7.99)	14.86 (8.05)	−5.21 (7.59)	0.25 (−0.18 to 0.65)	0.81	.37
Control	18.78 (9.84)	15.32 (10.34)	−3.46 (6.32)			
MANSA						
Intervention	3.27 (0.72)	3.52 (0.86)	0.25 (0.64)	0.18 (−0.22 to 0.58)	0.34	.56
Control	3.30 (0.85)	3.42 (0.83)	0.12 (0.79)			

[a] Between-group effect size from mean differences.

[b] Based on ANCOVA controlling for baseline scores (T0).

Table 4. Work absence indicators at baseline and at post-assessment for 160 participants.

Indicator	Baseline (T0)[a]	Post-assessment (T1)	Difference (%)	*P* value
Absence at least once, n (%)				
Intervention	75 (85.22)	25 (28.41)	−50 (−66.67)	<.001
Control	58 (80.55)	12 (16.67)	−46 (−79.31)	<.001
Absence frequency, mean (SD)				
Intervention	1.19 (0.09)	0.39 (0.08)	−0.80 (−67.23)	<.001
Control	1.15 (0.10)	0.20 (0.06)	−0.95 (−82.61)	<.001
Absence duration, mean (SD)				
Intervention	25.60 (2.03)	24.65 (3.80)	−0.95 (−3.71)	.79
Control	27.69 (2.37)	24.04 (4.36)	−3.65 (−13.18)	.34

[a] Baseline and post-assessment cover a period of 90 days each.

Treatment Response

In the intervention group, 63% (63/100) of the participants showed a reliable symptom change from baseline to post-intervention and were thus classified as responders. In the control group, 33% (27/80) were classified as responders. The difference in reliable symptom change between intervention and control group was significant (t_{178}=3.39, *P*<.001). This resulted in a NNT of 4.08. One participant in the intervention group experienced symptom deterioration, and five participants in the control group experienced symptom deterioration.

Sickness Absence

Information on sickness absenteeism was available for 160 participants (Intervention group: n=88, control group: n=72). For 20 participants, sickness absenteeism could not be retrieved from insurance records, and therefore the data from these participants was unavailable.

Table 4 shows persons who were absent at least once, with absence frequency and absence duration at baseline and at post-assessment. The majority of the participants were absent at least once during the baseline period: IV: 85% (77/85), CG: 80% (58/72). Overall, significantly fewer participants were absent at least once during post-assessment (IV: 28%, CG: 16%). The within-group absence reductions were significant (IV: t_{87}=6.54, *P*<.001, CG: t_{71}=6.17, *P*<.001).

Regarding absence frequency, participants in the intervention group were absent on average 1.2 times at baseline and 0.4 times at post-assessment. In the control group, participants were absent 1.2 times at baseline and 0.2 times at post-assessment. The within-group reductions in absence frequency were significant (IV: t_{87}=7.49, *P*<.001, CG: t_{71}=8.59, *P*<.001).

Similarly, high absence durations at both baseline and post-assessment were found. From the 90 days examined at each time point, participants in the intervention group were absent from work 26 days at baseline and 25 days at post-assessment. In the control group, participants were absent 28 days at baseline and 24 days at post-assessment. However, there were no significant differences in absence duration from baseline to post-assessment (IV: t_{87}=.26, *P*=.79, CG: t_{71}=.95, *P*=.34). For all three measurements, the between-group differences at post-assessment failed to reach statistical significance (absence at least once: $F_{1,159}$=.80, *P*=.37, absence frequency: $F_{1,159}$=3.24, *P*=.07, absence duration: $F_{1,159}$=.02, *P*=.88).

Secondary Outcome

No significant difference in quality of life as measured by MANSA was found ($F_{1,169}$=.71, *P*=.40, *d*=0.13, CI −0.21 to 0.41).

Long-term effect

No significant between-group difference for BDI-II depression scores at 24-week follow-up was found ($F_{1,85}$=.81, *P*=.33). However, significant within-group changes were sustained for both the intervention and the control group. In the intervention group, there was a mean reduction from baseline to follow-up assessment of 5.46 points on the BDI-II (t_{99}=6.81, *P*<.001), which corresponds to a moderate effect size (*d*=0.68, CI 0.46-0.90). In the control group, there was a mean reduction of 4.69 points (t_{79}=4.37, *P*<.001), corresponding to a small effect size (*d*=0.48, CI 0.26-0.72).

User Satisfaction

87 participants completed the user satisfaction survey at post-assessment. In the intervention group, 13.6% (6/44) rated the program overall as very good, 68.2% (30/44) as good, and 18.2% (8/44) as satisfactory. In the control group, 4.6% (2/43) rated the program as very good, 37.2% (16/43) as good, 34.9% as satisfactory, and 23.3% (10/43) as poor. Mean satisfaction scores were 2.04 in the intervention group and 2.76 in control group. There was a significantly higher mean satisfaction in the intervention group (t_{85}=4.60, *P*<.001).

Discussion

Principal Findings

This study compared the effectiveness of a Web-based intervention in reducing depressive symptoms and sickness absence among adults with immediate risk for sickness absence due to mild to moderate depression. When compared with a wait-list plus psycho-education control group, participants who used the Web-based intervention showed a significantly greater reduction in depressive symptoms. However, because of the low response and high attrition in this study, one primary outcome did not reach statistical significance among participants who completed the intervention (per-protocol analysis) and at follow-up, only within-group changes were sustained but the intervention effect was not. In terms of individual clinical significance, significantly more participants in the intervention group responded to the treatment. We used health insurance records to measure sickness absence and found that sickness absenteeism declined in both groups, but there were no statistical differences in work absence between groups. The achieved power of this trial was lower than calculated. Therefore, its results should be replicated in a larger sample.

Comparison With Prior Work

These findings are especially relevant when considering the increasing impact of mental ill-health on workforces across industrialized countries [55]. Depression is a significant cause for workday losses and produces more absence durations than other mental illnesses. To reduce the illness burden, widespread access to evidence-based treatment is needed to maintain workers' mental health before companies and individuals face more serious burdens as the illness progresses. In general, Web-based interventions provide a promising treatment tool because these interventions can be accessed at any time and at different locations at the users' own pace. Due to their relative anonymity, Web-based interventions may especially benefit employees with depression who wish to avoid the negative perceptions of being mentally ill in the workplace. However, risks associated with the dissemination of Web-based interventions in the health care system have been reported as well. According to meta-analysis, the effects of Web-based interventions vary, making it necessary to evaluate each intervention separately. If Web-based interventions that are effective in reducing depressive symptoms are more widely implemented and adopted, a positive impact on the burden and impairment caused by depression can be expected. It can also help to overcome the shortcomings of conventional treatment (eg, waiting lists).

This study contributes to the growing body of research that supports the effectiveness of Web-based interventions for depression. Within this research, a critical mass of efficacy studies is needed to identify subgroups for which these interventions work [56]. Adults who are at high risk for sick leave from work due to mild to moderate depression have not yet been targeted specifically by Web-based interventions. In terms of effect sizes, previous research has found significant heterogeneity between studies. This makes it necessary to evaluate each intervention separately. The effect sizes reported in this study are comparable to other studies evaluating guided interventions for depression, including those included in the meta-analysis by Andersson and Cuijpers [23] where a mean between-group effect size of d=0.41 is reported. The amount of therapy guidance that is necessary to increase intervention effectiveness and adherence remains a subject of debate in the field of Web-based interventions [18,57]. Considering that this intervention provided only minimal therapist support upon request and achieved similar outcomes to studies with more intense guidance, we speculate that merely having the option to contact a therapist during the intervention—versus regular therapist contact—is sufficient for the needs of many participants and works equally well. However, it is possible that the number of dropouts could have been reduced with regular guidance. The examination of support preferences was not within the scope of this study and further research on this subject is needed.

We found that health insurance records are a suitable outcome for effectiveness research in Web-based interventions. Both groups showed reductions in work absence frequency, however, no statistical difference in work absence between groups was found. Several explanations may account for this finding. First, we cannot rule out the possibility that the decline in work absence frequency over time was caused by regression toward the mean or spontaneous remission. Our sample was recruited during a period of high levels of sickness absence, as seen in the data (see Table 3). Consequently, due to statistical chance, the frequency of sickness absence tended to approach lower levels at post-assessment. A healthy control group is needed to compare baseline levels of sickness absence, which was unavailable in this study. Second, it is possible that the 90-day time period in our study was not sufficiently long to appropriately detect changes in work absence. Previous population studies on sickness absence due to mental health problems found a median absence duration of 79 days [58]. This indicates that sickness absence started or ended outside of the time period of this study. Similarly, our sample was not adequately powered to detect small differences in work loss days. Third, organizational factors (ie, high work demands, job security) could have influenced work absence in this study. Unfortunately, we could not measure organizational variables.

To disentangle these explanations, future studies on the effect of Web-based interventions on work absence should include a longer time period, information on organizational factors that may be related to sickness absence, and work absence data from a healthy control group for baseline comparisons. Integrating objective behavioral parameters (ie, sickness absence data from health insurance companies) can increase the validity of effectiveness studies and might be a valuable addition to self-reported outcome measurements.

Privacy and Data Security

In Web-based interventions, health-related information is processed and stored electronically. Therefore, data security and confidentiality issues need to be taken seriously. This study used several measures to ensure the privacy of the study participants. Person-related information and study data were stored on separate servers to ensure that individuals could not be identified. Communication between the users' Web browsers and the servers were encrypted via a Secure Sockets Layer (SSL) connection. All data were stored on servers located in Germany.

Limitations

There are several limitations to this study that must be acknowledged. First, although reporting the effect of the Web-based intervention was within the scope of our study, it remains unclear as to which specific elements and properties of the intervention contributed to its effectiveness. Regarding the length of the intervention, evidence on the dose-response relationship in psychotherapy points to the conclusion that most progress occurs in the first few sessions of an intervention [59]. Similarly, Web-based interventions with 8 or less sessions were found to be more effective than interventions with more than 8 sessions [21]. This indicates that the present intervention with its 12 sessions could be shortened in length while maintaining its effectiveness.

A second limitation concerns the response rate. Response from the pool of positively screened insurance members was low (5.8%, 226/3929). This raises the concern that participants were particularly motivated to use a Web-based intervention when compared with nonparticipants, especially because nonparticipants were positively screened for sickness absence due to depression and thus belonged to the target group. Ideally, data from nonparticipants should have been collected as a baseline comparison group, but this data was unavailable in this study because of a lack of informed consent. When the complete pool of insurance members was screened and all positively screened insurance members were invited, recruitment was stopped. This resulted in a substantially smaller sample than was previously calculated in the power analysis (calculated N=680 vs actual N=180).

Third, attrition during the study was high. At post-assessment, 45.5% of the participants had dropped out, and at follow-up, 67.7% of participants had dropped out. In general, dropout is a common problem in Web-based interventions [60]. However, the dropout in this study was remarkably higher than the average attrition rate for Web-based interventions with therapist support, as reported in the meta-analysis by Richards and Richardson (28%) [21]. Several recent studies on Web-based interventions showed dropout rates that were remarkably lower [27,32]. One possibility for the relatively high attrition in this study is that participants who failed to complete a weekly session were reminded via email only. Comparable studies, which used telephone reminders, achieved substantially higher participant compliance. A second explanation is that therapist guidance was available upon request only. Guided interventions have lower attrition rates as compared to unguided interventions. Thus, it is possible that participants who were at risk for dropout were less likely to use therapist guidance. Despite our analysis of per protocol and imputed data showing comparable results, which indicates no difference in intervention completers versus noncompleters, the risk remains that study dropout could have biased the results. Due to the associations of several baseline characteristics (age, education, and being in psychotherapy) with the likelihood to complete the outcome assessments, the missing at random assumption could have been violated. Overall, the high attrition rates limit the conclusions drawn from this study.

Third, the positive relationship of age and education with study dropout seen here limits the generalizability of the findings to younger and less educated groups. This is supported by the composition of the study sample, where highly educated participants and those in executive positions are over-represented. Further studies with more statistical power are needed to identify effectiveness among different subgroups.

Fourth, participants had access to treatment as usual during the study, including psychotherapy and medication. Therefore we cannot exclude the possibility that within-group changes in depression scores were affected by third factor variables. Thus, within-group changes must be interpreted with caution.

Fifth, no clinical interviews were conducted to assess depression. Structured clinical interviews represent the gold standard of clinical assessment, with superior validity and reliability. Due to limited resources, this study relied solely on participants' self-reports to assess clinical symptoms.

Sixth, the amount and duration of provided therapeutic support during the study was not measured, making it difficult to compare the results with other studies on Web-based interventions that have used different levels of support, ranging from no support to more intensive and regular support. In this study, support was provided upon request, which could have prevented some participants from using support, thus lowering adherence.

Seventh, we used a wait-list plus psycho-education control group. Wait-list control groups undermine internal validity and may lead to an over-estimation of the treatment effect [24]. Thus, active control groups are considered to be less biased. To maximize participant response, we decided to inform control group participants that they could access the intervention upon request after study completion. This may have lowered expectations with regard to the control condition. Furthermore, control group participants were active during the study period as they had access to psycho-education. As a result, we observed mean symptom reductions in the control group, which is consistent with the finding that Web-based psycho-education can reduce depressive symptoms [41]. At the same time, it is possible that psycho-education has adverse effects in some participants, because it sensitizes patients to the topic of depression, leading to an over-reporting of symptom severity at follow-up. For example, a study that used a psycho-education control group found that the incidence of depression was higher than usual [32]. Our results show that reliable symptom deterioration was low overall, but occurred more frequently in the control group (6.25%, 5/80) as compared with the intervention group (1%, 1/100). This suggests that adverse effects in the control groups were present, but were unlikely to bias the treatment effect. As we did not collect data on usage and engagement with the psycho-education in the control condition, the perceived credibility of the psycho-education remains unknown.

Eighth, this was an open-label trial, where participants and researchers were aware of which group was receiving which treatment. Furthermore, only questionnaire data was assessed as a proxy of use parameters, but no uptake data was available on the actual usage of the program (eg, frequency and length of website usage). Data on how the participants interacted with the program could provide valuable insights into the effectiveness of specific intervention elements.

Conclusions

The Web-based intervention reduced depressive symptoms among adults with sickness absence. As this trial achieved a lower power than calculated, its results should be replicated in a larger sample. Further validation of health insurance records as an outcome measure for eHealth trials is needed.

Acknowledgments

This research was funded by the European Union Innovation Incubator within the European Regional Development Fund. Novego AG is the developer and commercial distributor of the Web-based intervention, “HelpID.” Since July 2016, the copyright of “HelpID” has been owned by IVPNetworks GmbH. In cooperation with the Innovation Incubator, the developer made the intervention available at no cost to the participants in the trial. Funder, developer, and copyright owner had no role in the study design, data analysis, decision to publish, or preparation of the manuscript. KKH health insurance provided access to its insurance members for the purposes of this trial. KKH, the authors, and the funder have no financial interests in the distribution of the intervention. The authors wish to thank the entire team that worked on the study, including Jörn Moock, Sally Kindermann, Kai Kossow, Andreas Maier, Claudia Schmidt, Neele Garbers, Sharlina Spiering, Mathias von Waldenfels, Despina Lion, and Manuela Stockmann.

Conflicts of Interest

This research was funded by the European Union Innovation Incubator, and one goal of the Innovation Incubator was to facilitate regional economic development by evaluating the intervention, HelpID, produced by Novego AG, a regional commercial partner within the Innovation Incubator.

Multimedia Appendix 1

Patient information informed consent form.

[PDF File (Adobe PDF File), 113KB - jmir_v19i6e213_app1.pdf]

Multimedia Appendix 2

Intervention screenshots.

[ZIP File (Zip Archive), 631KB - jmir_v19i6e213_app2.zip]

Multimedia Appendix 3

CONSORT-EHEALTH-checklist.

[PDF File (Adobe PDF File), 806KB - jmir_v19i6e213_app3.pdf]

References

1. Sanderson K, Andrews G. Common mental disorders in the workforce: recent findings from descriptive and social epidemiology. Can J Psychiatry 2006 Feb;51(2):63-75. [doi: 10.1177/070674370605100202]
2. Wittchen H, Jacobi F. Epidemiologie der Depression. In: Stoppe G, editor. Volkskrankheit Depression?: Bestandsaufnahme und Perspektiven (German Edition). Heidelberg: Springer; 2006.
3. Goetzel RZ, Anderson DR, Whitmer RW, Ozminkowski RJ, Dunn RL, Wasserman J, Health Enhancement Research Organization (HERO) Research Committee. The relationship between modifiable health risks and health care expenditures. An analysis of the multi-employer HERO health risk and cost database. J Occup Environ Med 1998 Oct;40(10):843-854. [Medline: 9800168]
4. Welch CA, Czerwinski D, Ghimire B, Bertsimas D. Depression and costs of health care. Psychosomatics 2009;50(4):392-401. [doi: 10.1176/appi.psy.50.4.392] [Medline: 19687180]
5. Sobocki P, Jönsson B, Angst J, Rehnberg C. Cost of depression in Europe. J Ment Health Policy Econ 2006 Jun;9(2):87-98. [Medline: 17007486]
6. Kessler RC, Frank RG. The impact of psychiatric disorders on work loss days. Psychol Med 1997 Jul;27(4):861-873. [Medline: 9234464]
7. Kliner K, Rennert D, Richter M, editors. BKK Gesundheitsatlas 2015: Gesundheit in Regionen - Blickpunkt Psyche. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft und BKK Dachverband e.V; 2015.
8. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. JAMA 2003 Jun 18;289(23):3135-3144. [doi: 10.1001/jama.289.23.3135] [Medline: 12813119]
9. Wang PS, Beck AL, Berglund P, McKenas DK, Pronk NP, Simon GE, et al. Effects of major depression on moment-in-time work performance. Am J Psychiatry 2004 Oct;161(10):1885-1891. [doi: 10.1176/ajp.161.10.1885] [Medline: 15465987]

10. Blonk RW, Brenninkmeijer V, Lagerveld SE, Houtman ILD. Return to work: a comparison of two cognitive behavioural interventions in cases of work-related psychological complaints among the self-employed. Work & Stress 2006 Apr;20(2):129-144. [doi: 10.1080/02678370600856615]
11. Nieuwenhuijsen K, Bültmann U, Neumeyer-Gromen A, Verhoeven AC, Verbeek JH, van der Feltz-Cornelis CM. Interventions to improve occupational health in depressed people. Cochrane Database Syst Rev 2008(2):CD006237. [doi: 10.1002/14651858.CD006237.pub2] [Medline: 18425942]
12. Martin A, Sanderson K, Cocker F. Meta-analysis of the effects of health promotion intervention in the workplace on depression and anxiety symptoms. Scand J Work Environ Health 2009 Jan;35(1):7-18 [FREE Full text] [Medline: 19065280]
13. Kawohl W, Moock J, Heuchert S, Rössler W. Job maintenance by supported employment: an overview of the “supported employment plus” trial. Front Public Health 2015;3:140 [FREE Full text] [doi: 10.3389/fpubh.2015.00140] [Medline: 26075194]
14. Glozier N. Workplace effects of the stigmatization of depression. J Occup Environ Med 1998 Sep;40(9):793-800. [Medline: 9777563]
15. Greenberg PE, Kessler RC, Birnbaum HG, Leong SA, Lowe SW, Berglund PA, et al. The economic burden of depression in the United States: how did it change between 1990 and 2000? J Clin Psychiatry 2003 Dec;64(12):1465-1475. [Medline: 14728109]
16. Mohr DC, Ho J, Duffecy J, Baron KG, Lehman KA, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010 Apr;66(4):394-409 [FREE Full text] [doi: 10.1002/jclp.20659] [Medline: 20127795]
17. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme issue on e-mental health: a growing field in internet research. J Med Internet Res 2010;12(5):e74 [FREE Full text] [doi: 10.2196/jmir.1713] [Medline: 21169177]
18. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
19. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:1-13. [doi: 10.1136/bmj.h5627]
20. Kenter R, van de Ven PM, Cuijpers P, Koole G, Niamat S, Gerrits R. Costs and effects of Internet cognitive behavioral treatment blended with face-to-face treatment: results from a naturalistic study. Internet Interv 2015;2(1):77-83. [doi: 10.1016/j.invent.2015.01.001]
21. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
22. Van't HE, Cuijpers P, Stein DJ. Self-help and internet-guided interventions in depression and anxiety disorders: a systematic review of meta-analyses. CNS Spectr 2009 Feb;14(2 Suppl 3):34-40. [Medline: 19238128]
23. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
24. Kiluk BD, Sugarman DE, Nich C, Gibbons CJ, Martino S, Rounsaville BJ, et al. A methodological analysis of randomized clinical trials of computer-assisted therapies for psychiatric disorders: toward improved standards for an emerging field. Am J Psychiatry 2011 Aug;168(8):790-799 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10101443] [Medline: 21536689]
25. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
26. Buntrock C, Ebert D, Lehr D, Riper H, Smit F, Cuijpers P, et al. Effectiveness of a web-based cognitive behavioural intervention for subthreshold depression: pragmatic randomised controlled trial. Psychother Psychosom 2015;84(6):348-358. [doi: 10.1159/000438673] [Medline: 26398885]
27. Nobis S, Lehr D, Ebert DD, Baumeister H, Snoek F, Riper H, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care 2015 May;38(5):776-783. [doi: 10.2337/dc14-1728] [Medline: 25710923]
28. Ebert DD, Lehr D, Baumeister H, Boß L, Riper H, Cuijpers P, et al. GET.ON Mood Enhancer: efficacy of Internet-based guided self-help compared to psychoeducation for depression: an investigator-blinded randomised controlled trial. Trials 2014;15:39 [FREE Full text] [doi: 10.1186/1745-6215-15-39] [Medline: 24476555]
29. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. Effects of a psychological internet intervention in the treatment of mild to moderate depressive symptoms: results of the EVIDENT study, a randomized controlled trial. Psychother Psychosom 2016;85(4):218-228. [doi: 10.1159/000445355] [Medline: 27230863]
30. Thiart H, Lehr D, Ebert DD, Berking M, Riper H. Log in and breathe out: internet-based recovery training for sleepless employees with work-related strain - results of a randomized controlled trial. Scand J Work Environ Health 2015 Mar;41(2):164-174. [doi: 10.5271/sjweh.3478] [Medline: 25590336]

31. Geraedts AS, Kleiboer AM, Twisk J, Wiezer NM, van MW, Cuijpers P. Long-term results of a web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014;16(7):e168 [FREE Full text] [doi: 10.2196/jmir.3539] [Medline: 25008127]
32. Buntrock C, Ebert DD, Lehr D, Smit F, Riper H, Berking M, et al. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression: a randomized clinical trial. JAMA 2016 May 3;315(17):1854-1863. [doi: 10.1001/jama.2016.4326] [Medline: 27139058]
33. Hensing G, Alexanderson K, Allebeck P, Bjurulf P. How to measure sickness absence? literature review and suggestion of five basic measures. Scand J Soc Med 1998 Jun;26(2):133-144. [Medline: 9658514]
34. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
35. von Waldenfels M, Muth D, Pudack T. Novego. 2011. Online-unterstützungsprogramme von NovegoWochen-programm helpID URL: https://www.novego.de/programme/ [accessed 2016-08-22] [WebCite Cache ID 6jwzh2xt1]
36. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008;31(2):169-177. [doi: 10.1007/s10865-007-9144-1]
37. Kuyken W, Warren F, Taylor R, Whalley B, Crane C, Bondolfi G. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. JAMA psychiatry 2016;73(6):565-574. [doi: 10.1001/jamapsychiatry.2016.0076]
38. Kvillemo P, Brandberg Y, Bränström R. Feasibility and outcomes of an internet-based mindfulness training program: a pilot randomized controlled trial. JMIR Ment Health 2016 Jul 22;3(3):e33 [FREE Full text] [doi: 10.2196/mental.5457] [Medline: 27450466]
39. Schweitzer J, Von Schlippe A. Lehrbuch der systemischen Therapie und Beratung. Göttingen: Vandenhoeck & Ruprecht; 2007.
40. Leitlinien. 2015 Jun. S3-Leitlinie/Nationale VersorgungsLeitlinie: Unipolare Depression - Langfassung URL: http://www.leitlinien.de/mdb/downloads/nvl/depression/archiv/depression-1aufl-vers5-lang.pdf [accessed 2017-06-09] [WebCite Cache ID 6r5M8L6p8]
41. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
42. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
43. Löwe B, Spitzer R, Zipfel S, Herzog W. Gesundheitsfragebogen für Patienten (PHQ-D). Manual und Testunterlagen. Karlsruhe: Pfizer; 2002.
44. Beck A, Steer R, Brown G. Manual for the beck depression inventory-II. San Antonio, TX: Psychological Corporation; 1996.
45. Hautzinger M, Keller F, Kühner C. Beck Depressions Inventar II (BDI 2). Revision. Frankfurt: Harcourt Test Services; 2006.
46. Kühner C, Bürger C, Keller F, Hautzinger M. [Reliability and validity of the revised beck depression inventory (BDI-II). results from German samples]. Nervenarzt 2007 Jun;78(6):651-656. [doi: 10.1007/s00115-006-2098-7] [Medline: 16832698]
47. Priebe S, Huxley P, Knight S, Evans S. Application and results of the manchester short assessment of quality of life (Mansa). Int J Soc Psychiatry 1999 Mar 01;45(1):7-12. [doi: 10.1177/002076409904500102]
48. Björkman T, Svensson B. Quality of life in people with severe mental illness. Reliability and validity of the Manchester short assessment of quality of life (MANSA). Nord J Psychiatry 2005 Jul;59(4):302-306. [doi: 10.1080/08039480500213733] [Medline: 16195135]
49. Moher D, Hopewell S, Schulz K, Montori V, Gotzsche P, Devereaux P, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869. [doi: 10.1136/bmj.c869] [Medline: 20332511]
50. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
51. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]
52. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L. Erlbaum Associates; 1988.
53. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127]
54. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995 Feb 18;310(6977):452-454 [FREE Full text] [Medline: 7873954]
55. OECD. Sick on the job?: myths and realities about mental health and work. Paris, France: OECD Publishing; 2012.
56. Lehr D, Geraedts A, Asplund R, Khadjesari Z, Heber E, Bloom J. Occupational e-mental health: current approaches promising perspectives for promoting mental health in workers. In: Wiencke M, Cacace M, Fischer S, editors. Healthy at Work. Interdisciplinary Perspectives. Switzerland: Springer International Publishing; 2016.

57. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
58. Nystuen P, Hagen KB, Herrin J. Mental health problems as a cause of long-term sick leave in the Norwegian workforce. Scand J Public Health 2001 Sep;29(3):175-182. [Medline: 11680768]
59. Stiles WB, Barkham M, Connell J, Mellor-Clark J. Responsive regulation of treatment duration in routine practice in United Kingdom primary care settings: replication in a larger sample. J Consult Clin Psychol 2008 Apr;76(2):298-305. [doi: 10.1037/0022-006X.76.2.298] [Medline: 18377125]
60. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

Abbreviations

ANCOVA: analysis of covariance
BDI-II: Beck Depression Inventory
ITT: intention-to-treat
KKH: Kaufmännische Krankenkasse
MANSA: Manchester Short Assessment of Quality of Life
PHQ-9: Patient Health Questionnaire
PP: per-protocol
RCI: Reliable Change Index

Edited by G Eysenbach; submitted 25.08.16; peer-reviewed by T Krieger, B Meyer, C Buntrock; comments to author 05.12.16; revised version received 20.01.17; accepted 27.03.17; published 15.06.17.

Please cite as:
Beiwinkel T, Eißing T, Telle NT, Siegmund-Schultze E, Rössler W
Effectiveness of a Web-Based Intervention in Reducing Depression and Sickness Absence: Randomized Controlled Trial
J Med Internet Res 2017;19(6):e213
URL: http://www.jmir.org/2017/6/e213/
doi:10.2196/jmir.6546
PMID:28619701

Review

A Review of the Theoretical Basis, Effects, and Cost Effectiveness of Online Smoking Cessation Interventions in the Netherlands: A Mixed-Methods Approach

Kei Long Cheung[1,2]; Ben Wijnen[1]; Hein de Vries[2]

[1]Department of Health Services Research, Maastricht University, Maastricht, Netherlands

[2]Department of Health Promotion, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Kei Long Cheung
Department of Health Services Research
Maastricht University
Duboisdomein 30
Maastricht,
Netherlands
Phone: 31 43 38 82294
Fax: 31 43 38 84162
Email: kl.cheung@maastrichtuniversity.nl

Abstract

Background: Tobacco smoking is a worldwide public health problem. In 2015, 26.3% of the Dutch population aged 18 years and older smoked, 74.4% of them daily. More and more people have access to the Internet worldwide; approximately 94% of the Dutch population have online access. Internet-based smoking cessation interventions (online cessation interventions) provide an opportunity to tackle the scourge of tobacco.

Objective: The goal of this paper was to provide an overview of online cessation interventions in the Netherlands, while exploring their effectivity, cost effectiveness, and theoretical basis.

Methods: A mixed-methods approach was used to identify Dutch online cessation interventions, using (1) a scientific literature search, (2) a grey literature search, and (3) expert input. For the scientific literature, the Cochrane review was used and updated by two independent researchers (n=651 identified studies), screening titles, abstracts, and then full-text studies between 2013 and 2016 (CENTRAL, MEDLINE, and EMBASE). For the grey literature, the researchers conducted a Google search (n=100 websites), screening for titles and first pages. Including expert input, this resulted in six interventions identified in the scientific literature and 39 interventions via the grey literature. Extracted data included effectiveness, cost effectiveness, theoretical factors, and behavior change techniques used.

Results: Overall, many interventions (45 identified) were offered. Of the 45 that we identified, only six that were included in trials provided data on effectiveness. Four of these were shown to be effective and cost effective. In the scientific literature, 83% (5/6) of these interventions included changing attitudes, providing social support, increasing self-efficacy, motivating smokers to make concrete action plans to prepare their attempts to quit and to cope with challenges, supporting identity change and advising on changing routines, coping, and medication use. In all, 50% (3/6) of the interventions included a reward for abstinence. Interventions identified in the grey literature were less consistent, with inclusion of each theoretical factor ranging from 31% to 67% and of each behavior change technique ranging from 28% to 54%.

Conclusions: Although the Internet may provide the opportunity to offer various smoking cessation programs, the user is left bewildered as far as efficacy is concerned, as most of these data are not available nor offered to the smokers. Clear regulations about the effectiveness of these interventions need to be devised to avoid disappointment and failed quitting attempts. Thus, there is a need for policy regulations to regulate the proliferation of these interventions and to foster their quality in the Netherlands.

(J Med Internet Res 2017;19(6):e230) doi:10.2196/jmir.7209

KEYWORDS

Smoking cessation; telemedicine; review; online intervention; Internet-based intervention; behavioral change techniques; Netherlands

Introduction

Tobacco Smoking

Tobacco smoking is a worldwide public health problem, with more than 5 million deaths being attributable to direct smoking [1]. Nonsmokers are also impacted by tobacco smoke as a result of second-hand smoke (passive smoking); an estimated 600,000 deaths are caused by smoking behavior, which affects various related diseases such as lung cancer, heart diseases, and chronic obstructive pulmonary disease. In the Netherlands, smoking is a public health problem. In 2015, 26.3% of the Dutch population aged 18 years and older smoked, 74.4% of them daily [2]. Of all smokers, 15.6% were considered heavy smokers, meaning that these individuals smoke an average of 20 or more cigarettes per day [2]. Estimates showed that in 2013, 19,000 deaths were attributable to smoking-related diseases in the Netherlands [3].

Online Smoking Cessation Interventions

In an attempt to tackle the scourge of tobacco smoking, many smoking cessation interventions have been developed and proven effective, especially the more intensive interventions such as one-to-one behavioral therapy [4]. Each intervention has strengths and limitations. The more intensive interventions are often expensive, inconvenient to the recipient (eg, waiting list and the need to take time off work), and reach only a small proportion of smokers [5]. The Internet provides opportunities to address the smoking problem, as the Internet has grown to be an extremely important medium and is embedded in daily life in the Dutch population (and many other parts of the world). The opportunities provided by the Internet to enhance smoking cessation have led to the development of Internet-based interventions. Internet-based smoking cessation interventions (hereafter: online cessation interventions) are relatively new innovations and, due to the low costs per smoker, the accessibility (eg, home, work, and public access points), and 24-hour a day availability, have the potential to reach a large proportion of smokers [5]. Internet access is increasing worldwide. In 2016, approximately 94.4% of the Dutch population had access [6] (98.9% of those aged 25 to 45 years, 98.3% aged 45 to 65 years, and 77.6% aged 65 years or older [7]). The level of access for the younger population, aged between 12 and 25 years, was 99.1% in 2015 [7]. Moreover, individuals having access to the Internet in the Netherlands range from 87.0% for the lower educated to 99.5% for the higher educated [7]. However, reach may be a too narrow construct to reflect Internet usage, as activities in use differ [8]. Educational level might, for instance, play a role in online behavior because many individuals may not be sufficiently literate to understand the high literacy level of most written information on the Internet [9]. Yet, a recent study in the Netherlands indicated that usage of eHealth interventions, as recommended, did not differ among education levels. The study also found that eHealth interventions are more often used as recommended to people with a greater need for health care information [10]. As well as Internet access, online cessation interventions allow other opportunities, such as being interactive and tailoring messages to individuals, which has been shown to be effective in changing health behaviors such as enhancing smoking cessation [11]. In addition, due to the low costs per person, online cessation interventions can be cost effective [12]. Therefore, it is relevant to explore the online cessation interventions that are available in the Netherlands and investigate the extent to which these are effective and cost effective.

Cost effectiveness is explored in economic evaluation studies that determine the costs and effects associated with an intervention and compare these with the costs and effects of other interventions or current practice [13]. A typical economic evaluation consists of five steps: (1) identification of relevant costs and effects, (2) measurement of costs and effects, (3) valuation of measured costs and effects, (4) calculation of cost-effectiveness ratio, and (5) sensitivity analysis [14].

Evaluating Content: Theoretical Basis

Because many online cessation interventions may not have been tested in a randomized controlled trial (RCT), scholars have become interested in evaluating relevant characteristics of behavioral interventions, such as online cessation interventions. The literature shows that the theoretical basis of an intervention and behavior change techniques (BCTs) are characteristics that may influence the impact of behavior [15-17]. Three theories—the social cognitive theory [18], the transtheoretical model [19], and the theory of planned behavior [20]—were particularly important in developing the intervention; their usage in intervention development was associated with increases in effect size [21]. Various sociocognitive theories, including the three mentioned, are integrated in the Integrated Change (I-Change) Model [22-25]. This explains adoption of health behavior in (at least) three phases (ie, awareness, motivation, and action phase), each of which has phase-specific determinants, such as attitudes, social support, and self-efficacy for understanding motivation, and action planning and coping plans to understand the final step from intentions to behavior. Another way to evaluate the content of an intervention is by assessing whether or not the online cessation interventions use specific BCTs that are associated with higher success rates in smoking cessation [26-28]. Five BCTs that are applicable to online interventions and mobile-based interventions have been found to be associated with higher cessation rates [26].These are (1) strengthening exsmoker identity (eg, clear boundaries offered), (2) providing rewards contingent on stopping successfully (eg, use praise or advise on rewarding oneself for moving toward the goal of becoming an exsmoker), (3) advising on changing routines (eg, advising on avoiding cues that will trigger strong urges to smoke), (4) advising and assisting with ways of coping with urges to smoke (eg, develop effective ways of distracting attention from smoking cues in the environment and from urges to smoke), and (5) advising on use of stop-smoking medications (eg, explain the medication available, its benefits, and promote use). The theoretical underpinnings (theoretical factors and BCTs) of interventions, especially those that are not tested in RCTs, may thus provide insights into the content of online cessation interventions (ie, specifying interventions based on relevant theories and BCTs used).

Objective

Web-based interventions are usually developed for a specific platform, made accessible via computers, tablets, or mobile phones. Not all may be optimized for all platforms. A distinction is made between online interventions (accessed via computer), referred to as eHealth, and mobile-based interventions (apps), referred to as mHealth [29]. Online cessation interventions in this study are defined as online interventions—developed to be accessed via a computer—aimed at enhancing smoking cessation among individual smokers. To explore a broader range of online cessation interventions, combinations of an online intervention and a counselor were deemed relevant. This led to the study goal: to provide an overview of online cessation interventions in the Netherlands and explore their effectivity, cost effectiveness, and theoretical basis.

Methods

This study used a mixed-methods approach, including (1) a scientific literature search, (2) a grey literature search, and (3) expert input. The methods and the analysis are detailed subsequently.

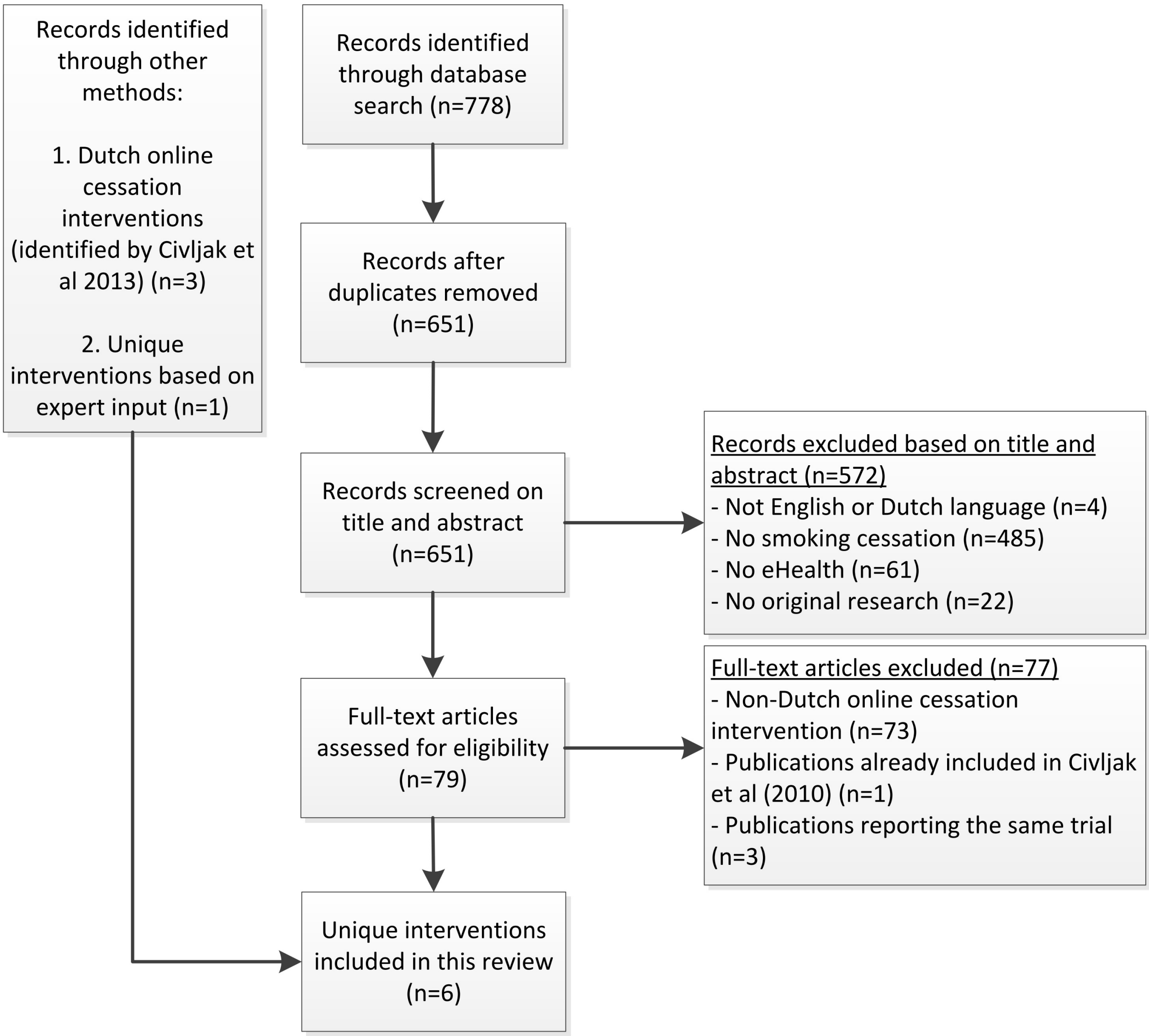
Scientific Literature Search

In the search for online cessation interventions, we used a systematic review of Cochrane [5] regarding the effectiveness of Internet-based interventions, to screen for the Dutch interventions. Because this review was last updated in April 2013, we conducted an additional systematic literature search to explore recent online cessation interventions in the Netherlands up to July 2016. The selection process is depicted in a flowchart (see Figure 1).

The Cochrane review included randomized or quasi-RCTs with smokers that participated in online cessation interventions [5]. All types of online cessation intervention were included; there was no exclusion with respect to method or duration. Also, combinations of interventions were included (when the Internet component was subject to evaluation), if the Internet intervention was an adjunct to behavioral therapy or pharmacotherapy. Trials were excluded if they used Internet solely for recruitment or as a reminder of offline appointments (eg, face-to-face therapy or pharmacotherapy). Moreover, text messaging interventions and mobile-based interventions were not covered. Also, trials with fewer than four weeks of follow-up were excluded.

Similar to the Cochrane review [5], the search strategy for the update of literature was based on the specialized register of the Cochrane Tobacco Addiction Group, including the terms “Internet” or “www” or “web” or “net” or “online” in the title, abstract, or as keywords since 2013. Databases of Cochrane Central Register of Controlled trials (CENTRAL), MEDLINE, and EMBASE were searched via OVID. For full search strategies, see the Tobacco Addiction Group Module [30]. The search led to a total of 651 papers (CENTRAL=96, MEDLINE=163, and EMBASE=392), which two researchers (KLC and BW) screened independently for title and abstract. Inclusion criteria were (1) English or Dutch language, (2) smoking cessation intervention, (3) eHealth (ie, Internet or computer), and (4) original research (excluding study protocols and conference abstracts). Interventions directed at indirect populations (eg, clinicians and nurses) were excluded, as well as lifestyle, telephone, and prevention interventions. The two researchers resolved disagreements through discussion, resulting in 79 potentially relevant papers. The full-text papers were screened by KLC and BW, resulting in exclusion of 10 additional papers. For the relevant Dutch online cessation interventions, the researchers searched through the full-text papers and the Cochrane review [5]. Excluding all non-Dutch intervention studies, the Cochrane review yielded three relevant papers, whereas the updated search (ie, between 2013 and July 2016) yielded six papers. The updated literature search had six papers on Dutch online cessation interventions, from which one was a cost-effectiveness study aiding a study included in the Cochrane review, and four papers were studies on the same intervention (with the same lead author). Due to expert input (described in “Expert Input” section), one intervention was added to the list. Hence, this search led to a final list of six unique online cessation interventions in the Netherlands that were reported in the scientific literature.

Figure 1. Flowchart of the selection process in the scientific literature.



Grey Literature Search

Most online cessation interventions in the Netherlands were not described in peer-reviewed journals. Hence, a Google search was conducted to capture interventions in the grey literature. In line with a previous Dutch study on identifying online tobacco control methods [31], we conducted a Google search to identify actual interventions rather than a search strategy aiming for identifying documents. This is now elaborated in the text. The free-text terms *stoppen met roken* ("smoking cessation") and *online roken* ("online smoking") were used, with the intention to capture a broad range of online cessation interventions. For each of the text terms, the first five pages (50 hits) of Google were explored, yielding 100 results (September 25, 2016). These were then screened by two researchers together (KLC and BW) looking at the title (of the page) and content of the site (for potential online cessation interventions), leading to 47 results. Lifestyle interventions, websites targeting indirect groups (eg, clinicians), or telephone-only interventions were excluded. These websites were then screened independently (KLC and BW) for content, including online cessation interventions in the Dutch language. For each website where there was disagreement, consensus was reached via discussion. Here we included a broad range of interventions, including informative static websites and websites promoting any intervention in online form (eg, online counseling). Websites that solely referred to other websites were excluded, as were standalone apps. Six interventions were added due to expert input (described in "Expert Input" section). This led to a final list of 39 online cessation interventions in the Netherlands that were reported in the grey literature.

Expert Input

To check and complement the previously mentioned searches, Dutch experts in smoking cessation (eg, smoking cessation research and/or online smoking cessation interventions) were recruited using a convenience sampling strategy, in which researchers created a list of potential participants based on personal networks. Twenty experts were approached via email, outlining the goal of the study and asked whether they were able to provide input. Five (25% response rate) provided input for the questionnaire. We asked experts to list the online cessation interventions they knew, adding to the list from the literature search. They provided a list of 21 interventions in total, with seven unique interventions added to the final list (compiled from the literature and grey literature search). One online cessation intervention was added to the scientific literature search, and six to the grey literature search.

Data Extraction and Analysis

For each included online cessation intervention, two independent reviewers (KLC and BW) systematically extracted data using a predefined extraction form that included general characteristics, four theoretical components (ie, attitudes, social support, self-efficacy, and action planning and coping plans), and five BCTs (ie, strengthening exsmoker identity, providing rewards contingent on stopping successfully, advising on changing routines, advising and assisting with ways of coping with urges to smoke, and advising on use of stop-smoking medications) for the content analysis. Two separate extraction forms were developed (one for scientific literature and one for grey literature), based on literature (theoretical underpinnings) and discussions in the research team. The two reviewers piloted and discussed the extraction form for two interventions, which led to minor adjustments to the data extraction forms.

The extracted data for the interventions identified in the scientific literature included four categories: general study and intervention information; effectiveness, cost effectiveness, and outcome information; theoretical factors; and BCTs. General study and intervention information included intervention name, brief description, and target population. Effectiveness, cost effectiveness, and outcome information included effectiveness (yes, no, and not reported [NR]), effect (in relative risk [RR]), number and percentage abstinent in the intervention condition and in the control condition, control group, outcome measure (for smoking cessation), and cost effectiveness (yes, no, NR). Theoretical factors included were based on the I-Change Model and assessed the presence or absence of the core factors for motivation and behavioral change (yes, no): attitude (discussing advantages and disadvantages of smoking), social influence (addressing social influence or social support), self-efficacy (enhancing one’s capability to become an exsmoker), and action/coping planning (supporting creation or advising on action or coping plans). Literature concerning BCT revealed a large range of potential BCTs that could be selected for reviewing the efficacy of interventions, amounting to 40 different BCTs [21]. Therefore, five BCTs that were especially applicable and relevant to online interventions and mobile-based interventions were included [26]. Consequently, the BCTs included the five BCTs discussed (yes, no): supporting identity change, rewarding abstinence, advising on changing routines, advising on coping, and advising on medication use.

The extracted data for the interventions identified in the grey literature included general intervention information, theoretical factors, and BCTs. General intervention information included intervention name, provider, and target population. The same theoretical factors and BCTs were extracted as for the scientific literature. Websites with only a page of static information about smoking cessation were grouped in the results.

Data were independently extracted by the two researchers and the first author then checked and compared the extracted data. Any disagreements between the researchers were resolved through discussion until consensus was reached [5,26]. Extracted data are presented in tables. Whenever an intervention was shown to be cost effective, this is elaborated in the results section. Information was extracted from the literature and, when available, the researchers checked the online cessation intervention. Suggested interventions by experts are indicated in the tables.

Results

The mixed-methods yielded 45 unique online cessation interventions (scientific literature: n=6; grey literature: n=39).

Scientific Literature

The literature revealed six interventions in the Netherlands, which typically targeted adult smokers, who were motivated to quit smoking (thus willing to set a quitting date) (see Table 1). These interventions typically involved tailoring health messages based on the I-Change Model. The exception was the Web-based Attentional Bias Modification (ABM) self-help intervention [32], which aimed to reduce attentional bias and generalize to influence an approach bias for cigarettes and success in quitting. The effectiveness, cost effectiveness, and outcome information of these interventions are detailed in Table 2.

Table 1. General study and intervention details of smoking cessation interventions in the scientific literature (n=6).

Intervention	Target population	Brief description
Quit Smoking 2.0 [33]	Smokers (≥18 years, smoker of cigarettes and/or loose-cut tobacco and intending to quit within a year or sooner)	One 7-to-9 page computer-tailored email letter, based on I-Change Model, addressing motivational and postmotivational factors.
Stay Quit For You (SQ4U) [34,35]	Smokers 18-65 years, smoked daily, willing to set a quit date within 1 month, and motivated to quit smoking	The action planning (AP) program provided tailored feedback at baseline and invited respondents to do 6 preparatory and coping planning assignments (the first 3 assignments prior to quit date and the final 3 assignments after quit date). The action planning plus (AP+) program was an extended version of the AP program that also provided tailored feedback at 11 time points after the quit attempt.
Personal Advice in Stopping smoking (PAS) [22,36]	Smokers (Dutch adults) with intention to stop smoking within 6 months	A Web-based computer-tailored smoking cessation intervention with 4 sessions, based on I-Change Model.
Smoke Alert[a] [37]	Nonsmoker or former smoker; 10 and 20 years; having computer/Internet literacy; having sufficient command of Dutch; no previous exposure to the earlier version of Smoke Alert	Web-based computer-tailored smoking prevention and cessation intervention for adolescents. Based on an intervention shown to be effective [33]. Website intervention providing one-time feedback after filling in a questionnaire.
Support to Quit (STQ) [38-41]	Smokers motivated to stop smoking and aged 18 years	Web-based text and a Web-based video-driven computer-tailored approach for low and high SES smokers, this incorporates 3 to 6 computer-tailored feedback moments with the aim to support smoking cessation.
Attentional Bias Modification (ABM)[a] [32]	Smokers aged 18-65 years, reporting smoking on a daily basis for at least 1 year and not having made a quit attempt yet	Online 6 sessions of ABM training with the aim to reduce attentional bias and induce more distant generalization effects regarding approach bias and success in quitting.

[a] Currently available on the Internet.

Table 2. Effectiveness, cost effectiveness, and outcome information of smoking cessation interventions in the scientific literature (n=6)

Intervention	Effectiveness	RR (95% CI)	Percentage abstinent, n (%)[a]	Control group	Outcome measure	CE[b]
Quit Smoking 2.0 [33]	Yes	2.48 (1.11-5.55)	Int: 224 (8.5), Con: 234 (3.4)	Nontailored email	7-day point prevalence abstinence at 6 months follow-up	NR
SQ4U [34,35]	Yes	AP+: 1.37 (0.97-1.92); AP: 1.49 (1.07-2.06)	Int (AP+): 53 (7.6), Con: 45 (7.1); Int (AP): 63 (9.0), Con: 45 (7.1)	Questionnaires only	Continued abstinence at 12 months follow-up	NR
PAS [22,36]	Yes	1.50 (0.77-2.94)	Int: 20 (15.2); Con: 12 (10.1)	Usual care	Prolonged abstinence at 12-month follow-up	Yes
Smoke Alert[c] [37]	NR	NA	NA	NA	NA	NR
STQ [38-41]	Yes	Video: 1.54 (1.08-2.22); text: 1.15 (0.78-1.69)	Int (video): 66 (9.9), Con: 46 (6.4); Int (text): 52 (7.3), Con: 46 (6.2)	Nontailored generic advice	Prolonged abstinence at 12-month follow-up	Yes
ABM[c] [32]	No	NR, for subgroup: 2.33 (1.14-4.76)	Int (subgroup): 22 (14.2); Con: 10 (6.1)	Placebo-training, continued assessments	Continued abstinence at 6 months follow-up	NR

[a] Percentage abstinent: n (%) in the intervention condition (Int) and in the control condition (Con).

[b] CE: Cost effectiveness (yes, no, NR=not reported). If yes, this is detailed in the text.

[c] Shown effective in Te Poel et al [33].

[d] Shown effective for subgroup heavy smokers [37].

The effectiveness was shown in four interventions. For Smoke Alert (2.0) [37], the study did not investigate the effectiveness of smoking cessation, but rather investigated and found effects on prevention. Yet, a previous version did show effects on smoking cessation with 17.2% quitters in the control group and 26.8% in the online computer-tailored condition ($P<.03$) [42,43]. Results indicate no overall effectiveness evidence for the ABM intervention except only for heavy smokers. The effective interventions incorporated health messages tailored to responses in an initial questionnaire, which was based on the I-Change Model. Long-term effects were shown in Stay Quit For You (SQ4U) [34,35], Personal Advice in Stopping smoking (PAS) [22,36], and Support to Quit (STQ) [38-41], using a 12-month follow-up with either continued or prolonged abstinence outcome measure. For these interventions with long-term follow-ups, we conducted a meta-analysis in which the Mantel-Haenszel fixed-effect model was used to pool the effectiveness of these interventions. This resulted in a pooled RR of 1.39 (95% CI 1.18-1.65). For two interventions (PAS and STQ), studies also found support for its cost effectiveness from a societal perspective. Using prolonged abstinence as the outcome measure, the cost-effectiveness study for PAS showed that the mere multiple computer-tailored program had the highest probability of being cost effective as in this group €5100 had to be paid for each additional abstinent participant (compared to usual care) [22]. Using prolonged abstinence, the study for STQ indicated that with a willingness to pay €1500, the video-based intervention was likely to be the most cost-effective treatment. For each additional abstinent participant, €1500 had to be paid within the video-based intervention [38]. Given the fact that, in the Netherlands, cost effectiveness is concluded when the additional costs (for an additional quality-adjusted life year [QALY]) are between €20,000 and €80,000, the conclusion is that these interventions were highly cost effective [44]. Normally, this threshold is dependent on the severity of the disease. In prevention, a threshold of €20,000 is often considered [45].

The interventions were also evaluated and analyzed with regard to the theoretical factors and BCTs used (see Table 3). For the theoretical factors, all interventions (with one exception) addressed attitudes, social influence, self-efficacy, and action or coping planning (83%, 5/6). The exception was the ABM intervention [32], which had a different theoretical basis, namely retraining implicit associations (automatic attentional processes). Attentional bias is the tendency of certain cues in a person's environment to attract and/or hold the individual's attention preferentially, such as cigarettes. This attentional bias could be retrained toward or away from the substance-related cues [32]. The SQ4U [34,35] and STQ [38-41] interventions included all BCTs, whereas half of the interventions included all BCTs, except rewarding abstinence (50%, 3/6).

Table 3. Theoretical factors and behavior change techniques (BCTs) of smoking cessation interventions in the scientific literature (n=6).

Intervention	Attitude	Social influence	Self-efficacy	Action & coping planning	BCTs[a]				
					1	2	3	4	5
SQ4U[b] [34,35]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PAS[b] [22,36]	Yes	Yes	Yes	Yes	Yes	NR	Yes	Yes	Yes
Smoke Alert [37]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quit Smoking 2.0[b] [33]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
STQ[b] [38-41]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ABM[b] [32]	No	No	No	No	No	No	No	No	No
Percentage; n (%)[c]	5 (83)	5 (83)	5 (83)	5 (83)	5 (83)	3 (50)	5 (83)	5 (83)	5 (83)

[a] BCT1: Supporting identity change; BCT2: rewarding abstinence; BCT3: advising on changing routines; BCT4: advising on coping; and BCT5: advising on medication use.

[b] Mentioned by experts.

[e] Percentage (%) interventions of scientific literature including this factor/BCT.

Grey Literature

The grey literature revealed 39 interventions in the Netherlands, which typically were targeted at smokers (see Table 4). The online cessation interventions were grouped as websites providing only static information (n=23) and websites incorporating an interactive component (n=17). Seven unique interventions were added due to expert input (see Tables 3 and 4). Some experts indicated that it was difficult to evaluate the effectiveness of the online cessation interventions, with the exception of a few reported in scientific literature. The grey literature yielded a variety of online cessation interventions with different providers, but only minimal research has been conducted to test their effects, resulting in little scientific support for their evidence. One intervention was identified in both the grey literature and in the scientific literature (Smoke Alert [37], which—as outlined previously—was effective in preventing the initiation of smoking among adolescents [37]). An intervention (Online zelfhulp tabak [46]) identified in the grey literature search and via expert input was reported to be effective; it was to be tested in a RCT. After contacting the provider of the intervention, it was revealed that the claim on effectiveness was based on a trial of a similar intervention on alcohol [47]. It was stated that ‘’due to the generalizability of the self-help module, the self-help tobacco module was recognized as well.” Hence, the quality of the interventions identified in the grey literature search seems to lack scientific basis because the number of effectiveness studies is limited.

Table 4. General intervention information of grey literature (translation in brackets).

Intervention	Provider	Target population	Static?
De StopSite (The QuitSite)[a]	Luchtsignaal	Smokers aged ≥18 years	Interactive
uQuit.nl[a]	Universiteit Nijmegen, VU Amsterdam, and IVO	Smokers / student smokers	Interactive
Tabakstop (Tobaccostop)	Stichting tegen kanker	Smokers	Interactive
ExSmokers (iCoach)[a]	European Commission	Smokers	Interactive
Stoppen met roken (Smoking cessation)	Pfizer	Smokers	Interactive
Roken de Baas (Boss of your smoking)[a]	Tactus Verslavingszorg	Smokers	Interactive
CZ Stoppen met roken coach (CZ smoking cessation coach)[a]	CZ	Smokers	Interactive
Stoppen met roken (Smoking cessation)[a]	Minddistrict	Smokers	Interactive
StopExpert[a]	ExaCare	Smokers	Interactive
Online zelfhulp tabak (Online self-help tobacco)[a]	Jellinek	Smokers	Interactive
Stoppen met roken, in één dag van het roken af (Smoking cessation, quit in one day)[a]	De opluchting	Smokers	Interactive
Wat doe je om te stoppen met roken? De PZP helpt (What do you do to quit smoking? PZP helps)	PZP	Smokers with an insurance for the police (PZP)	Interactive
Home Roken—Ja (Home Smoking—Yes)	Victas	Smokers	Interactive
Training stoppen met roken—Kentra (Training smoking cessation—Kentra)	Kentra	Smokers	Interactive
stoppen met roken (smoking cessation) | iLifecoach	iLifeCoach	Smokers	Interactive
Online cursus stoppen met roken | Zo stop je wel (Online course smoking cessation | You’ll quit like this)	Zostopjewel.nl	Smokers	Interactive
Stoppen met roken (Smoking cessation)	Gezondheidsnet.nl	Smokers	Static
ikstopnu.nl (Iquitnow.nl)[a]	Ikstopnu.nl	Smokers	Static
Ik stop! (I quit!)	Ikstop.nl	Smokers	Static
NuStoppenmetRoken.nl-Stoppen met Roken (Quitsmokingnow.nl—Smoking cessation)	NuStoppenmetRoken.nl	Smokers	Static
Stoppen met roken (Smoking cessation)	Nederlands Huisartsengenootschap	Smokers	Static
Stoppen met roken (Smoking cessation)	Medical Media BV	Smokers	Static
Hoe kan ik stoppen met roken? (How can I quit smoking?)	Stichting Opvoeden.nl	Smokers	Static
Rokeninfo.nl (Smokinginfo.nl)[a]	Trimbos Instituut	Smokers	Static
Ex rokers (Former smokers)	Ex rokers	Smokers	Static
Tips stoppen met roken (Tips to quit smoking)	Christelijke Mutualiteit	Smokers	Static
Hoe kan ik stoppen met roken? (How to quit smoking?)	GGD	Smokers	Static
Ik Wil Stoppen Met Roken.NU (I Want To Quit Smoking.NOW)[a]	Ik Wil Stoppen Met Roken.NU	Smokers	Static
Welkom bij de stoppen met roken test! (Welcome to the smoking cessation test)[a]	ProStop	Smokers	Static
soChicken	soChicken	Smokers	Static
Waarom stoppen met roken? (Why quit smoking?)—Watchtower ONLINE LIBRARY	Watchtower	Jehovah’s Witnesses	Static
Stoppen met Roken (Smoking cessation)	Verslaving.nu	Smokers	Static
Stoppen met Roken? (Quit smoking?)	Dokteronline.com	Smokers	Static
Stoptober[a]	Stoptober[b]	Smokers	Static

Intervention	Provider	Target population	Static?
Stoppen met Roken.nl (Smoking Cessation.nl)	Stichting stop bewust	Smokers	Static
Stoppen met roken-GGD Fryslan (Smoking cessation—CHS of Fryslan)	GGD	Smokers	Static
Welkom bij nl.support.stop-met-roken (Welcome to nl.support smoking cessation)	NSSMR	Smokers	Static
Stoppen met roken (Smoking cessation)—YouTube	Stichting Gezondheid	Smokers	Static
Eenrookvrijleven.nl (Smokefreelife.nl)[a]	Eric Eraly	Smokers	Static

[a] Mentioned by experts.

[b] Stoptober: KWF Kankerbestrijding, Hartstichting, Longfonds, Ministerie van Volksgezondheid, Welzijn en Sport (VWS), Trimbos Instituut, GGD GHOR Nederland, and Alliantie Nederland Rookvrij.

The interventions were extracted on theoretical factors and BCTs (see Multimedia Appendixes 1 and 2) with the exception of those that were not accessible to the researchers. This was especially the case for interactive interventions, with tailored messages or combined with feedback from a counselor. For those that could be evaluated, the results show that at least 67% (26/39) addressed attitude, mentioning the advantages and disadvantages of smoking cessation. Many interventions also addressed social influence and action or coping planning (at least 44%, 17/39 and 51%, 20/39, respectively), by advising on social support, planning to continue to abstain from smoking, or dealing with difficult situations. At least 31% (12/39) of the interventions addressed the self-efficacy of smokers by persuasion or via modeling, enhancing beliefs that the smoker is able to become an exsmoker. Regarding BCTs, most online cessation interventions provided information and advice on the importance of changing routines (at least 49%, 19/39) and medication usage (at least 54%, 21/39). A moderate number of interventions supported identity change (at least 28 %, 11/39), provided tips or provided rewards for abstinence (at least 38%, 15/39), and advised on coping strategies (at least 44%, 17/39).

Discussion

This study provides an overview of online cessation interventions in the Netherlands, up to 2016. They are broadly categorized in two lists: interventions reported in the scientific literature and those identified in the grey literature. Expert input overlapped with the lists and added a few unique interventions.

Summary of Evidence

The first goal of this study was to assess the effectiveness of online interventions concerning smoking cessation as defined by the most strictly reported outcome measure. Our search revealed that six online cessation interventions in the Netherlands were investigated for their effectiveness in trials. With one exception, these were all interactive individually tailored interventions based on sociocognitive models (ie, the I-Change Model). Smokers filled in questionnaires dealing with demographics, smoking behavior, and sociocognitive constructs. The answers were used to yield tailored motivational messages. The exception was the ABM intervention, which focused on reducing attentional bias and inducing more distant generalization effects regarding approach bias and success in quitting [32]. The effects indicated that smokers using an online cessation intervention are 1.15 to 2.84 times more likely to become a former smoker compared to the control condition. This range and the meta-analysis (with a pooled RR 1.39, 95% CI 1.18-1.65) seem in line with the Cochrane review with a pooled RR (Mantel-Haenszel fixed-effect) of 1.41 for computer-tailored online smoking cessation interventions (95% CI 1.11-1.78) [5]. These reported effects of online smoking cessation interventions may perhaps be underestimated because conservative analyses were used, including penalized imputation which categorizes missing data as smoking. A recent study showed that penalized imputation biases the reported effects of online smoking cessation interventions, favoring the condition with the lowest proportion of missing data [48]. Interestingly, many more interventions were identified using grey literature; yet, none of them were evaluated in terms of their effectiveness.

A second goal was to identify the cost effectiveness of online cessation interventions in the Netherlands. Of the six interventions identified, two were tested and shown to be cost effective, meaning that the costs for an additional quitter (or QALY) did not surpass a conservative threshold of €20,000. PAS was cost effective, with costs of €5100 for each additional quitter [22], and STQ with its tailored video-based intervention was cost effective with costs of €1500 per additional quitter [38]. When the additional costs for an additional QALY is between €20,000 and €80,000, an intervention is considered to be cost effective in the Netherlands. Normally, this threshold is dependent on the severity of the disease. In prevention, often a threshold of €20,000 is considered [45]. There is, however, no cut-off point with regard to smoking abstinence rates as outcome, which suggests future research should identify an acceptable cut-off point for the willingness to pay per abstinent participant [49]. Extracting data on theoretical components and BCTs revealed that these I-Change Model-based interventions all addressed attitudes, social influence, self-efficacy, and action or coping planning. It also revealed that these interventions support identity change, and advise on changing routines and on coping. Because not all interventions were available to the public or researchers, there is some uncertainty about the percentage interventions rewarding abstinence and advising on medication. Other interventions that have potential to be implemented are SQ4U (scoring high on all BCTs and shown to be effective) [34] and Quit Smoking 2.0 (shown to be effective and recommended based on quality, practicability, and effectiveness on loketgezondheid.nl) [33]. However, data concerning their cost effectiveness are lacking, and for Quit Smoking 2.0 the follow-up is 6 months only [33]. Hence, because a follow-up of 12 months indicates long-term effects, SQ4U [34,35] and especially PAS [22,36] and STQ [38-41] appear to be promising interventions for implementation in the Netherlands. The effective interventions (ie, Quit Smoking 2.0, SQ4U, PAS, and STQ), are no longer publicly available because they were part of a research study. Yet, if organizations are interested, they can be obtained via Vision2Health (a health consultancy initiative offering evidence-based health promotion interventions). For instance, Quit Smoking 2.0 has been offered by several regional health authorities on request. The youth intervention Smoke Alert is still online because it is offered by Trimbos and has received approximately 13,000 unique visitors. Interestingly, it was shown that such online interventions did not differ in usage as recommended by individuals with different educational levels [10]. This indicates the potential to be effectively implemented for smokers with different levels of education [50].

A third goal was to assess the theoretical underpinnings of the interventions. The majority of the interventions that were also evaluated for their effectiveness and cost effectiveness used most, if not all, of the theoretical factors and the five BCTs. The interventions identified from the grey literature also used some of these factors and techniques. Some appeared to have included several, from which one might conclude that they are more likely to be successful. They could be candidates for testing in RCTs. But the theoretical factors and BCTs have to be applied according to parameters, which impact on the degree of effectiveness [51].

In conclusion, it would be beneficial to many smokers, and even cost effective from a societal perspective, if effective and cost-effective interventions were to become available to the public after completion of a RCT study. For public health impact, it is important that these evidence-based interventions are not only available online, with no marketing (passive availability), but that they be promoted proactively, involving marketing to various stakeholders within health care (eg, general practitioners, dental practices, and other primary health care providers) and public health organizations (eg, regional health education authorities). Moreover, it is important for universities and research institutes to consider implementation of the intervention and to create a business model [52]. Researchers need to include this step, enhancing availability and usage of the developed intervention. Furthermore, more funding is needed to implement interventions that are proven to be effective and cost effective, in order to enhance usage and availability of those already online, such as PAS [22,36] and STQ [38-41]. In addition, future research may also look at the cost effectiveness to bridge the gap between research and practice because cost-effectiveness studies provide a financial argument for investment in effective interventions, such as SQ4U [34,35]. Many interventions were identified in the grey literature, more than in the scientific literature. Some may be promising if they address theoretical factors and BCTs, but they lack scientific support for effectiveness. This may be an argument for introducing a quality mark to reflect the extent of the scientific evidence. Providers could then submit their intervention for evaluation. Such a mark is important for health care stakeholders and smokers to identify quality online cessation interventions.

Limitations

This study has several limitations. One is the lack of an evaluation tool to investigate the quality or effectiveness of nontested reports identified in the grey literature. To evaluate the content, we assessed whether the interventions implemented relevant theoretical components and BCTs, identified as being associated with smoking cessation [21,26-28]. However, BCTs have limitations because they simplify reality [51]. They are generic and their application varies. Current practice often does not recognize that these BCTs have parameters for effectiveness and that methods can interact with one another. Context factors may also impact the effectiveness of BCTs [51]. Hence, the evaluation of whether interventions include BCTs may result in overly simplistic or even incorrect conclusions about the quality of the identified interventions. As well as the evaluation of the content via BCTs, the delivery, such as engagement and ease-of-use features, is important [26]. Delivery evaluation was not included in this study. Similarly, previously mentioned criticism of BCTs also applies with regard to the inclusion of theoretical factors in the interventions. This study is cautious in its recommendations based on the inclusion of theoretical factors and BCTs. Another limitation of the study may be the limited generalizability of results to other countries. Findings reflect the online cessation interventions in the Netherlands; other countries may have implemented and tested different interventions. It would, therefore, be interesting to conduct similar studies in other countries and compare potential differences. Moreover, no risk of bias assessment (eg, a quality checklist) was conducted, which may have over- or underemphasized the strength of the evidence of some studies when calculating the pooled effect. Yet, our estimates were in line with others reported in, for example, the Cochrane review on online cessation interventions [5]. Nevertheless, study results should be interpreted with some caution.

The usage of mobile phones in everyday life is becoming increasingly important, indicating that online cessation interventions (developed to be accessed via a computer) may also need to function on a mobile platform. In the most recent Cochrane review on mobile-based smoking cessation interventions, only five studies were included [53]. None of them was developed for the Dutch population; results were heterogeneous with three of five interventions crossing the line of no effects. Hence, more research is required to explore opportunities for mobile-based interventions by investigating effectiveness and cost effectiveness. The lack of Dutch mobile phone-based interventions—tested for effectiveness—indicates the need to transfer the effective online cessation interventions to the mobile phone. Whereas mHealth interventions may have great potential, a recent Dutch study found that eHealth was more effective in realizing physical activity [54]. This does not imply that mHealth cannot be effective, but rather that we need to identify how to use the two modalities optimally. The potential of mHealth, as well as issues such as effectiveness, cost effectiveness, and usability [55,56], should be topics for future research, as well as exploring ways of improving computer tailoring of health messages (via different algorithms) [57].

Conclusions

This study provides an overview of Dutch online cessation interventions, while assessing their effectivity, cost effectivity, and theoretical basis. This mixed-methods study may also serve as a vantage point for future overviews in other countries. Although the Internet may well provide the possibility of offering various smoking cessation interventions, the user in the Netherlands is left bewildered about its efficacy because most of the data are not available or are not offered to the smokers. If the Internet wants to live up to its promise, clear regulations governing effectiveness of interventions have to be devised to avoid disappointment and failed quit attempts. Policy regulations may be needed to regulate the proliferation of interventions and foster quality. As nonadoption of cost-effective eHealth interventions is both detrimental at the micro level (smokers not profiting from effective interventions) as well as the macro level (unnecessary high costs of smoking due to nonimplementation of effective methods), reasons for nonadoption, as well as strategies enhancing such an adoption, are a prerequisite to ensure a significant public health impact of effective eHealth and mHealth interventions.

Acknowledgments

The study was funded by Miro. The funding agency had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Authors' Contributions

KLC planned and managed the work, analyzed and interpreted results, and produced the first draft of the manuscript with support from BW and HV. KLC and BW screened titles, abstracts, and full text for the online cessation interventions in the Netherlands. Different versions of the manuscript have been reviewed and conceptualized by all authors. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Theoretical factors of grey literature.

[PDF File (Adobe PDF File), 27KB - jmir_v19i6e230_app1.pdf]

Multimedia Appendix 2

BCTs of grey literature.

[PDF File (Adobe PDF File), 27KB - jmir_v19i6e230_app2.pdf]

References

1. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]
2. Aspect Consortium. European Commission, Directorate General for Health and Consumer Protection. 2004. Tobacco or health in the European Union: past, present and future URL: http://ec.europa.eu/health/archive/ph_determinants/life_style/tobacco/documents/tobacco_fr_en.pdf [accessed 2017-06-05] [WebCite Cache ID 6r05bELj2]
3. Volksgezondheidenzorg. Roken. 2013. Sterfte en verloren levensjaren door roken URL: https://www.volksgezondheidenzorg.info/onderwerp/roken/cijfers-context/oorzaken-en-gevolgen [accessed 2017-06-06] [WebCite Cache ID 6r11v7l8A]
4. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2005 Apr 18(2):CD001292. [doi: 10.1002/14651858.CD001292.pub2] [Medline: 15846616]
5. Civljak M, Stead L, Hartmann-Boyce J, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2013 Jul 10(7):CD007078. [doi: 10.1002/14651858.CD007078.pub4] [Medline: 23839868]
6. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
7. Centraal Bureau voor de Statistiek. 2016. Statline: Internet; toegang, gebruik en faciliteiten URL: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=83429NED&D1=0&D2=0,3-6,20-22&D3=0&D4=a&HDR=T&STB=G1,G2,G3&VW=T [accessed 2017-08-01]
8. van Deursen AJ, Van Dijk JA. The digital divide shifts to differences in usage. New Med Soc 2014;16(3):507-526.

9. Zarcadoolas C, Blanco M, Boyer J, Pleasant A. Unweaving the Web: an exploratory study of low-literate adults' navigation skills on the World Wide Web. J Health Commun 2002;7(4):309-324. [doi: 10.1080/10810730290088157] [Medline: 12356289]
10. Reinwand D, Schulz D, Crutzen R, Kremers S, de Vries H. Who follows eHealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. J Med Internet Res 2015 May 11;17(5):e115 [FREE Full text] [doi: 10.2196/jmir.3932] [Medline: 25963607]
11. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [Medline: 10223015]
12. Swartz LH, Noell JW, Schroeder SW, Ary DV. A randomised control study of a fully automated internet based smoking cessation programme. Tob Control 2006 Feb;15(1):7-12 [FREE Full text] [doi: 10.1136/tc.2003.006189] [Medline: 16436397]
13. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press; 2005.
14. Evers S, Wolfs C, Ponds R, Fasotti L, Wekking E. Economische evaluatie van neuropsychologische behandeling. In: Neuropsychologische Behandeling. Amsterdam: Uitgeverij Boom; 2010:115-140.
15. Bootzin R. Behavior modification and therapy: An introduction. In: Behavior modification and therapy: an introduction. Cambridge, Mass: Winthrop Publishers; 1975.
16. Kanfer F, Goldstein A. Helping People Change: A Textbook of Methods. New York: Pergamon Press; 1975.
17. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
18. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [Medline: 2782727]
19. Prochaska J, DiClemente C. The Transtheoretical Approach: Crossing Traditional Boundaries of Therapy. Malabar, FL: Krieger Pub; 1994.
20. Ajzen I. The theory of planned behavior. Organ Behav Human Dec 1991;50(2):179-211.
21. Webb T, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
22. Smit E, Evers SM, de Vries H, Hoving C. Cost-effectiveness and cost-utility of Internet-based computer tailoring for smoking cessation. J Med Internet Res 2013 Mar 14;15(3):e57 [FREE Full text] [doi: 10.2196/jmir.2059] [Medline: 23491820]
23. de Vries H, Eggers SM, Bolman C. The role of action planning and plan enactment for smoking cessation. BMC Public Health 2013 Apr 26;13:393 [FREE Full text] [doi: 10.1186/1471-2458-13-393] [Medline: 23622256]
24. de Vries H, Eggers SM, Jinabhai C, Meyer-Weitz A, Sathiparsad R, Taylor M. Adolescents' beliefs about forced sex in KwaZulu-Natal, South Africa. Arch Sex Behav 2014 Aug;43(6):1087-1095 [FREE Full text] [doi: 10.1007/s10508-014-0280-8] [Medline: 24729133]
25. de Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, et al. The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. Health Educ Res 2003 Oct;18(5):611-626. [Medline: 14572020]
26. Ubhi H, Michie S, Kotz D, van Schayck OC, Selladurai A, West R. Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features. Transl Behav Med 2016 Sep;6(3):410-417 [FREE Full text] [doi: 10.1007/s13142-015-0352-x] [Medline: 27528530]
27. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
28. West R, Evans A, Michie S. Behavior change techniques used in group-based behavioral support by the English stop-smoking services and preliminary assessment of association with short-term quit outcomes. Nicotine Tob Res 2011 Dec;13(12):1316-1320. [doi: 10.1093/ntr/ntr120] [Medline: 21742650]
29. Borrelli B, Ritterband LM. Special issue on eHealth and mHealth: challenges and future directions for assessment, treatment, and dissemination. Health Psychol 2015 Dec;34S:1205-1208. [doi: 10.1037/hea0000323] [Medline: 26651461]
30. Cochrane Tobacco Addiction Group. URL: http://tobacco.cochrane.org/about-us [accessed 2017-06-20] [WebCite Cache ID 6rMnWPJfc]
31. de Gee EA, Blankers M, Duivis H, Spijkerman M, Fischer E, Leone S. Trimbos Institute. 2014. Programmeringsstudie Technologisch Ondersteunde Innovatieve Tabaksontmoediging URL: https://assets.trimbos.nl/docs/75a53af6-c3ed-487b-a8fd-b5135fa5a3e5.pdf [accessed 2017-06-06] [WebCite Cache ID 6r0tHU58I]
32. Elfeddali I, de Vries H, Bolman C, Pronk T, Wiers RW. A randomized controlled trial of Web-based Attentional Bias Modification to help smokers quit. Health Psychol 2016 Dec;35(8):870-880. [doi: 10.1037/hea0000346] [Medline: 27505210]
33. Te Poel F, Bolman C, Reubsaet A, de Vries H. Efficacy of a single computer-tailored e-mail for smoking cessation: results after 6 months. Health Educ Res 2009 Dec;24(6):930-940. [doi: 10.1093/her/cyp036] [Medline: 19574405]

34. Elfeddali I, Bolman C, Candel MJ, Wiers R, de Vries H. Preventing smoking relapse via Web-based computer-tailored feedback: a randomized controlled trial. J Med Internet Res 2012 Aug 20;14(4):e109 [FREE Full text] [doi: 10.2196/jmir.2057] [Medline: 22903145]
35. Elfeddali I, Bolman C, de Vries H. SQ4U-a computer tailored smoking relapse prevention program incorporating planning strategy assignments and multiple feedback time points after the quit-attempt: development and design protocol. Contemp Clin Trials 2012 Jan;33(1):151-158. [doi: 10.1016/j.cct.2011.09.008] [Medline: 22015574]
36. Smit E, Candel MJ, Hoving C, de Vries H. Results of the PAS study: a randomized controlled trial evaluating the effectiveness of a Web-based multiple tailored smoking cessation program combined with tailored counseling by practice nurses. Health Commun 2016 Sep;31(9):1165-1173. [doi: 10.1080/10410236.2015.1049727] [Medline: 26934538]
37. de Jong S, Candel M, Segaar D, Cremers H, de Vries H. Efficacy of a Web-based computer-tailored smoking prevention intervention for Dutch adolescents: randomized controlled trial. J Med Internet Res 2014 Mar 21;16(3):e82 [FREE Full text] [doi: 10.2196/jmir.2469] [Medline: 24657434]
38. Stanczyk N, Smit E, Schulz D, de Vries H, Bolman C, Muris JW, et al. An economic evaluation of a video- and text-based computer-tailored intervention for smoking cessation: a cost-effectiveness and cost-utility analysis of a randomized controlled trial. PLoS One 2014;9(10):e110117 [FREE Full text] [doi: 10.1371/journal.pone.0110117] [Medline: 25310007]
39. 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 Mar 03;16(3):e69 [FREE Full text] [doi: 10.2196/jmir.3016] [Medline: 24589938]
40. Stanczyk N, Bolman C, Smit E, Candel MJ, Muris JW, de Vries H. How to encourage smokers to participate in web-based computer-tailored smoking cessation programs: a comparison of different recruitment strategies. Health Educ Res 2014 Feb;29(1):23-40. [doi: 10.1093/her/cyt104] [Medline: 24287402]
41. Stanczyk NE, Crutzen R, Bolman C, Muris J, de Vries H. Influence of delivery strategy on message-processing mechanisms and future adherence to a Dutch computer-tailored smoking cessation intervention. J Med Internet Res 2013 Feb 06;15(2):e28 [FREE Full text] [doi: 10.2196/jmir.2153] [Medline: 23388554]
42. Dijk F, de Vries H. Internet-based smoking-cessation aids for young people: practical experience and current evidence. In: Res Pract Health. 2008 Presented at: FCHE Workshop; Sep 18-19, 2008; Berlin p. 35-46.
43. de Jong S, Segaar D, de Vries H. Effectiviteit en bereik van e-health interventie Smoke Alert. 2012. Onderzoeksrapport URL: http://www.alliantienederlandrookvrij.nl/wp-content/uploads/2012/docs/rapporten/jongeren/Onderzoeksrapport_SmokeAlert2%200.pdf [accessed 2017-06-06] [WebCite Cache ID 6r0wpZIuv]
44. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 Presented at: CHI '93 Conference on Human Factors in Computing Systems; Apr 24-29, 1993; Amsterdam.
45. van den Berg M, van Baal P, de Wit G, Schuit A. literatuursignalering en modellering. RIVM rapport 27009. 2008. Kosteneffectiviteit van preventie URL: http://www.rivm.nl/dsresource?objectid=cde704df-3d29-4e08-8f51-c425501a4a01 [accessed 2017-06-06] [WebCite Cache ID 6r0xtyqpt]
46. Online Zelfhulp Tabak. URL: http://www.zelfhulptabak.nl/Portal [accessed 2017-06-20] [WebCite Cache ID 6rMl5NQo9]
47. Blankers M, Koeter MW, Schippers G. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341. [doi: 10.1037/a0023498] [Medline: 21534652]
48. Blankers M, Smit E, van der Pol P, de Vries H, Hoving C, van Laar M. The missing=smoking assumption: a fallacy in Internet-based smoking cessation trials? Nicotine Tob Res 2016 Jan;18(1):25-33. [doi: 10.1093/ntr/ntv055] [Medline: 25744969]
49. Feenstra T, Hamberg-van Reenen HH, Hoogenveen R, Rutten-van Mölken MP. Cost-effectiveness of face-to-face smoking cessation interventions: a dynamic modeling study. Value Health 2005;8(3):178-190 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04008.x] [Medline: 15877590]
50. Elder JP, Ayala GX, Harris S. Theories and intervention approaches to health-behavior change in primary care. Am J Prev Med 1999 Nov;17(4):275-284. [Medline: 10606196]
51. Peters G, de Bruin M, Crutzen R. Everything should be as simple as possible, but no simpler: towards a protocol for accumulating evidence regarding the active content of health behaviour change interventions. Health Psychol Rev 2015;9(1):1-14 [FREE Full text] [doi: 10.1080/17437199.2013.848409] [Medline: 25793484]
52. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
53. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
54. Quiñonez SG, Walthouwer MJ, Schulz D, de Vries H. mHealth or eHealth? Efficacy, use, and appreciation of a Web-based computer-tailored physical activity intervention for Dutch adults: a randomized controlled trial. J Med Internet Res 2016 Nov 09;18(11):e278 [FREE Full text] [doi: 10.2196/jmir.6171] [Medline: 27829576]

55. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
56. Hassandra M, Lintunen T, Kettunen T, Vanhala M, Toivonen H, Kinnunen K, et al. Effectiveness of a mobile phone app for adults that uses physical activity as a tool to manage cigarette craving after smoking cessation: a study protocol for a randomized controlled trial. JMIR Res Protoc 2015 Oct 22;4(4):e125 [FREE Full text] [doi: 10.2196/resprot.4600] [Medline: 26494256]
57. Sadasivam R, Borglund E, Adams R, Marlin B, Houston T. Impact of a collective intelligence tailored messaging system on smoking cessation: the Perspect randomized experiment. J Med Internet Res 2016 Nov 08;18(11):e285 [FREE Full text] [doi: 10.2196/jmir.6465] [Medline: 27826134]

Abbreviations

ABM: Attentional Bias Modification
BCT: behavior change technique
PAS: Personal Advice in Stopping smoking
QALY: quality-adjusted life year
RCT: randomized controlled trial
RR: relative risk
SQ4U: Stay Quit for You
STQ: Support to Quit
VWS: Volksgezondheid, Welzijn en Sport

Edited by G Eysenbach; submitted 13.01.17; peer-reviewed by O Perski, T McFadyen, J Apolinário-Hagen; comments to author 01.03.17; revised version received 10.04.17; accepted 11.04.17; published 23.06.17.

Please cite as:
Cheung KL, Wijnen B, de Vries H
A Review of the Theoretical Basis, Effects, and Cost Effectiveness of Online Smoking Cessation Interventions in the Netherlands: A Mixed-Methods Approach
J Med Internet Res 2017;19(6):e230
URL: http://www.jmir.org/2017/6/e230/
doi:10.2196/jmir.7209
PMID:28645889

Review

Web-Based Interventions for Weight Loss or Weight Loss Maintenance in Overweight and Obese People: A Systematic Review of Systematic Reviews

Angela Sorgente[1,2], MSc; Giada Pietrabissa[1,2], PhD; Gian Mauro Manzoni[1,3], PhD; Federica Re[2], MSc; Susan Simpson[4], PhD; Sara Perona[2], BSc; Alessandro Rossi[1], MSc; Roberto Cattivelli[1,2], PhD; Marco Innamorati[5,6], PhD; Jeffrey B Jackson[7], PhD; Gianluca Castelnuovo[1,2], PhD

[1]Psychology Research Laboratory, Ospedale San Giuseppe, IRCCS, Istituto Auxologico Italiano, Verbania, Italy
[2]Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy
[3]eCampus University, Faculty of Psychology, Como, Italy
[4]University of South Australia, School of Psychology, Social Work & Social Policy, Adelaide, Australia
[5]European University of Rome, Rome, Italy
[6]Skinner Institute, Rome, Italy
[7]Virginia Tech, Falls Church, VA, United States

Corresponding Author:
Giada Pietrabissa, PhD
Department of Psychology, Catholic University of the Sacred Heart
L.go Gemelli, 1
Milan,
Italy
Phone: 39 3495324303
Email: giada.pietrabissa@unicatt.it

Abstract

Background: Weight loss is challenging and maintenance of weight loss is problematic. Web-based programs offer good potential for delivery of interventions for weight loss or weight loss maintenance. However, the precise impact of Web-based weight management programs is still unclear.

Objective: The purpose of this meta-systematic review was to provide a comprehensive summary of the efficacy of Web-based interventions for weight loss and weight loss maintenance.

Methods: Electronic databases were searched for systematic reviews and meta-analyses that included at least one study investigating the effect of a Web-based intervention on weight loss and/or weight loss maintenance among samples of overweight and/or obese individuals. Twenty identified reviews met the inclusion criteria. The Revised Assessment of Multiple SysTemAtic Reviews (R-AMSTAR) was used to assess methodological quality of reviews. All included reviews were of sufficient methodological quality (R-AMSTAR score ≥22). Key methodological and outcome data were extracted from each review.

Results: Web-based interventions for both weight loss and weight loss maintenance were more effective than minimal or control conditions. However, when contrasted with comparable non-Web-based interventions, results were less consistent across reviews.

Conclusions: Overall, the efficacy of weight loss maintenance interventions was stronger than the efficacy of weight loss interventions, but further evidence is needed to more clearly understand the efficacy of both types of Web-based interventions.

Trial Registration: PROSPERO 2015: CRD42015029377; http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015029377 (Archived by WebCite at http://www.webcitation.org/6qkSafdCZ)

(J Med Internet Res 2017;19(6):e229) doi:10.2196/jmir.6972

KEYWORDS

Internet; review; delivery of health care; obesity; overweight; telemedicine; weight loss; body weight maintenance; treatment outcome

Introduction

Obesity and overweight have reached epidemic proportions globally and pose a major risk for serious chronic diseases, including type 2 diabetes, cardiovascular disease, hypertension, sleep apnea, osteoarthritis, and certain forms of cancer [1]. Such conditions may further impact individuals' quality of life and well-being [2]. Moreover, people suffering from weight disorders are at greater risk of social, emotional, and psychological problems such as depression, poor self-esteem, and social isolation [3]. Functional interventions aimed at reducing weight and maintaining weight loss, while working on related pathologies, are typically combined treatment options (nutritional, physical, behavioral, psychological, pharmacological, surgical) [4]. Although these usually lead to short-term weight loss, long-term maintenance of results is rarely achieved [5,6]. Consequently, alternative integrative programs aimed at supporting long-lasting weight loss are typically needed. As a result, a number of Web-based interventions for weight loss or weight loss maintenance have been recently developed, and their efficacy has been tested in a number of randomized controlled trials (RCTs). Web-based therapy could help patients overcome barriers to treatment such as long distances to clinics and long waiting times. Most Web-based interventions have zero waiting time, and all are considerably cheaper than face-to-face therapy, enabling widespread dissemination of treatment [7]. Furthermore, Web-based interventions are cost-effective and provide greater user access, flexibility, and anonymity [8]. Therefore, Web-based interventions are especially relevant for patients who might not otherwise access treatment for reasons such as fear of social stigma associated with seeking treatment.

The published systematic reviews and meta-analyses of Web-based interventions for weight loss and weight loss maintenance reveal conflicting conclusions. Thus, the purpose of this meta-review was to (1) examine the published systematic reviews that included at least one study assessing the efficacy of a Web-based intervention for weight loss and/or weight loss maintenance for samples of participants who are either overweight or obese, (2) produce a summary of the scientific evidence, (3) identify the strengths and weaknesses of Web-based interventions to help clinicians select the best treatment option for their patients, and (4) provide empirically supported suggestions for practice.

Methods

This review was carried out according to the guidelines proposed by Smith et al [9]. The protocol for this study was registered in 2015 in the International Prospective Register of Systematic Reviews (PROSPERO).

Inclusion and Exclusion Criteria

Given the absence of an established standard definition for systematic reviews, the following inclusion and exclusion criteria provide the parameters used for defining systematic reviews for this meta-review. Only reviews that satisfied the following criteria were included: (1) used a systematic review method (eg, critical review, literature review, meta-analysis), (2) indicated the method for identifying and evaluating studies for inclusion, (3) included at least one study assessing the efficacy of a Web-based intervention for weight loss and/or weight loss maintenance on the absolute variation and/or the change in percentage of body weight or body mass index (BMI) for a sample of overweight and/or obese people, and (4) received a methodological quality score of 22 or higher on the Revised Assessment of Multiple SysTemAtic Reviews (R-AMSTAR; see methodological quality assessment section for details). There were no restrictions for participant age, publication year, or publication language to obtain the maximum number of reviews possible. Non-English publications were translated to facilitate data extraction.

Search Methods

As suggested by Smith et al's guidelines [9], the following electronic databases were searched: PubMed, the Cochrane Library, PsycINFO (ProQuest platform), and the Centre for Review and Dissemination (CRD), which includes the Database of Abstracts of Reviews of Effects (DARE). Search terms were identified for each of the following relevant categories: population (obese, obesity, overweight), intervention (online, Web, computer), outcome (weight loss, weight loss maintenance), and review type (review, meta-analysis). Boolean searches were then conducted to systematically link the various combinations of category terms (and their variations through truncation) as search terms, Medical Subject Headings (MeSH) keywords and Emtree keywords to identify potential systematic reviews [10]: ("Review"[MeSH]) OR ("Meta-Analysis"[MeSH]) OR review OR meta-analysis AND ("Computers"[MeSH]) OR online OR web OR computer AND ("Weight Loss"[MeSH]) OR weight loss OR weight loss maintenance AND ("Obesity"[MeSH]) OR ("Overweight"[MeSH]) OR obese OR obesity OR overweight.

In addition, the contents of *Obesity Reviews*, *Annual Review of Public Health*, and the *Journal of Medical Internet Research* were searched using the following syntax: (review OR meta-analysis) AND (online OR web OR computer) AND ("weight loss") AND (obes* OR overweight).

As a supplement to electronic searching, reference lists were checked to identify additional potential systematic reviews. The search was performed for records published through December 2015.

Selection Process

Titles and abstracts of records resulting from the literature search were independently screened by authors FR and SP. When further clarification was needed, the full text was retrieved. Disagreements were resolved by a third author (AS). In accordance with one of Smith et al's recommendations [9], the review team included at least one person with methodological expertise in conducting systematic reviews (GMM and AS) and at least two experts on the topic under review (GC, GMM, and GP).

Data Extraction and Management

Authors SP and FR independently extracted the following data and resolved any disagreements in consultation with a third author (AS): (1) authorship and publication-related information; (2) aims of the review; (3) searched databases; (4) inclusion criteria; (5) number of included studies; (6) overall sample size and participant age, gender, race, and BMI; (7) overall length of treatment, including follow-up time points; (8) country in which the interventions were developed; and (9) outcomes of the interventions.

Reviews that included studies that did not investigate the efficacy of weight loss and/or weight loss maintenance programs among obese and overweight participants were coded for the total number of included studies and the number of included studies involving treatments for weight loss and/or weight loss maintenance in a sample of obese and/or overweight persons. Additional relevant information was obtained by retrieving original studies and contacting review authors as necessary for coding purposes.

Methodological Quality Assessment

The R-AMSTAR [11] was used to quantitatively measure the methodological quality of included systematic reviews by assessing the presence of the following 11 domains: (1) an a priori design, (2) duplicate study selection and data extraction, (3) a comprehensive literature search, (4) the use of status of publication as an inclusion criteria, (4) a list of included/excluded studies, (5) characteristics of included studies, (6) documented assessment of the scientific quality of included studies, (7) appropriate use of the scientific quality in forming conclusions, (8) the appropriate use of methods to combine findings of studies, (8) assessment of the likelihood of publication bias, and (9) documentation of conflicts of interest. Each domain's score ranged between 1 and 4, and the R-AMSTAR total scores had a range of 11 to 44.

A total score of 22 (ie, a mean of two criteria for each item were satisfied) was required for systematic review inclusion, thus excluding low-scoring systematic reviews [11]. The authors in charge of extracting data from the selected reviews (SP and FR) also preliminarily and independently assessed the methodological quality of the contributions. A third author (AS) resolved any discrepancies.

Data Synthesis

First, reviews were analyzed and relevant information was extracted and recorded. Then, the results across the different reviews were aggregated through a second-order qualitative synthesis of treatment efficacy conclusions for weight loss interventions and then for weight loss maintenance interventions. Quantitative results were recorded but no second-order overall effect was calculated from the included meta-analyses including similar sets of studies because a meta-analysis of meta-analyses is possible only if the data from individual studies have not been used in more than one meta-analysis [9]. Thus, pooled effects of overlapping reviews were only compared in order to investigate the consistency of results.

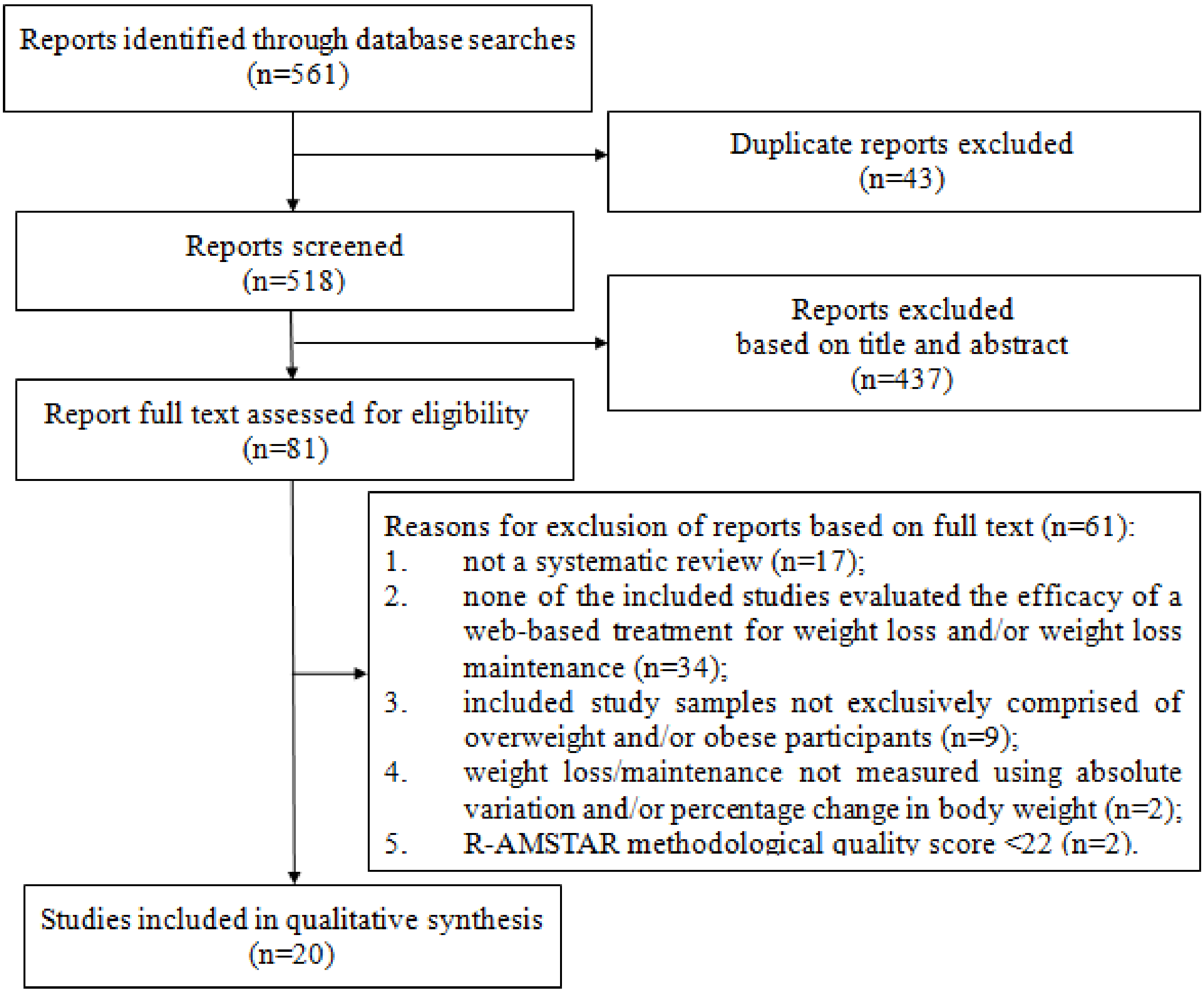
Ultimately, the strengths and weaknesses of the various Web-based interventions listed across the reviews were summarized.

Results

Included Reviews

A flowchart indicating the selection of included systematic reviews is presented in Figure 1. Searches of electronic databases identified 561 reports, of which 43 were duplicate and 437 were excluded based on information from the title and abstract. The remaining 81 reports were then evaluated for inclusion by reviewing the full text of each report, resulting in the exclusion of 61 reports for the following reasons (3 reports were omitted for more than one reason): (1) no systematic review was presented (n=17), (2) none of the included studies evaluated the efficacy of a Web-based treatment for weight loss and/or weight loss maintenance (n=34), (3) the included study samples were not exclusively comprised of overweight and/or obese participants (ie, study samples were also comprised of normal-weight participants; n=9), (4) weight change (loss or maintenance) was not measured or summarized in terms of absolute variation and/or change in percentage of body weight or BMI (n=2), and (5) the review R-AMSTAR methodological quality score was less than 22 (n=2). A total of 20 systematic reviews were finally included [12-31]. Multimedia Appendix 1 details the reasons for exclusion and whether inclusion or exclusion was based on information from the title and abstract or full text for each of the evaluated reports.

Figure 1. Systematic review selection flowchart.



Description of Included Systematic Reviews

Tables 1 and 2 provide a summary of the characteristics of each included systematic review. Overall, 10 [12-21] of the 20 systematic reviews examined the effects of Web-based interventions for weight loss and/or weight loss maintenance, whereas the other 10 systematic reviews [22-31] examined the effects of both Web-based and traditional interventions for weight loss and/or weight loss maintenance.

Inclusion criteria for the majority of the systematic reviews consisted of age restrictions (only adults [12-18,20,22-24,26,27,31]), research design restrictions (only RCT [12-17,19,21,24,28-30]), and outcome restrictions (differences in weight loss or weight loss maintenance as primary treatment effects [12,13,16-20,26,29,31]). In terms of other age-related inclusion criteria, five systematic reviews had no age restriction [19,21,25,29,30] and only one restricted inclusion to participants younger than 18 years [28]. In terms of BMI score inclusion criteria, with the exception of two systematic reviews that restricted study inclusion to participants with a BMI greater than 28 [26] and 30 or greater [29], the BMI cut-off was 25 or greater in the other 18 systematic reviews.

Table 1. Characteristics of the included systematic reviews (N=20).

Author(s), publication year	Aim of the review	Searched databases[a]	Inclusion criteria of studies	Included (relevant) studies, n (n)
Tsai et al, 2005 [22]	Describe the components, costs, and efficacy of weight loss programs	Medline	Only adults, in USA, ≥1-year follow-up assessment(s), ≥10 participants and treatment stated and program lasted ≥12 weeks	10 (1)
Weinstein, 2006 [12]	Describe the efficacy of Web-based weight loss and weight loss management programs	PubMed, CINAHL, CL, NIHCT	Only adults, in USA, RCT, ≥1 Web-based intervention, BMI≥25, published in peer-reviewed journal, primary outcome weight loss or weight loss management	8 (8)
Sharma, 2007 [23]	Review behavioral interventions for prevention and treatment of overweight and obesity	Medline	Only adults, English, published 2000-2006, educational approach	23 (2)
Neve et al, 2009 [14]	Assess the effectiveness of Web-based interventions for weight loss and weight loss management	CL, Medline, EMBASE, CINAHL, Web of Science, Scopus, PsycINFO	Only adults, RCT, ≥1 Web-based intervention, BMI≥25	18 (18)
Osei-Assibey et al, 2010 [24]	Evaluate the effectiveness of dietary and lifestyle weight loss interventions	Medline, EMBASE, CCTR, CINAHL, DARE	Only adults, RCT, ≥50% of participants were minorities, treatment lasted ≥6 months	19 (1)
Arem and Irwin, 2011 [13]	Summarize the state of the science of Internet-delivered weight loss interventions and highlight their strengths and weaknesses	PubMed	Only adults, RCT, BMI≥25, primary outcome weight loss or weight loss management, website or Web-based programming	9 (9)
Burke et al, 2011 [25]	Evaluate the effect of self-monitoring diet, physical activity level, and weight management program on weight loss in behavioral treatment studies	Medline, PsycINFO	In USA, published 1989-2009, studies on effect and use of self-monitoring	22 (3)
Manzoni et al, 2011 [16]	Evaluate the effectiveness of Web-based interventions for weight loss and weight loss management	PubMed, PsycINFO, CL, NIH	Only adults, RCT, published in peer-reviewed journal, primary outcome weight loss or weight loss management	25 (25)
Kodama et al, 2012 [15]	Review the weight loss or weight loss management effect of the Internet component in obesity treatment programs	Medline, EMBASE	Only adults, RCT, BMI≥25, website or Web-based programming	23 (23)
Reed et al, 2012 [17]	Evaluate the impact of computer-based technology on interventions for weight loss	Medline, CC, CINAHL, PsycINFO	Only adults, RCT, BMI≥25, used computer/interactive technologies, primary outcome weight loss or weight loss management, control group received non-computer-based intervention	11 (11)
Wieland et al, 2012 [18]	Assess the effect of interactive computer-based interventions for weight loss or weight loss management	CC, Medline, EMBASE, CINAHL, LILACS, PsycINFO	Only adults, BMI≥25, includes RCTs or quasi-RCTs, primary outcome weight loss or weight loss management, website or Web-based programming, lasted ≥4 weeks	18 (18)

Author(s), publication year	Aim of the review	Searched databases[a]	Inclusion criteria of studies	Included (relevant) studies, n (n)
Young et al, 2012 [26]	Investigate the effectiveness of weight loss and weight loss management interventions and identify the characteristics associated with effectiveness	CINAHL, EMBASE, Medline, PsycINFO, PubMed, Sport Discus, Scopus, Web of Science	Only adults, BMI>28, primary outcome weight loss or weight loss management, only male participants	24 (6)
Chang et al, 2013 [19]	Describe the use and impact of social media in online weight management program	PubMed, PsycINFO, EMBASE, Web of Science, Scopus	RCT, published in peer-reviewed journal, primary outcome weight loss or weight loss management, social media component	20 (20)
Grunenberg et al, 2013 [21]	Investigate the effectiveness of Web-based psychological interventions for weight loss	Medline, PsycINFO, Psyndex	RCT, BMI≥25, primary outcome weight loss or weight loss management, website or Web-based programming, control either waitlist or standard waiting treatment, psychologically based intervention for behavioral modification	5 (5)
Bennett et al, 2014 [20]	Evaluate the efficacy of eHealth weight management programs	PubMed, EMBASE, CINAHL, CL, Web of Science	Only adults, in USA, English, BMI≥25, primary outcome weight loss or weight loss management, used computer/interactive technologies	6 (6)
Hartmann-Boyce et al, 2014 [27]	Determine the clinical effectiveness of multicomponent behavioral weight management program	BIOSIS, CL, CC, CP, DR, EMBASE, HT, Medline, PsycINFO, SCI	Only adults, BMI≥25	8 (1)
Altman and Wilfley, 2015 [28]	Evaluate the evidence for overweight and obesity treatments	PubMed, PsycINFO, Google Scholar	RCT, treated children and adolescents, estimated costs for childhood obesity treatments	9 (1)
Gilmartin and Murphy, 2015 [29]	Evaluate the effectiveness of behavioral weight loss management interventions	CL, Medline, EMBASE, PsycINFO, Web of Science	RCT, primary outcome weight loss or weight loss management, BMI≥30	13 (1)
Levine et al, 2015 [30]	Examine technology-assisted weight loss interventions and highlight innovation, impact, and pragmatism	PubMed, Medline, EMBASE, CD, CC	RCT, used computer/interactive technologies, ambulatory setting	16 (8)
Raaijmakers et al, 2015 [31]	Evaluate the effectiveness of technology-based interventions on weight loss and quality of life	PubMed, PsycINFO, Web of Science, Science Direct, CINAHL, EMBASE	Only adults, BMI≥25, used computer/interactive technologies, primary outcome weight loss or weight loss management	27 (12)

[a] BIOSIS: BIOSIS Preview; CC: Cochrane Central; CCTR: Centre for Care Technology Research; CD: Cochrane Database of Systematic Reviews; CINAHL: Cumulative Index to Nursing and Allied Health Literature; CL: Cochrane Library; CP: Cochrane Public Health Group and Evidence for Policy and Practice Information Centre; DARE: Database of Abstracts of Reviews of Effects; HT: Health Technology Assessment database; DR: Database of Abstracts of Reviews and Effects; LILACS: Latin American and Caribbean Health Sciences Literature; NIH: The National Institutes of Health; NIHCT: National Institutes of Health Clinical Trials database; RC: review of company Web sites; SCI: Science Citation Index.

Table 2. Characteristics of the included systematic reviews (continued) (N=20).[a]

Author, publication year	N	Age (years)	Women, %	BMI	Duration of the intervention, mean or range[b]	Country (race)[c]	Outcomes
Tsai et al, 2005 [22]	1877	—	—	—	12 wk-2 y	US	Weight
Weinstein, 2006 [12]	791	30-62	55		22 wk-12 mo	US	Weight
Sharma, 2007 [23]	—	—	—	—	3 mo-9 y	AU, BE, CN, FI, IT, JP, NL, SE, UK, US	BMI, weight, waist circumference, body fat
Neve et al, 2009 [14]	5700	≥18	77	≥25	6 wk-2 y	UK, US	Weight
Osei-Assibey et al, 2010 [24]	—	Mean=47.2	—	—	>6 mo	Western countries (people of color)	BMI
Arem and Irwin, 2011 [13]	—	34-54	50-100	Mean=29	3-18 mo	(White but 2 studies did not report race)	Weight
Burke et al, 2011 [25]	9668	—	41-100	—	—	US (>white)	
Manzoni et al, 2011 [16]	8324	≥18	76.7	—	6 wk-2 y	US and other unspecified countries	Weight
Kodama et al, 2012 [15]	8697	≥18	66.1	26.2-35.7	3-30 mo	—	Weight
Reed et al, 2012 [17]	1866	≥18	71.64	—	2-12 mo	—	Weight
Wieland et al, 2012 [18]	4140	≥18	73	>25	4 wk-30 mo	—	Weight
Young et al, 2012 [26]	1869	18-65	>0	>28	3-24 mo	AU, CN, FI, JP, NL, SE, UK, US	Weight
Chang et al, 2013 [19]	—	—	—	—	—	AU, CN, UK, US	Waist circumference, BMI, physical activity level, dietary intake
Grunenberg et al, 2013 [21]	727	≥18	57	≥25	3-12 mo	—	Weight, BMI, waist circumference
Bennett et al, 2014 [20]	4899	≥18	—	≥25	3-30 mo	US (people of color)	Weight
Hartmann-Boyce et al, 2014 [27]	>3700	40-52	—	25 (≥23 among Asians)	—	AU, CH, DE, UK, US	Weight
Altman and Wilfley, 2015 [28]	—	6-18	—	—	—	(White/people of color)	Weight
Gilmartin and Murphy, 2015 [29]	—	>18	—	≥30	>2 y	CN, FI, SE, UK, US	Weight
Levine et al, 2015 [30]	6786	Middle-aged	62	—	3-36 mo	(71% White)	Weight
Raaijmakers et al, 2015 [31]	—	—	—	>20, ≥30, ≥40	—	AU, CN, DE, JP, UK, US	Weight, quality of life, adherence

[a] —: Information that was not reported.

[b] Mo: month; wk: week; y: year.

[c] AU: Australia; BE: Belgium; CH: Switzerland; CN: China; DE: Germany; FI: Finland; IT: Italy; JP: Japan; NL: Netherlands; SE, Sweden; UK: United Kingdom; US: United States.

The total number of participants in each systematic review ranged from 727 [21] to 9668 [25]. In terms of gender, nine systematic reviews reported no gender-specific data [32-36]; the majority of reviews that included information on participant gender had samples with more female participants than male participants.

The number of databases that were searched for each systematic review ranged from 1 [13,23] to 10 [27]. A total of 351 studies were evaluated across the 20 systematic reviews, of which only

83 evaluated the effects of Web-based interventions for weight loss and/or weight loss maintenance. Of the 83 studies, 73 evaluated Web-based interventions for weight loss and 10 evaluated Web-based interventions for weight loss maintenance (see Multimedia Appendix 2 for details). The 83 studies were mostly conducted in the United States, Europe, or Australia; 51 of 83 studies were only included in one systematic review and the remaining 32 studies were included in more than one systematic review. The study that was included in the most systematic reviews [37] was included in a total of eight reviews.

Methodological Quality of Included Reviews

The R-AMSTAR scores of the 20 included reviews (Table 3) ranged from 23 to 43 points, with a mean of 30.5 (SD 5.5), a median of 30.5 (IQR 9.25). The highest mean score across the 20 systematic reviews (mean 4, SD 0) was for providing the characteristics of the included studies (item 6), whereas the lowest mean score was for the inclusion of conflicts of interest (item 11; only Hartmann-Boyce et al [27] fully satisfied this criterion).

Table 3. Systematic review quality (N=20).

Systematic review	R-AMSTAR Item[a]											Score
	1	2	3	4	5	6	7	8	9	10	11	
Tsai et al, 2005 [22]	3	4	3	1	3	4	4	4	1	1	1	29
Weinstein, 2006 [12]	3	1	4	2	4	4	1	2	1	1	1	24
Sharma, 2007 [23]	3	4	1	3	3	4	1	1	1	1	1	23
Neve et al, 2009 [14]	4	4	4	1	3	4	4	4	4	4	1	37
Osei-Assibey et al, 2010 [24]	3	4	4	3	2	4	4	4	2	1	1	32
Arem and Irwin, 2011 [13]	3	4	4	2	3	4	1	1	1	1	1	25
Burke et al, 2011 [25]	3	1	4	4	3	4	1	1	1	1	1	24
Manzoni et al, 2011 [16]	3	4	4	1	3	4	1	1	1	1	1	24
Young et al, 2011 [26]	3	4	4	3	4	4	4	4	2	1	1	34
Kodama et al, 2012 [15]	3	4	3	1	1	4	1	1	4	4	1	27
Reed et al, 2012 [17]	4	4	4	1	3	4	4	1	4	4	1	34
Wieland et al, 2012 [18]	4	4	4	4	4	4	4	1	4	4	1	38
Chang et al, 2013 [19]	4	4	4	3	3	4	4	4	1	1	1	33
Grunenberg et al, 2013 [21]	3	1	3	2	1	4	4	4	4	2	3	31
Bennett et al, 2014 [20]	3	4	4	4	4	4	4	4	1	1	1	34
Hartmann-Boyce et al, 2014 [27]	4	4	4	3	4	4	4	4	4	4	4	43
Altman and Wilfley, 2015 [28]	3	4	3	1	3	4	1	1	2	1	1	24
Gilmartin and Murphy, 2015 [29]	3	4	3	2	3	4	4	4	1	1	1	30
Levine et al, 2015 [30]	4	4	4	4	3	4	1	1	4	4	1	34
Raaijmakers et al, 2015 [31]	3	4	4	1	3	4	4	4	1	1	1	30
Mean	3.32	3.55	3.60	2.30	3.00	4.00	2.80	2.55	2.20	1.95	1.25	30.50
Median	3.00	4.00	4.00	2.00	3.00	4.00	4.00	3.00	1.50	1.00	1.00	30.50
SD	0.48	1.10	0.75	1.17	0.86	0.00	1.51	1.50	1.40	1.39	0.79	5.54
IQR	1	0	0.5	2	0.25	0	3	3	3	3	0	9.25

[a] Item 1: a priori design; item 2: duplicate study selection and data extraction; item 3: comprehensive literature search; item 4: publication status as an inclusion criteria; item 5: list of included and excluded studies; item 6: characteristics of included studies; item 7: documented assessment of the scientific quality of included studies; item 8: appropriate use of the scientific quality in forming conclusions; item 9: appropriate use of methods to combine study findings; item 10: assessment of publication bias likelihood; item 11: conflict of interest documentation.

Efficacy of Web-Based Interventions for Weight Loss and/or Weight Loss Maintenance

Effect sizes of Web-based interventions for weight loss and weight loss maintenance, together with the specific comparison interventions, are reported in Table 4. The intervention purpose (ie, weight loss or weight loss maintenance) is also specified in Table 4. In addition, details for each meta-analysis, such as the number of studies used to calculate effect sizes, the heterogeneity among included studies, and the combined sample size are also reported in Table 4. Except for Kodama et al [15], all meta-analyses performed quantitative data synthesis separately for both the type of condition compared to the Web-based intervention and whether the purpose of the intervention was weight loss or weight loss maintenance. Given that several primary studies were included in more than one meta-analysis, issues related to statistical independence prevented meta-meta-analysis of the effect sizes across the meta-analyses. Overall, the meta-analysis effect sizes were relatively small in magnitude, suggesting that although Web-based interventions were significantly more or less effective than the comparison conditions, this difference may have little clinical relevance.

Web-Based Interventions for Weight Loss

Web-Based Interventions Versus Control Conditions (Minimal Interventions)

Across reviews, Web-based interventions were found to be significantly more effective than minimal treatments in reducing weight. Specifically, Wieland et al [18] found that Web-based interventions were significantly more effective than minimal treatments in reducing weight and BMI at 3- and 6-month follow-ups. Young et al [26] and Weinstein [12] also obtained a significant difference in weight change favoring Web-based interventions over controls. Additionally, in Raaijmakers et al’s review [31], six technological-based interventions generated a significantly greater effect in terms of weight loss than no treatment conditions. In Bennett et al’s review [20], more than half of the identified trials reported significantly greater weight loss outcomes for eHealth interventions compared to control conditions. Similar results were found in Levine et al’s review [30], in which 12 Web-based interventions (75%) resulted in greater weight loss compared to control conditions. Finally, Grunenberg et al [21] found a Web-based intervention to be more effective than control groups (waitlist and standard waiting treatment) at reducing both BMI and weight. Neve et al [14] was the only review to report no significant difference in weight loss between Web-based interventions and control groups at treatment termination. This contrasting finding may be attributable to Neve et al [14] including fewer studies that tested this comparison (n=3) than the other reviews that found Web-based intervention to be more effective at promoting weight loss compared to control conditions. Also, the meaning of the term “control condition” varied across reviews from no intervention [31] to providing participants with a weight loss manual [14]. Additionally, Neve et al [14] combined treatment effects irrespective of time points (from 16 weeks to 12 months), whereas other reviews pooled the studies’ effects separately for each follow-up point.

Web-Based Interventions Versus Non-Web-Based Comparable Interventions

Included systematic reviews that included studies comparing Web-based treatments with non-Web-based comparable interventions presented inconsistent results. For example, Raaijmakers et al [31] found Web-based interventions to be more effective than usual care, and Tsai et al [22] found greater weight reduction among the participants assigned to a Web-based condition (ie, Weight Watchers) than those receiving self-help interventions. Levine et al [30] also concluded that technology-based interventions can successfully supplement primary care interventions for weight loss outcomes. Finally, Weinstein [12] found that Web-based interventions are significantly more effective than their non-Web-based counterparts both when the latter consists of usual care or when participants receive information from a manual.

On the other hand, other reviews found Web-based interventions to be as effective as non-Web-based comparable interventions. Specifically, Burke et al [25] examined three studies on online dietary self-monitoring and found that online treatments resulted in significant within-group weight loss; however, when compared with a paper diary self-monitoring condition, the pooled effect size was no longer statistically significant. In addition, Bennett et al [20] found that eHealth approaches led to relatively modest weight loss outcomes with undetermined clinical significance when compared with traditional individual and group-based interventions. Finally, Reed et al [17] found that computer-based technology led to significantly less weight loss than comparable interventions. Therefore, the research on the efficacy of Web-based interventions compared to similar non-Web-based interventions is inconclusive. This lack of consistency may be due to the large heterogeneity of non-Web-based comparison interventions in the primary studies. For example, the non-Web-based comparison interventions ranged from manualized interventions to a counseling program in the studies included by Raaijmakers et al [31].

Table 4. A summary of meta-analyses.

Review and comparison[a]	Number of included articles	Outcome (units) and follow-up	N	Heterogeneity[b]				Effect size[c] (95% CI)	*P*
				χ^2 (df)	*P*	T^2	I^2		
Neve et al, 2009 [14]									
Web vs control	3	Weight loss (kg)	151	12.8 (2)	.002	—	84.4%	0.73 (–0.6, 1.51) WMD	.07
Enhanced Web vs basic Web	3	Weight loss (kg)	217	3.8 (3)	.28	—	21%	2.24 (1.27, 3.21)[d] SMD	<.001
Web vs control	2	Weight loss maintenance (kg)	409	0.02 (1)	.90	—		–0.30 (–0.34,–0.26)[d] WMD	<.001
Web vs face-to-face	2	Weight loss maintenance (kg)	182	12.2 (3)	.007	—	76%	1.80 (–1.18, 4.79) WMD	.24
Kodama et al, 2012 [15]	23	Weight loss, weight loss maintenance (kg)	8697	—	<.001	—	84%	–0.68 (–1.29,–0.08) WMD	.03
Reed et al, 2012 [17]									
Intervention with Web vs without Web	5	Weight loss (kg)	336	0.7 (5)	.98	0.00	0%	–1.48 (–2.52,–0.43) WMD	.006
Web vs non-Web	5	Weight loss (kg)	544	14.2 (5)	.01	3.61	65%	0.36 (–1.80, 2.53) WMD	.74
Web vs non-Web	4 (articles published in 1995 or later)	Weight loss (kg)	538	1.7 (4)	.78	0.00	0%	1.47 (0.13, 2.81) WMD	.03
Intervention with Web vs without Web	4	Weight loss (kg); short-term follow-up	100	0.2 (4)	>.99	0.00	0%	–1.89 (–3.41,–0.38) WMD	.01
Intervention with Web vs without Web	1	Weight loss (kg); long-term follow-up	236	NA	NA	NA	NA	–1.10 (–2.55, 0.35) WMD	.14
Intervention with Web vs without Web	2	Weight loss (kg); short-term follow-up	53	0.04 (1)	.85	0.00	0%	–1.95 (–3.50,–0.40) WMD	.01
Intervention with Web vs without Web	3	Weight loss (kg) long-term follow-up	283	0.03 (3)	>.99	0.00	0%	–1.08 (–2.50,–0.34) WMD	.14
Intervention with Web vs without Web	2 (articles published prior to 1995)	Weight loss (kg)	47	0.02 (2)	.99	0.00	0%	–0.63 (–7.91, 6.66) WMD	.87
Intervention with Web vs without Web	3 (articles published in 1995 or later)	Weight loss (kg)	289	0.6 (2)	.72	0.00	0%	–1.50 (–2.55, 0.44) WMD	.006
Intervention with Web vs without Web	3	Weight loss (BMI)	380	0.8 (2)	.67	0.00	0%	–0.43 (–0.83,–0.03) WMD	.04
Web vs non-Web	2	Weight loss (BMI)	51	0.3 (2)	.88	0.00	0%	0.44 (–1.15, 2.03)[c] WMD	.59
Web vs control	2	Weight loss (kg)	511	0.04 (1)	.84	0.00	0%	−1.5 (−2.1, −0.9) MD	<.001
Wieland et al, 2012 [18]									
Web vs control	2	Weight loss maintenance (kg)	897	0.7 (1)	.41	0.00	0%	−0.7 (−1.2, −0.2) MD	.004
Web vs face-to-face	2	Weight loss maintenance (kg)	897	2.9 (1)	.09	0.41	66%	0.5 (−0.5, 1.6) MD	.32
Grunenberg et al, 2013 [21]									
Web vs control	5	Weight loss (BMI)	727	10.5 (4)	.03	0.15	62%	–0.49 (–0.95,–0.03) MD	.04
Web vs control	5	Weight loss (kg)	727	16.7 (4)	.002	1.46	76%	–1.32 (–2.59,–0.06) MD	.04

[a] Web vs control: Web-based intervention vs control condition (minimal intervention); enhanced Web vs basic Web: enhanced Web-based interventions

vs basic Web-based interventions; Web vs face-to-face: Web-based intervention vs face-to-face intervention; intervention with Web vs intervention without Web: adding a Web-based component to an intervention vs the same intervention without the Web-based component; Web vs non-Web: Web-based interventions vs non-Web-based comparable interventions.

[b] I^2: Percentage of the variation across studies attributable to study heterogeneity rather than chance, indicating the level of inconsistency across study results; T^2: between-study variance.

[c] Effect sizes were retrieved from original articles reporting a statistically significant pooled effect estimated from at least two trials. All studies except for those indicated used a random effects model to calculate the aggregated effect size. MD: mean difference; SMD: standardized mean difference (Cohen *d*; standardized weighted aggregated average difference score between conditions across primary studies that use different outcome measures/metrics; to facilitate aggregation across measures/metrics, the between-condition difference for each primary study is converted to standard deviation units that are then weighted with primary studies with more precise estimates carrying more weight in aggregation); WMD: weighted mean difference (unstandardized weighted aggregated average difference score between conditions across primary studies that use the same outcome measure/metric; the between-condition difference for each primary study is weighted with primary studies with more precise estimates carrying more weight in aggregation).

[d] A fixed effect model was used to calculate the aggregated effect size.

Web-Based Interventions Versus Face-to-Face Interventions

This section summarizes results from systematic reviews in which Web-based interventions were compared with non-Web-based counterparts involving face-to-face interventions. In Wieland et al [18], face-to-face interventions were more effective at promoting weight loss than Web-based interventions. Also, Raaijmakers et al [31] reviewed a primary study in which face-to-face treatment led to a significantly greater reduction in weight than Web-based intervention. Similarly, Kodama et al [15] concluded that using a Web-based intervention as a substitute for a face-to-face intervention produced unfavorable results.

Web-Based Interventions Versus Hybrid Interventions Versus Face-to-Face Interventions

Web-based interventions were further compared with hybrid interventions (ie, including both Web-based and non-Web-based components) in several systematic reviews. For example, Kodama et al [15] came to the conclusion that adding face-to-face interventions to Web-based interventions increases the impact of the Web-based interventions on weight loss. In contrast, Wieland et al [18] reported that Web-based interventions and hybrid conditions (ie, Web-based intervention face-to-face treatment) did not differ significantly in their effects. In the study reported by Wieland et al [18], the hybrid condition was also compared with the face-to-face intervention without Web-based components. This pairwise comparison indicated that mean weight loss achieved by face-to-face treatments was significantly greater than mean weight loss achieved by hybrid conditions. In comparison, Reed et al [17] determined that computer-based treatments combined with standard interventions (ie, behavioral programs, face-to-face treatments) resulted in significantly more weight loss than standard interventions only, at least when short-term effects were considered. Similarly, Tsai et al [22] found significantly greater weight loss in participants receiving a Web-based treatment (ie, Weight Watchers program) combined with individualized contacts than in participants receiving a face-to-face intervention. Due to these contrasting results, it is not clear if hybrid interventions are more effective in increasing weight loss than single component interventions (ie, either only Web-based or only non-Web-based).

Enhanced Web-Based Interventions Versus Basic Web-Based Interventions

Several systematic reviews also compared the effects of Web-based interventions that differed on both the interaction level and the extent to which they were tailored to users' needs. Osei-Assibey et al [24] and Hartmann-Boyce et al [27] reported that Web-based tailored programs were more effective in weight loss than information-only websites, despite the disappearance of this difference by 18 months after treatment [27]. Levine et al [30] concluded that interventions including clinician-guided software or feedback from personnel promoted greater weight loss than fully automated interventions, thus underlining the importance of the interactive component. Both Neve et al [14] and Wieland et al [18] were also in agreement about Web-based interventions with interactive components being effective in reducing weight. Specifically, the enhanced Web-based interventions considered in both reviews included additional programs, such as email-based behavioral therapy delivered by a doctoral-level therapist (including feedback and behavioral lessons), behavioral e-counseling provided by a counselor (weekly email behavioral counseling and feedback), and automated e-counseling (weekly automated and tailored messages). Sharma [23] reported a greater weight reduction in behavioral e-counseling conditions compared to basic Web-based programs, and Manzoni et al [16] concluded that Web-based behavioral programs enhanced by tailored feedback or self-monitoring resulted in more effective weight reduction than education-only Web-based interventions. Similarly, Osei-Assibey et al [24] found that weight change was greater for Web-based programs supporting collaborative interactions than for Web-based educational interventions. Furthermore, Weinstein [12] concluded that online counseling may be a valid alternative to time-consuming clinical programs and health care costs. Still, in Altman and Wilfley [28], an included study revealed a Web-based lifestyle behavior modification program to be more effective that a Web-based health education program at treatment termination, but not at 2-year follow-up (probably because program usage decreased over time). There was some evidence that website usage was associated with enhanced outcomes. For example, one study included in Chang et al [19] reported a Web-mediated walking program that was administered both alone and in conjunction with online community components. No differences were found in physical activity outcomes between participants who had access to social media versus those who did not; however, among participants using online communities, higher use of social media was associated with greater weight loss. Overall, these findings suggest that tailored and interactive Web-based interventions promote greater weight loss than basic Web-based interventions (ie, delivering information via the Internet). However, results also indicate that utilization of Web-based resources has potential to boost treatment effectiveness.

Within-Subject Comparisons

Arem and Irwin's review [13] summarized the results of studies measuring within-group effects of Web-based interventions by comparing weight outcomes before and after treatment. Findings indicate Web-based interventions caused a decrease in weight ranging from 0.8 kg (considered to be natural noise) to 4.9 kg. The authors concluded that the large degree of treatment heterogeneity across studies reduced their ability to make reliable conclusions.

Web-Based Interventions for Weight Loss Maintenance

Web-Based Interventions Versus Control Conditions (Minimal Interventions)

Six reviews compared Web-based weight loss maintenance interventions with control conditions with consistent results. Specifically, Neve et al [14], Manzoni et al [16], Gilmartin and Murphy [29], Young et al [26], Bennett et al [20], and Wieland et al [18] found that Web-based interventions were, on average, significantly more effective than minimal interventions in promoting weight loss maintenance.

Web-Based Interventions Versus Non-Web-Based Comparable Interventions

Two systematic reviews reported results of studies comparing Web-based interventions for weight loss maintenance with non-Web-based comparable interventions. Kodama et al [15] concluded that, in comparison with non-Web-based conditions, Web-based programs were ineffective. In contrast, Bennett et al [20] reported on a study in which an interactive Web-based intervention was compared to a monthly face-to-face or telephone-based intervention. In this case, the amount of weight regained did not differ significantly between the two interventions. Overall, the inconsistent results for this particular comparison of treatments may be due to the diverse characteristics of the non-Web-based interventions that were provided.

Web-Based Interventions Versus Face-to-Face Interventions

Change et al [19], Weinstein [12], Neve et al [14], and Manzoni et al [16] reported that maintenance of weight loss was similar between Web-based and non-Web-based face-to-face interventions. In comparison, Gilmartin and Murphy [29] and Wieland et al [18] concluded that Web-based treatments were less effective than face-to-face interventions, especially if the latter were intensive and not minimal [12]. Specifically, Gilmartin and Murphy [29] stated that face-to-face interventions and facilitator-led interventions were more effective than remotely delivered methods such as Web-based interventions. Finally, Wieland et al [18] referred to three studies comparing face-to-face interventions with Web-based interventions for weight loss maintenance. Both minimal (once monthly or less) and intensive (more than once per month) face-to-face interventions were found to be more effective than computer-based interventions. However, the amount of weight lost by persons assigned to the control (minimal) conditions was relatively small and was not maintained in the long term, making the clinical significance of these differences unclear.

Strengths and Weaknesses of Web-Based Interventions for Weight Loss and Weight Loss Maintenance

Eleven systematic reviews provided information about the strengths and weaknesses of the evaluated Web-based interventions, with seven of them specifically identifying three advantages of Web-based interventions:

They may enhance perceived self-control within treatment. Specifically, Levine et al [30], Manzoni et al [16], Kodama et al [15], and Raaijmakers et al [31] pointed out that Web-based interventions allow people to self-monitor their weight and behaviors, thereby increasing their perceived sense of control and ultimately reducing the number of dropouts [15].

They may facilitate patient-patient and patient-expert interactions, thus allowing people to receive regular consistent feedback on their behaviors and answers to questions [16,30,31].

Web-based interventions for weight loss are more cost-effective than standard treatments [31].

Only three systematic reviews reported weaknesses associated with Web-based interventions for weight loss and weight loss maintenance. Arem and Irwin [13] reported that the limited effectiveness of Web-based interventions may be due to the restricted range of programs and updates that are available, which may not always be suitable to meet users' needs. Bennett et al [20] and Chang et al [19] indicated that Web-based treatments may be affected by low levels of familiarity and self-efficacy associated with managing Web technologies, as well as by limitations associated with access to the Internet.

Discussion

Principal Results

To our knowledge, this systematic review of systematic reviews represents the first state-of-the-art analysis of Web-based intervention efficacy for weight loss and weight loss maintenance. According to the selection criteria, 20 systematic reviews were deemed eligible for inclusion. All 20 systematic reviews were published in 2005 or later. They mainly investigated Web-based interventions for weight loss, with only a few investigating Web-based interventions for weight loss maintenance. Findings from the meta-systematic review regarding Web-based interventions for weight loss and weight loss management were mixed; in fact, the findings within the included systematic reviews are often conflicting, particularly in relation to the efficacy of Web-based weight loss interventions. The conflicting results are likely due to the notable heterogeneity of inclusion criteria across the systematic reviews for selecting primary studies. Nevertheless, all the included systematic reviews demonstrated methodological rigor (R-AMSTAR score ≥22), although none received the highest possible score for methodological quality. Specifically, Hartmann-Boyce et al [27] was the only systematic review that fully met the 11th R-AMSTAR criterion of disclosing conflicts of interest, ensuring the validity of the systematic review results. Indeed, by not declaring conflicts of interest, it is impossible to rule out the existence of publication bias. The synthesis of the included systematic reviews identified both strengths and weaknesses of the Web-based interventions for both weight loss and weight loss maintenance. Web-based interventions may facilitate continuous and automated tracking of health-related behaviors by supporting self-regulatory techniques, patient involvement, and patient commitment to treatment. Moreover, Web-based connectivity permits the sharing of information among health professionals and peers. However, the efficacy and dissemination of Web-based interventions may be affected by the gap in access to computers and the Internet, as well as the lack of technological literacy among potential users.

Limitations

In conducting this systematic review of systematic reviews, it was sometimes difficult to make a clear distinction between Web-based interventions (delivered over the Internet) and computer-based interventions (delivered over the Internet or by installing computer software) because these terms are often used interchangeably or defined differently [18]. It was also difficult to compare the overall effects across systematic reviews since they were calculated differently (ie, weighted mean difference vs standardized mean difference). Furthermore, conclusions of a second-order review are not drawn from results of primary studies, but from reviews that have synthesized the results of primary studies. Because the same primary studies were often included in more than one systematic review, not only did this overrepresentation prevent meta-meta-analysis of the efficacy of Web-based interventions for weight loss and weight loss maintenance, it also compromised the accuracy of the meta-systematic review findings, thus affecting the actual reliability of findings based on second-order data synthesis. In addition, because the number of primary studies on which each systematic review was based varied substantially (from 1 to 25), the findings from some systematic reviews were based on more evidence than the findings of other systematic reviews. Given the limitations associated with this meta-systematic review, the conclusions should be interpreted with some caution.

Conclusion

Are Web-Based Interventions for Weight Loss and Weight Loss Maintenance Effective?

This systematic review of systematic reviews concludes that Web-based interventions for weight loss are often more effective than minimal treatments (only Neve et al [14] reached different conclusions); however, when compared with non-Web-based or hybrid interventions, results appear inconsistent across reviews. More encouraging results in terms of weight loss were obtained when Web-based interventions were enhanced (ie, more interactive and tailored) than when they were basic (ie, information website). Nevertheless, Web-based interventions for weight loss were less effective than face-to-face interventions across the selected reviews. Results were more encouraging in relation to Web-based weight loss maintenance interventions, which were found to be more effective than minimal interventions across all the reviews and, in some reviews, as effective as the non-Web-based counterpart. The decision of whether or not to substitute an in-person intervention for weight loss maintenance with a comparable Web-based treatment mainly depends on patient costs, needs, and preferences.

These conclusions should be considered cautiously. Reported effect sizes were small; for example, weight loss of 1 to 2 kg may not by clinically significant, irrespective of significance level. Also, conclusions might be affected by heterogeneity across primary studies. In fact, research designs differed in terms of type of intervention, sample size, duration, control condition, etc. Although the conclusions from this meta-systematic review are of significant interest, the real impact of Web-based interventions for weight loss remains unclear, suggesting the need for greater clarity in both the definition and specificity of the different types of Web-based treatments available, as well as how each intervention can be best matched to users' needs. Further evidence is therefore necessary.

Suggestions and Implications for Future Research

Authors interested in providing a new summary review of the literature on the efficacy of Web-based interventions for weight loss and weight loss maintenance for obese and overweight patients can refer to the list of included records reported in Multimedia Appendix 2. A total of 83 primary studies investigating the effectiveness of online interventions for weight loss and weight loss maintenance were identified, analyzed, and compared across systematic reviews. A single study-level review of these primary studies that pinpoints differences and inconsistences across the primary studies would be beneficial. In addition, meta-analysis of these primary studies would provide a quantitative summary of the efficacy of Web-based treatments for weight loss and weight loss maintenance.

Future systematic reviews should provide a high level of detail when reporting primary study effect sizes. Specifically, detailed information about the nature of the comparison conditions (especially for instances in which there are multiple comparison conditions) and the various types of efficacy outcomes is necessary to allow other researchers and practitioners to more clearly interpret the results and to facilitate replication of these studies. Also, the effects of Web-based interventions for weight loss and Web-based interventions for weight loss maintenance should not be compared with each other [15] because they differ in both aims and outcomes. In addition, researchers and practitioners should carefully consider the cost of Web-based intervention; although technology-based treatments are fundamental in reducing health care costs, cost-effectiveness is often not adequately evaluated (if at all) in comparing Web-based and face-to-face interventions. Therefore, interventions should be compared in terms of both efficacy and cost-effectiveness (see Raaijmakers et al [31] and Wieland et al [18] for examples of reviews that evaluated treatment efficacy and cost-effectiveness).

Authors' Contributions

GMM conceived the study. SP and FR conducted the literature searches and extracted, tabulated, and interpreted data. AS resolved any controversy, outlined the study design, and wrote the first draft of the manuscript. GP contributed to data interpretation and manuscript writing. SS revised the first and last drafts of the manuscript for content and the English language. AR, RC, and MI revised the advanced versions of the manuscript. JBJ revised the final version of the manuscript for both content and the English language. GC approved the final version of the manuscript. No author received financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Reasons for exclusion for each record.

[XLSX File (Microsoft Excel File), 112KB - jmir_v19i6e229_app1.xlsx]

Multimedia Appendix 2

Target studies included in the reviews.

[XLSX File (Microsoft Excel File), 25KB - jmir_v19i6e229_app2.xlsx]

References

1. Bray GA. Medical consequences of obesity. J Clin Endocrinol Metab 2004 Jun 01;89(6):2583-2589 [FREE Full text] [doi: 10.1210/jc.2004-0535]
2. Taylor VH, Forhan M, Vigod SN, McIntyre RS, Morrison KM. The impact of obesity on quality of life. Best Pract Res Clin Endocrinol Metab 2013 Apr;27(2):139-146. [doi: 10.1016/j.beem.2013.04.004]
3. Puhl RM, Heuer CA. The stigma of obesity: a review and update. Obesity 2009 May 17;17(5):941-964. [doi: 10.1038/oby.2008.636]
4. Castelnuovo G, Pietrabissa G, Manzoni GM, Corti S, Ceccarini M, Borrello M, et al. Chronic care management of globesity: promoting healthier lifestyles in traditional and mHealth based settings. Front Psychol 2015 Oct 15:1557-1558. [doi: 10.3389/fpsyg.2015.01557]
5. Wadden TA, Butryn ML, Wilson C. Lifestyle modification for the management of obesity. Gastroenterology 2007 May;132(6):2226-2238. [doi: 10.1053/j.gastro.2007.03.051]
6. Castelnuovo G, Manzoni GM, Villa V, Cesa G, Pietrabissa G, Molinari E. The STRATOB study: design of a randomized controlled clinical trial of Cognitive Behavioral Therapy and Brief Strategic Therapy with telecare in patients with obesity and binge-eating disorder referred to residential nutritional rehabilitation. Trials 2011 May 09;12(114):1-7. [doi: 10.1186/1745-6215-12-114]
7. Kersting A, Schlicht S, Kroker K. Internet therapy. Opportunities and boundaries. Nervenarzt 2009 Jul;80(7):797-804. [doi: 10.1007/s00115-009-2721-5]
8. Dolemeyer R, Tietjen A, Kersting A, Wagner B. Internet-based interventions for eating disorders in adults: a systematic review. BMC Psychiatry 2013 Aug 06;13(270):1-16. [doi: 10.1186/1471-244X-13-207]
9. Smith V, Devane D, Begley C, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011 Feb;11(15):1-6. [doi: 10.1186/1471-2288-11-15]
10. Johnson K. Database Search Tips and Tricks. Los Angeles, CA: Meriam Library, California State University; 2002 Feb. URL: https://www.csuchico.edu/lins/handouts/Database_Search_Tips.pdf [accessed 2017-06-01] [WebCite Cache ID 6qu6CTOLx]
11. Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, et al. From systematic reviews to clinical recommendations for evidence-based health care: validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J 2010;4:84-91. [doi: 10.2174/1874210601004020084]
12. Weinstein PK. A review of weight loss programs delivered via the Internet. J Cardiovasc Nurs 2006;21(4):259-260. [Medline: 16823276]
13. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):236-243 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]
14. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x]
15. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of Web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012;36(5):675-685. [doi: 10.1038/ijo.2011.121]

16. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011 Mar 04;7:19-28. [doi: 10.2174/1745017901107010019]
17. Reed VA, Schifferdecker KE, Rezaee ME, O?Connor S, Larson RJ. The effect of computers for weight loss: a systematic review and meta-analysis of randomized trials. J Gen Intern Med 2012;27(1):99-108. [doi: 10.1007/s11606-011-1803-9]
18. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:1-157. [doi: 10.1002/14651858.CD007675.pub2]
19. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov 28;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
20. Bennett GG, Steinberg DM, Stoute C, Lanpher M, Lane I, Askew S, et al. Electronic health (eHealth) interventions for weight management among racial/ethnic minority adults: a systematic review. Obes Rev 2014 Oct;15(4):146-158. [doi: 10.1111/obr.12218]
21. Grunenberg E, Lin J, Baumeister H. [Effectiveness of web-based psychological interventions for weight loss--a systematic review]. Rehabilitation (Stuttg) 2013 Jun;52(3):182-187. [doi: 10.1055/s-0033-1343489] [Medline: 23761207]
22. Tsai AG, Wadden TA. Systematic review: an evaluation of major commercial weight loss programs in the United States. Ann Intern Med 2005;142(1):56-66. [doi: 10.7326/0003-4819-142-1-200501040-00012]
23. Sharma M. Behavioural interventions for preventing and treating obesity in adults. Obes Rev 2007 Sep;8(5):441-449. [doi: 10.1111/j.1467-789X.2007.00351.x] [Medline: 17716301]
24. Osei-Assibey G, Kyrou I, Adi Y, Kumar S, Matyka K. Dietary and lifestyle interventions for weight management in adults from minority ethnic/non-White groups: a systematic review. Obes Rev 2010;11(11):769-776. [doi: 10.1111/j.1467-789X.2009.00695.x]
25. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011;111(1):92-102. [doi: 10.1016/j.jada.2010.10.008]
26. Young MD, Morgan PJ, Plotnikoff RC, Callister R, Collins CE. Effectiveness of male-only weight loss and weight loss maintenance interventions: a systematic review with meta-analysis. Obes Rev 2012 May;13(5):393-408. [doi: 10.1111/j.1467-789X.2011.00967.x]
27. Hartmann-Boyce J, Johns DJ, Jebb SA, Summerbell C, Aveyard P, Behavioural Weight Management Review Group. Behavioural weight management programmes for adults assessed by trials conducted in everyday contexts: systematic review and meta-analysis. Obes Rev 2014 Nov;15(11):920-932 [FREE Full text] [doi: 10.1111/obr.12220] [Medline: 25112559]
28. Altman M, Wilfley DE. Evidence update on the treatment of overweight and obesity in children and adolescents. J Clin Child Adolesc Psychol 2015;44(4):521-537. [doi: 10.1080/15374416.2014.963854] [Medline: 25496471]
29. Gilmartin J, Murphy M. The effects of contemporary behavioural weight loss maintenance interventions for long term weight loss: A systematic review. J Res Nurs 2015;20(6):481-496. [doi: 10.1177/1744987115599671]
30. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. Technology-assisted weight loss interventions in primary care: a systematic review. J Gen Intern Med 2015 Jan;30(1):107-117. [doi: 10.1007/s11606-014-2987-6]
31. Raaijmakers LC, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: A systematic review. Appetite 2015 Dec;95:138-151. [doi: 10.1016/j.appet.2015.07.008]
32. Webb TL, Joseph J, Yardley L, Michie S. Using the Internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376]
33. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009;28(6):690-701. [doi: 10.1037/a0016136]
34. Dellifraine JL, Dansky KH. Home-based telehealth: a review and meta-analysis. J Telemed Telecare 2008;14(2):62-66. [doi: 10.1258/jtt.2007.070709]
35. Galani C, Schneider H. Prevention and treatment of obesity with lifestyle interventions: review and meta-analysis. Int J Public Health 2007;52(6):348-359. [Medline: 18368998]
36. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727]
37. Womble LG, Wadden TA, McGuckin BG, Sargent SL, Rothman RA, Krauthamer-Ewing ES. A randomized controlled trial of a commercial internet weight loss program. Obes Res 2004 Jun;12(6):1011-1018. [doi: 10.1038/oby.2004.124]

Abbreviations

BMI: body mass index
CRD: Centre for Review and Dissemination
DARE: Database of Abstracts of Reviews of Effects
MeSH: Medical Subject Headings
PROSPERO: Prospective Register of Systematic Reviews

R-AMSTAR: Revised Assessment of Multiple SysTemAtic Reviews
RCT: randomized controlled trial

Edited by M Focsa; submitted 09.11.16; peer-reviewed by M Rethlefsen, J Apolinário-Hagen; comments to author 28.11.16; revised version received 20.01.17; accepted 26.01.17; published 26.06.17.

Please cite as:
Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, Rossi A, Cattivelli R, Innamorati M, Jackson JB, Castelnuovo G
Web-Based Interventions for Weight Loss or Weight Loss Maintenance in Overweight and Obese People: A Systematic Review of Systematic Reviews
J Med Internet Res 2017;19(6):e229
URL: http://www.jmir.org/2017/6/e229/
doi:10.2196/jmir.6972
PMID:28652225

Original Paper

Web-Based Alcohol Intervention: Study of Systematic Attrition of Heavy Drinkers

Theda Radtke[1], PhD; Mathias Ostergaard[2,3], MSc Psych; Richard Cooke[4], PhD; Urte Scholz[1,5], PhD

[1]Department of Psychology, Applied Social and Health Psychology, University of Zurich, Zurich, Switzerland
[2]Forel Clinic, Ellikon an der Thur, Switzerland
[3]Department of Psychology, Clinical Psychology and Clinical Neuropsychology, University of Konstanz, Konstanz, Germany
[4]Department of Psychology, School of Life and Health Sciences, Aston University, Birmingham, United Kingdom
[5]University Research Priority Program Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland

Corresponding Author:
Theda Radtke, PhD
Department of Psychology
Applied Social and Health Psychology
University of Zurich
Binzmühlestrasse 14/14
Zurich, 8050
Switzerland
Phone: 41 44 635 72 55
Fax: 41 44 635 72 59
Email: theda.radtke@uzh.ch

Abstract

Background: Web-based alcohol interventions are a promising way to reduce alcohol consumption because of their anonymity and the possibility of reaching a high numbers of individuals including heavy drinkers. However, Web-based interventions are often characterized by high rates of attrition. To date, very few studies have investigated whether individuals with higher alcohol consumption show higher attrition rates in Web-based alcohol interventions as compared with individuals with lower alcohol consumption.

Objectives: The aim of this study was to examine the attrition rate and predictors of attrition in a Web-based intervention study on alcohol consumption.

Methods: The analysis of the predictors of attrition rate was performed on data collected in a Web-based randomized control trial. Data collection took place at the University of Konstanz, Germany. A total of 898 people, which consisted of 46.8% males (420/898) and 53.2% females (478/898) with a mean age of 23.57 years (SD 5.19), initially volunteered to participate in a Web-based intervention study to reduce alcohol consumption. Out of the sample, 86.9% (781/898) were students. Participants were classified as non-completers (439/898, 48.9%) if they did not complete the Web-based intervention. Potential predictors of attrition were self-reported: alcohol consumption in the last seven days, per week, from Monday to Thursday, on weekends, excessive drinking behavior measured with the Alcohol Use Disorder Identification Test (AUDIT), and drinking motives measured by the Drinking Motive Questionnaire (DMQ-R SF).

Results: Significant differences between completers and non-completers emerged regarding alcohol consumption in the last seven days (B=−.02, *P*=.05, 95% CI [0.97-1.00]), on weekends (B=−.05, *P*=.003, 95% CI [0.92-0.98]), the AUDIT (B=−.06, *P*=.007, 95% CI [0.90-0.98], and the status as a student (B=.72, *P*=.001, 95% CI [1.35-3.11]). Most importantly, non-completers had a significantly higher alcohol consumption compared with completers.

Conclusions: Hazardous alcohol consumption appears to be a key factor of the dropout rate in a Web-based alcohol intervention study. Thus, it is important to develop strategies to keep participants who are at high risk in Web-based interventions.

(J Med Internet Res 2017;19(6):e217) doi:10.2196/jmir.6780

KEYWORDS

attrition; dropout; alcohol drinking; intervention study; Internet; surveys and questionnaires; university student drinking; motivation

Introduction

Background

According to the World Health Organization, hazardous alcohol consumption is one of the world’s leading health risks [1]. A common definition of hazardous drinking is the consumption of at least 4 (for girls or women) or 5 (for boys or men) standard glasses of alcohol (eg, 0.3 liter beer, 0.2 liter wine, or 0.04 liter spirits) on a single occasion [2]. Hazardous alcohol consumption is most common among young people [1]. For example, in Germany, young people aged 18 to 29 (1790/7649, 23.40%) are most likely to engage in hazardous drinking compared with other age groups [3]. Long-term consequences of hazardous consumption include increased risk of developing diseases such as breast cancer, coronary heart disease, and liver cirrhosis [1]. Short-term consequences of hazardous consumption include poor educational attainment and involvement in other health risk behaviors such as riding with a driver who had been drinking, using illicit drugs, and unplanned and unsafe sexual behavior [4,5]. Against the background of such health consequences, the question of why young people consume alcoholic beverages arises. According to the Motivational Model of Alcohol Use [6] 2 dimensions are relevant to classify the motivation to drink alcohol: the valence (positive or negative) and the source (internal or external) of the outcomes individuals expect to achieve from drinking alcohol. On the basis of these 2 dimensions, 4 classes of drinking motives are generated: social motives (eg, drinking to enjoy social gatherings), enhancement motives (eg, drinking to have fun), coping motives (eg, drinking to handle worries or personal problems), and conformity motives (eg, drinking to be part of the group). Among hazardous drinkers, social motives seem to be more important relative to the other motives [7].

Due to the high health risks of alcohol consumption and the high consumption among young individuals, alcohol intervention programs in young people are of key importance. Due to the increased use of the Internet by adolescents and young adults, digital media may be an effective strategy for reaching out to young people [8,9]. Furthermore, online delivery offers a cost-effective and an easy way to deliver alcohol intervention to a large number of people [10]. Web-based interventions can be as effective as traditional face-to-face interventions at reducing alcohol consumption [11,12] with the advantages of higher availability, greater convenience, and greater anonymity [13,14] when compared with the more traditional modes of delivery.

A serious concern with Web-based delivery of interventions is the potentially high rates of attrition. According to Postel, de Haan, ter Huurne, van der Palen, Becker, and Jong [15] attrition rates of Web-based alcohol interventions can be up to 92% whereas face-to-face interventions show premature termination rates of around 50% [16].

Many different reasons for attrition in Web-based intervention studies have been previously reported. Technical reasons include visual appeal, the complexity of the visual presentation, or the number of questions [17]. Other reasons include low material incentives [18,19], asking for personal information [20], non-familiarity with the Internet [15,21-23], or low interest and motivation to take part [24]. A further aspect comprises the access to guidance support (eg, via messages) instead of a self-help Web-based intervention. As shown, attrition rates are higher in Web-based intervention studies without guidance from a professional compared with Web-based studies with a counselor [25,26]. Regarding socio-demographic characteristics, evidences exist that men are more likely to be non-completers than women (eg, [15,27]) and that attrition rates are lower among participants with a lower or middle educational level compared with highly educated participants [28].

A key factor that may cause attrition is engaging in the behaviour that is the focus of the intervention [29,30]. For example, Bewick, Trusler, Mulhern, Barkham, and Hill [31] reported that among participants in their online alcohol intervention, those who completed the study reported lower alcohol consumption at baseline compared with participants who did not complete the study. Similar patterns were found in other studies (eg, [32,33]). These results support our proposition that heavier drinkers are less likely to remain in alcohol interventions [29].

Attrition threatens the internal and external validity of any intervention [34]. A key consequence of a high attrition rate among individuals at high risk is that they are unlikely to benefit from the intervention. A further key consequence of retaining people at lower risk is that intervention results, and interpretation thereof, are likely to be highly biased [35].

The Study

Few studies to date have provided detailed analyses of attrition rates and possible reasons [36,37], and to the best of our knowledge, there are only a few studies investigating systematic attrition in Web-based alcohol interventions (eg, [15,38]) despite a high need [39]. In line with previous research [29-32] we expect that non-completers will report consuming more alcohol than completers. Relatedly, as it has been shown that hazardous drinkers report drinking for social motives more than other motives, we also expect that non-completers will be more likely to report drinking for social motives versus completers. Finally, in previous alcohol intervention studies [15,28], men were more likely to dropout than women. The current study tests the following hypotheses (H):

1. Non-completers will have a higher level of alcohol consumption compared with completers.
2. Non-completers will report higher social motives compared with completers.
3. Non-completers will be more likely to be men.

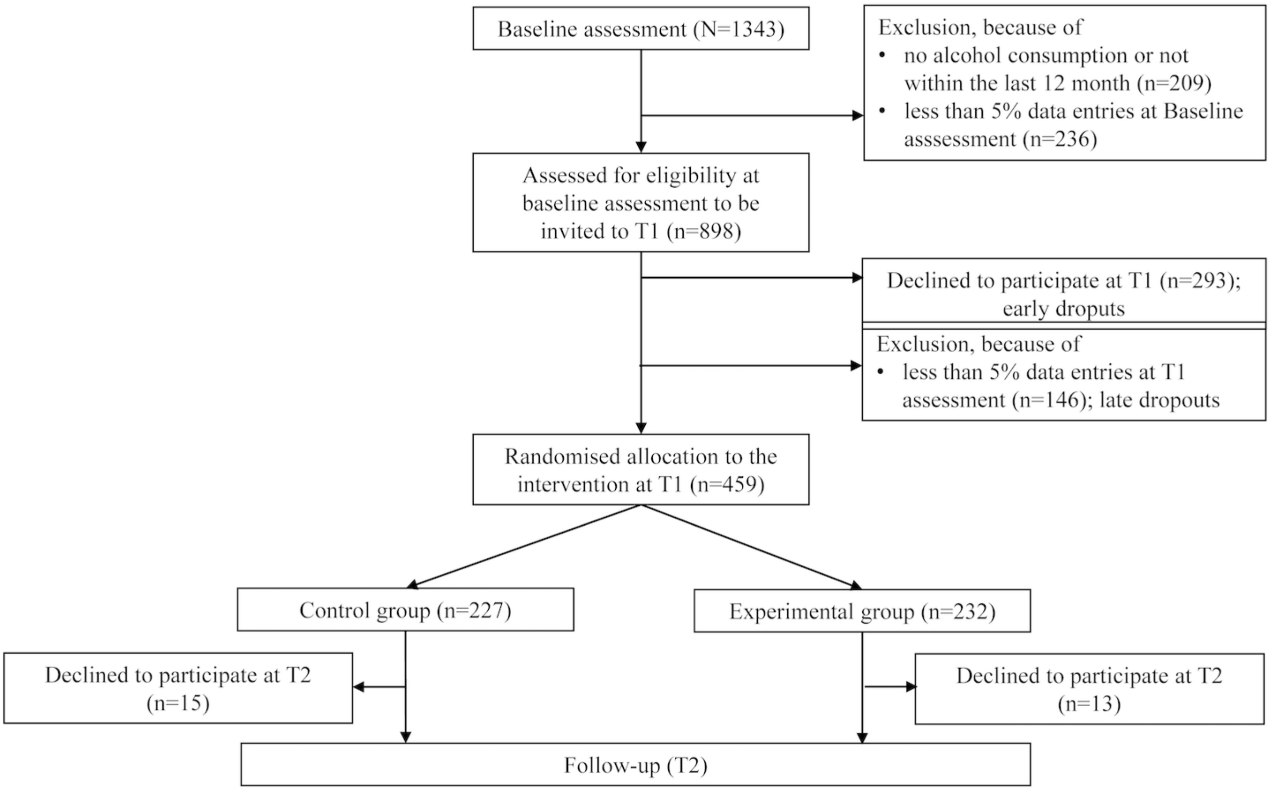
Methods

Participants

Participants were recruited in Germany via the University of Konstanz student email list, flyers, posters on campus notice boards, and via social media networks (eg, Facebook). The study was advertised as a study about individual alcohol consumption habits to recruit participants who consume alcoholic beverages. The study took place during an exam free period of the semester (May and June 2012). Overall, N=1343 people clicked on the link to the online questionnaire (see Figure 1). Of those, n=209 had to be excluded because they reported that they had not consumed any alcohol in the last 12 months. In addition, n=236 were excluded as they completed less than 5% (3/53) of the questionnaire. Although one could argue that this is also a form of attrition, it was not possible to include these participants in the analyses due to too many missing values to analyze the differences between completers and non-completers. Thus, the baseline sample consisted of n=898 participants out of which 46.8% (420/898) were male and 53.2% (478/898) were female with a mean age of 23.57 years (SD 5.19). Out of the total sample, 86.9% (781/989) were students whereas the others were employees or freelancers, PhD students, school children, or interns. Women and men reported high alcohol consumption rates per week (women: mean=5.74, SD 5.01; men: mean=12.33, SD 10.93) and high alcohol consumption rates within the last seven days (women: mean=7.92, SD 7.81; men: mean=16.28, SD 15.28); this indicated hazardous alcohol consumption according to the German national drinking guidelines [2].

As an incentive to encourage participation, 10 Amazon gift cards worth 20 Euros were raffled among participants who completed all three points of measurement. Students received no credits for taking part in the study. All individuals participated voluntarily, gave informed consent, and were treated in accordance with the ethical standards of the Declaration of Helsinki [40]. This implied that the confidentiality and privacy of all participants was assured all the time. Only the main investigator had access to personal information like the e-mail address. Personal data were stored separately from data files to ensure a strictly anonymous data analysis. Furthermore, it was not possible for the participants to fill in the questionnaire twice. However, participants were able to interrupt filling out the questionnaire and return at a later time point to complete the questionnaire. All participants who did not answer the questionnaires got a reminder by email after two days.

Figure 1. Study flow chart. T1=Time point 1, one week after baseline. T2=Follow-up, one week after T1.



Design

This study was designed as a Web-based self-affirmation alcohol intervention study [41] to test possible moderators and mediators that may underlie the effects of self-affirmation on health behavior. The study design is only described briefly because the present analyses focus on the systematic dropout of participants from the intervention, rather than the intervention itself (for further information see [42]). The study consisted of 3 measurement points. The first (baseline) point of measurement took place in May and June 2012. Only participants who consumed alcoholic beverages were included in the study. After one week, participants were invited by email to take part in the study (T1) and randomly assigned to the experimental or the control group of the self-affirmation intervention (adopted from [41,43]). Participants in the experimental condition had to choose one of their most important values from a list of several values (eg, reliability) and to write down why this value is important to them. In the control condition, participants were instructed to write about their least important value from the listed values and why it might be important to another person. After this intervention task, all participants had to read health messages about alcohol consumption. This information included a description of short-term and long-term health risks of heavy alcohol consumption (eg, cancer), a definition of heavy drinking, and the recommended intake of alcoholic beverages. Afterwards, participants had to fill in several measures. The final follow-up

(T2) took place one week later. Again, participants were invited by email to take part in the study.

Measures

The following instruments were included in the online questionnaire at the baseline measurement one week prior to the intervention.

Alcohol Consumption (adapted from [41]). Alcohol consumption was assessed with 4 items. Before answering the items, participants were provided with a brief definition and examples of standard glasses of different alcoholic beverages in line with the German drinking guidelines [44]. An example is: “One German standard drink equals 10g to 12g of pure alcohol, which equates to a small beer (250 ml), or a small glass of wine (100 ml).” The items concerning alcohol-consumption were ‘How many standard drinks do you consume per week?’ (per week), ‘Referring to the last 7 days, how many standard drinks did you consume?’ (last seven days), ‘How many standard drinks do you consume from Monday to Thursday?’ (Monday to Thursday), and ‘How many standard drinks do you consume from Friday to Sunday?’ (on weekends).

The Alcohol Use Disorders Identification Test (AUDIT, [45]) measures excessive drinking behavior. The 10-item questionnaire screens for harmful alcohol use. An item example is, ‘How often do you have six or more drinks on one occasion?’ A sum score of eight or above refers to hazardous alcohol consumption, a score between 16 and 19 refers to harmful alcohol consumption, and a score higher than 20 indicates a possible alcohol dependence. Cronbach alpha of the scale was equal to .83.

The Drinking Motive Questionnaire Revised Short Form (DMQ-R SF, [46]) measures 4 different motives to drink alcohol. The social (Cronbach alpha =.84), enhancement (Cronbach alpha=.76), coping (Cronbach alpha=.86), and conformity (Cronbach alpha=.79) motive dimensions were assessed with three items each. The 5-point Likert-scale ranges from 1 = (almost) never to 5 = (almost) always. The stem for each item was ‘I have drunk alcohol in the past 12 months’ followed by different item options, for example, ‘...in order to be liked’. Cronbach alpha for the overall scale was equal to .83.

Data Analysis

All analyses were run using SPSS version 20 (IBM, Somers, NY, USA) to test the hypotheses. In a first step, *t* tests and Chi-square tests (threshold for statistical significance $P<.05$) were conducted to reveal group differences concerning alcohol consumption, excessive drinking behavior (AUDIT), drinking motives (DMQ-R SF), and socio-demographics like gender, age, and student status. After the detection of differences between completers and non-completers, the relationship between dropout classification and possible predictors was tested with logistic regressions in a second step. Logistic regression analyses were performed with completion of the Web-based self-affirmation intervention as the dependent variable. Predictor variables showing a significant association with dropout (based on correlations) were entered into the model, whereas non-significant variables were removed. In addition, it was decided to include only one variable regarding alcohol consumption in the model to avoid multicollinearity as the variables correlated to a high degree. Thus, 4 models were calculated. The goodness of fit of the models was evaluated by the Hosmer-Lemeshow test, which needs to be non-significant.

Results

Overall, the baseline sample consisted of n=898 participants. Of these, 293 participants did not participate at T1 (early dropouts). A further n=146 participants who began completing the online questionnaire at T1 dropped out (ie, completed less than 5% data of the questionnaire) before being allocated to one of the experimental conditions (late dropouts; see Figure 1).

Overall, no significant differences were found between the early and late dropouts regarding most of the measures. The only two significant differences were found regarding the AUDIT score and the amount of alcohol consumption on weekends. Early dropouts reported significantly higher AUDIT scores compared with the late dropouts (mean=9.60 vs mean=8.47; $t_{317.76}=2.19$, $P=.03$, $d=0.21$, 95% CI [0.12-2.13]) as well as higher alcohol consumption on weekends (mean=8.37 vs mean=6.90; $t_{335.46}=2.16$, $P=.03$, $d=0.21$, 95% CI [0.13-2.82]). On the basis of these few differences and the fact that the results of the logistic regression revealed an almost identical pattern of results when running the analyses separately for early and late dropouts, it was decided not to further distinguish between both groups in the following analyses. Overall, the dropout rate from baseline to the randomized allocation to the intervention at T1 was 48.9% (439/898). At T2, a further n=28 dropped out. However, those non-completers did not differ significantly from the completers regarding all measures, except for the sub-dimension coping motives of the DMQ-R SF (mean=2.08 vs mean=1.59; $t_{457}=-3.44$, $P=.001$, $d=0.67$, 95% CI [−0.77 to −0.21]).

Tables 1 and 2 present the descriptive statistics for completers (n=459) and non-completers (n=439). Significant differences between completers and non-completers were found regarding alcohol consumption, AUDIT score, coping, enhancement and social drinking motives, gender, and student status. Non-completers showed a significantly higher alcohol consumption per week, from Monday to Thursday, on weekends and in the last seven days. In addition, non-completers reported significantly higher AUDIT scores and higher mean values regarding social, enhancement, and coping motives for drinking alcohol compared with completers. Furthermore, women and students were more likely to be completers compared with men and non-students. No significant differences between completers and non-completers were found for age, or the conformity motive.

Table 1. Sample characteristics (mean, SD) and differences concerning systematic attrition.

Sample characteristics	Attrition							
	Completers N=459		Non-completers N=439		*All*	*All*		
	Mean	SD	Mean	SD	T	*P*	95% CI	Cohen d
Alcohol consumption (standard drinks)								
per week (range 0-70)	7.20	6.85	10.60	10.50	−5.71	<.001	−4.56 to −2.23	.39
Monday to Thursday (range 0-40)	2.60	3.58	3.96	5.37	−4.45	<.001	−1.96 to −0.76	.30
Weekends (range 0-45)	5.18	4.76	7.88	7.15	−6.63	<.001	−3.50 to −1.90	.45
last seven days (range 0-95)	9.65	9.54	14.17	14.83	−5.41	<.001	−6.16 to −2.88	.36
AUDIT[a] (range 0-27)	7.19	4.20	9.22	5.27	−6.35	<.001	−2.65 to −1.40	.43
Drinking motives (range 1-5)								
social motives	3.15	.92	3.30	.98	−2.33	.02	−0.27 to −0.02	.13
enhancement motives	2.88	.92	3.08	.98	−3.22	.001	−0.32 to −0.08	.21
coping motives	1.62	.74	1.78	.85	−2.98	.003	−0.26 to −0.05	.20
conformity motives	1.33	.57	1.35	.61	−.55	.58	−0.10 to 0.06	-
Age (range 16-67)	23.51	5.17	23.63	5.22	−.33	.74	−0.80 to 0.57	-

[a] AUDIT: Alcohol Use Disorders Identification Test.

Table 2. Distribution of sample characteristics and differences concerning systematic attrition.

Sample characteristics	Attrition					
	Completers N=459		Non-completers N=439			
	N	%	N	%	χ^2	*P*
Female	269	58.6	208	47.6	10.90	.001
Student	416	90.6	360	82.0	12.61	<.001

Table 3. Intercorrelations between variables separated by completers (upper half) and non-completers (lower half).

	Completers											
Non-Completers	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	12[l]
1[a]	-	.82 <.001	.86 <.001	.74 <.001	.71 <.001	.23 <.001	.28 <.001	.21 <.001	.08 .071	−.03 .57	.34 <.001	.02 .75
2[b]	.83 <.001	-	.57 <.001	.59 <.001	.57 <.001	.13 .004	.19 <.001	.21 <.001	.06 .19	.00 .99	.26 <.001	.04 .37
3[c]	.87 <.001	.63 <.001	-	.71 <.001	.68 <.001	.24 <.001	.28 <.001	.16 .001	.04 .35	−.06 .24	.29 <.001	−.01 .91
4[d]	.80 <.001	.66 <.001	.73 <.001	-	.61 <.001	.26 <.001	.26 <.001	.16 <.001	.05 .31	−.03 .51	.29 <.001	.01 .82
5[e]	.61 <.001	.47 <.001	.61 <.001	.56 <.001	-	.43 <.001	.39 <.001	.36 <.001	.09 .06	−.12 .009	.33 <.001	.04 .35
6[f]	.25 <.001	.18 <.001	.29 <.001	.21 <.001	.48 <.001	-	.57 <.001	.16 .001	.25 <.001	−.25 <.001	.18 <.001	.13 .006
7[g]	.31 <.001	.24 <.001	.36 <.001	.26 <.001	.47 <.001	.63 <.001	-	.23 <.001	.21 <.001	−.11 .02	.12 .01	.07 .16
8[h]	.14 .004	.11 .02	.17 <.001	.14 .003	.41 <.001	.27 <.001	.31 <.001	-	.13 .006	−.01 .94	−.07 .13	−.03 .49
9[i]	.03 52	.02 .70	.03 .50	.02 .70	.26 <.001	.36 <.001	.28 <.001	.20 <.001	-	−.09 .07	.07 .17	.07 .12
10[j]	.02 .65	.05 .28	−.02 .69	.01 .83	−.09 .06	−.29 <.001	−.11 .02	.01 .84	−.02 .69	-	−.05 .34	−.51 <.001
11[k]	.38 <.001	.25 <.001	.32 <.001	.35 <.001	.34 <.001	.13 .008	.08 .10	−.06 .20	.08 .09	.03 .58	-	−.05 .30
12[l]	−.02 .74	.05 .26	−.06 .19	−.01 .97	.02 .69	.16 .001	.03 .56	−.04 .44	.01 .78	−.46 <.001	−.10 .04	-

[a] Alcohol consumption per week.

[b] Alcohol consumption Monday-Tuesday.

[c] Alcohol consumption weekends.

[d] Alcohol consumption last seven days.

[e] Alcohol Use Disorders Identification Test (AUDIT).

[f] Social drinking motives.

[g] Enhancement drinking motives.

[h] Coping drinking motives.

[i] Conformity drinking motives.

[j] Age.

[k] Gender; 1=male, 0=female.

[l] Student; 1=student, 0=non-student.

In Table 3 the intercorrelations separated by completers and non-completers are presented. As evident, the intercorrelations among the great majority of the study variables are similar across completers and non-completers. Overall, completion of the study was significantly negatively related with all 4 alcohol consumption measures (r=−.15 to −.22, P<.001), the AUDIT score (r =−.21, P<.001), all drinking motives except for conformity drinking motives (r =−.08 to r =−.11, P=.001 to .02), gender (r =−.11, P=.001), and status as a student (r =−.12, P<.001).

Four multivariate logistic regression analyses were performed to predict the completion of Web-based alcohol intervention using the variables outlined above. Results across the 4 analyses revealed that alcohol consumption on weekends and in the last seven days, the AUDIT score, and being a student predicted completion of the study (see Table 4). The higher the alcohol consumption on weekends and in the last seven days as well as the AUDIT score, the more likely an individual was to be a non-completer. Students were more likely to complete the study than non-students. All other variables included in the model did not significantly predict study completion. Overall, the regression models showed sufficient goodness of fit.

Table 4. Logistic regression of completion[a] of the Web-based self-affirmation intervention.

Scale	B	SE	*P*	Odds ratio	95% CI for odds ratio
Constant Model 1	.30	.33	.36	1.35	
Alcohol consumption (standard drinks)					
per week	−.02	.01	.08	.98	0.96-1.00
AUDIT[b]	−.06	.02	.02	.95	0.90-0.99
Drinking motives					
social motives	.03	.10	.75	1.03	0.85-1.25
enhancement motives	−.06	.09	.55	.95	0.79-1.14
coping motives	−.07	.10	.46	.93	0.77-1.13
Gender[c]	−.11	.15	.48	.90	0.66-1.21
Student[d]	.72	.21	.001	2.04	1.35-3.11
R^2=.06 (Cox & Snell), .08 (Nagelkerke). Model Model χ^2_8=5.05, *P*=.75. Correct classification 60.7%.					
Constant Model 2	.28	.33	.38	1.33	
Alcohol consumption (standard drinks)					
Monday to Thursday	−.03	.02	.18	.97	0.94-1.01
AUDIT[b]	−.07	.02	.002	.94	0.90-0.98
Drinking motives					
social motives	.04	.10	.69	1.04	0.86-1.26
enhancement motives	−.07	.09	.49	.94	0.78-1.13
coping motives	−.06	.10	.52	.94	0.78-1.14
Gender[c]	−.14	.15	.36	.87	0.65-1.17
Student[d]	.73	.21	.001	2.08	1.34-3.15
R^2=.06 (Cox & Snell), .08 (Nagelkerke). Model χ^2_8=4.36, *P*=.82. Correct classification 51.3%.					
Constant Model 3	.33	.33	.31	1.39	
Alcohol consumption (standard drinks)					
Weekends	−.05	.02	.003	.95	0.92-0.98
AUDIT[b]	−.04	.02	.08	.96	0.92-1.0
Drinking motives					
social motives	.03	.10	.75	1.03	0.85-1.25
enhancement motives	−.04	.10	.67	.96	0.80-1.16
coping motives	−.08	.10	.38	.92	0.76-1.11
Gender[c]	−.11	.15	.46	.89	0.66-1.21
Student[d]	.68	.22	.002	1.97	1.29-2.99
R^2=.07 (Cox & Snell), .09 (Nagelkerke). Model χ^2_8=8.69, *P*=.37. Correct classification 61.0%.					
Constant Model 4	.30	.33	.36	1.35	
Alcohol consumption (standard drinks)					
last seven days	−.02	.01	.05	.99	0.97-1.00
AUDIT[b]	−.06	.02	.007	.94	0.90-0.98
Drinking motives					
social motives	.04	.10	.68	1.04	0.86-1.26

Scale	B	SE	P	Odds ratio	95% CI for odds ratio
enhancement motives	−.06	.09	.52	.94	0.78-1.13
coping motives	−.07	.10	.47	.93	0.77-1.13
Gender[c]	−.11	.15	.46	.89	0.66-1.21
Student[d]	.72	.21	.001	2.05	1.35-3.11
R^2=.06 (Cox & Snell), .08 (Nagelkerke). Model χ^2_8=4.65, *P*=.80. Correct classification 60.3%.					

[a] Completion: 0=non-completers, 1=completers.

[b] AUDIT: Alcohol Use Disorders Identification Test.

[c] Gender: 1=male, 0=female.

[d] Student: 1=student, 0=non-student.

Discussion

Principal Findings

This study found that individuals who reported higher alcohol consumption on weekends and in the last seven days were less likely to complete a Web-based alcohol intervention over and above other predictors, providing support for Hypothesis 1. In addition, higher likelihood of problematic alcohol consumption as indicated by higher AUDIT scores positively predicted the dropout rate over and above other predictors. In line with our Hypotheses 2 and 3, non-completers reported higher levels of social drinking motives compared with completers, and men were more likely to dropout compared with women. However, social drinking motives and gender did not predict completion of the study over and above other predictors, so Hypotheses 2 and 3 have to be rejected. However, we found that students were more likely to complete a Web-based alcohol intervention than non-students.

As assumed in Hypothesis 1, non-completers showed higher levels of alcohol consumption compared with completers. There are several explanations for the higher attrition rate among individuals with higher alcohol consumption and higher AUDIT scores. Participants with higher alcohol consumption and AUDIT scores may have responded with defensive reactivity [41] when confronted with the definition of a standard drink and further information about acceptable and excessive alcohol consumption, which could have led to dissatisfaction with the study in general or the intervention in particular and consequently to a dropout (see also [15]). Another explanation for early termination of the study could be the experience of cognitive dissonance [47] in individuals with higher alcohol consumption and AUDIT scores when learning about the German drinking guidelines and evaluating their drinking against these guidelines. Early termination of study participation could have been one way to resolve their cognitive dissonance. In line with this, participants might also have terminated the intervention as they might have recognized that a reduction of their level of alcohol consumption is not realistic. However, all these explanations remain speculative and can account only in some parts for the finding of a higher attrition rate among individuals with a high alcohol consumption. Therefore and because only very few studies exist that examined reasons for dropout [48], future research should examine in more detail why individuals at risk (especially regarding alcohol consumption) have higher attrition rates.

However, some differences emerged regarding the different measures of alcohol consumption included in this study. Despite significant mean differences between completers and non-completers concerning alcohol consumption per week as well as the alcohol consumption from Monday to Thursday, these two alcohol measures did not significantly contribute to the prediction of the dropout rate in the regression analyses in comparison to other predictors. This provides an important information on the characteristics of a group that might be especially hard to reach: this study's results indicate that it might be especially important to address interventions in young adults engaging in excessive weekend drinking [49]. These results also emphasize the importance of assessing more than a single indicator of alcohol consumption as was done in this study. This is of high importance especially for heavy alcohol drinkers or alcohol abusers. Several explanations have been offered why those individuals tend to underreport their alcohol consumption. Other than forgetting, the way how alcohol consumption was measured might also lead to biased self-reports [50]. To resolve such problems with self-reports, Sobell and Sobell [50] recommend including a minimum set of essential items, such as, usual quantity of drinking. In addition, future studies might also want to complement self-reported alcohol consumption with objective measures such as blood alcohol concentration, the enzymes gamma glutamyl transpeptidase (gamma-GTP), glutamic-oxaloacetic transaminase (GOT), or glutamic-pyruvic transaminase (GPT) [50,51]. However, objective measurements are not suitable for all alcohol-related studies like Web-based surveys or interventions like in this study. The reason is that an objective measurement of alcohol consumption requires a face-to-face contact with participants. In addition, there are also shortcomings of objective measurements. For example, the measurement of gamma-GTP does not guarantee to be a sole indicator of excessive alcohol consumption as an increased serum gamma-GTP activity is also associated with overweight [51]. The same problem of the causal relationship applies to the assessment of the GOT and GPT levels. These objective measures provide information about the severity of the liver disease, which might be a consequence of excessive alcohol consumption [52].

Contrary to Hypothesis 2 drinking motives were not significant predictors of study completion in contrast to other predictors like alcohol consumption as revealed by logistic regression analyses. However, significant mean differences between completers and non-completers emerged regarding drinking motives. As expected, non-completers reported higher social drinking motives compared with completers. In addition, coping and enhancement drinking motives were higher among non-completers [7,53], whereas there was no difference regarding conformity motives. As non-completers in this study are characterised as individuals with high alcohol consumption, this result is in line with research showing that conformity motives are generally not related to alcohol consumption (eg, [54]). Concerning Hypothesis 3, it was found that more men than women dropped out in the current study. This is in line with the result that men reported higher alcohol consumption compared with women as also found in other studies [15,28]. However, as gender did not emerge as a significant predictor in the logistic regression analysis, it appears that alcohol consumption is a more important factor for attrition than gender.

An additional predictor of completing the study was student status. Students were more likely to complete the study as compared with non-students. This result is in contrast to previous research, which did not find a difference regarding the level of education and attrition in Web-based studies [15,55]. However, previous studies also reported that individuals with higher education and higher socioeconomic status are more likely to take part in psychological surveys and interventions (eg, [56,57]). Thus, further research is recommended to focus on the socioeconomic status and the level of education as a predictor of completion of a Web-based intervention; as alcohol-related harms are more likely in individuals with lower educational attainment and or lower socioeconomic status this is a key issue to resolve. Furthermore, it might also be the case that students are more familiar with the participation in Web-based studies. The reason for that might be that numerous studies, lecture evaluations, or surveys via Internet take place at the University of Konstanz. This might enhance the probability that students have lower attrition rates in contrast to non-students. In addition, it might also be the case, that students are more willing to spend time filling out further questionnaires compared with non-students as non-students might be less flexible due to a more stringent working schedule. Furthermore, it might be possible that amazon vouchers that were raffled among participants were more attractive for students due to a tight budget compared with non-students.

Limitations

This study is not without limitations. First, all the information was derived from self-reports, which might cause a bias due to social desirability and memory effects like forgetting [50,58]. Therefore, future research is recommended to use objective measurements (eg, breathalyzer for estimating blood alcohol concentration, see discussion above) although self-reported alcohol consumption has been shown to be reliable and valid [50,59]. However, the validity and reliability depends upon the accurate measure of alcohol consumption. Therefore, standardized alcohol consumption measures should be used. In this study one of the 4 alcohol measures (AUDIT) was a standardized one. Nonetheless, the other measures included in this study followed Sobell and Sobell`s recommendations how to best measure alcohol consumption [50].

A second limitation of the study is that the sample consisted mainly of students (781/898, 86.4%). As outlined above, a greater heterogeneity in socioeconomic status would be desirable to learn more about the characteristics of people at most risk for hazardous alcohol consumption. However, as alcohol consumption in students is a problem per se [7,31,32,60], this study is nonetheless able to make a significant contribution to the identification of risk factors for attrition of alcohol-intervention studies.

Conclusions

Reducing attrition is essential for any intervention. Usually, interventions are designed to reach people who are at the most risk. Unfortunately, as shown by this study, individuals with higher alcohol consumption were more likely to dropout before or at the beginning of the intervention. Thus, it is important to develop and test different strategies to keep participants who are at high risk in Web-based interventions. Only after taking this into account can the effectiveness of Web-based interventions for individuals at high risk be evaluated correctly.

Acknowledgments

We would like to thank all students who worked on this project for their assistance in recruiting participants.

Conflicts of Interest

None declared.

References

1. WHO. Global status report on alcohol and health. Geneva, Switzerland: WHO Press; 2014.
2. Stolle M, Sack P, Thomasius R. Binge drinking in childhood and adolescence: epidemiology, consequences, and interventions. Dtsch Arztebl Int 2009 May;106(19):323-328 [FREE Full text] [doi: 10.3238/arztebl.2009.0323] [Medline: 19547732]
3. Hapke U, V der Lippe E, Gaertner B. [Alcohol consumption, at-risk and heavy episodic drinking with consideration of injuries and alcohol-specific medical advice: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2013 May;56(5-6):809-813. [doi: 10.1007/s00103-013-1699-0] [Medline: 23703502]

4. Miller JW, Naimi TS, Brewer RD, Jones SE. Binge drinking and associated health risk behaviors among high school students. Pediatrics 2007 Jan;119(1):76-85 [FREE Full text] [doi: 10.1542/peds.2006-1517] [Medline: 17200273]
5. Townshend JM, Kambouropoulos N, Griffin A, Hunt FJ, Milani RM. Binge drinking, reflection impulsivity, and unplanned sexual behavior: impaired decision-making in young social drinkers. Alcohol Clin Exp Res 2014 Apr;38(4):1143-1150. [doi: 10.1111/acer.12333] [Medline: 24428268]
6. Cox W, Klinger E. A motivational model of alcohol use: determinants of use and change. In: Cox WM, Klinger E, editors. Handbook of motivational counseling: Concepts, approaches, and assessment. Chichester, UK: John Wiley & Sons; 2004:121-138.
7. Van Damme J, Maes L, Clays E, Rosiers JF, Van Hal G, Hublet A. Social motives for drinking in students should not be neglected in efforts to decrease problematic drinking. Health Educ Res 2013 Mar 13;28(4):640-660. [doi: 10.1093/her/cyt036]
8. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
9. Norman CD, Yip AL. eHealth promotion and social innovation with youth: using social and visual media to engage diverse communities. Stud Health Technol Inform 2012;172:54-70. [Medline: 22910502]
10. 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]
11. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
12. 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]
13. Murray E, Khadjesari Z, Linke S, Hunter R, Freemantle N. Health on the web: randomised trial of work-based online screening and brief intervention for hazardous and harmful drinking. BMC Public Health 2013;13:505 [FREE Full text] [doi: 10.1186/1471-2458-13-505] [Medline: 23706155]
14. Tensil M, Jonas B, Strüber E. Two fully automated web-based interventions for risky alcohol use: randomized controlled trial. J Med Internet Res 2013;15(6):e110 [FREE Full text] [doi: 10.2196/jmir.2489] [Medline: 23742808]
15. Postel MG, de Haan HA, ter Huurne ED, van der Palen J, Becker ES, de Jong CA. Attrition in web-based treatment for problem drinkers. J Med Internet Res 2011;13(4):e117 [FREE Full text] [doi: 10.2196/jmir.1811] [Medline: 22201703]
16. de Weert-van Oene G, Burger H, Grobbee D, Schrijvers A. Identification of patients at high risk of noncompliance in substance dependence treatment. Eur Addict Res 2007 Mar 16;13(2):74-80. [doi: 10.1159/000097936]
17. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries N, Brug J. What makes people decide to visit and use an internet - delivered behavior - change intervention? Health Educ 2009 Oct 16;109(6):460-473. [doi: 10.1108/09654280911001149]
18. Göritz A. Incentives in web studies: Methodological issues and a review. Int J Internet Sci 2006;1:58-70.
19. Jander A, Crutzen R, Mercken L, De Vries H. Web-based interventions to decrease alcohol use in adolescents: a Delphi study about increasing effectiveness and reducing drop-out. BMC Public Health 2015 Apr 9;15(1):340. [doi: 10.1186/s12889-015-1639-z]
20. O'Neil KM, Penrod SD, Bornstein BH. Web-based research: methodological variables' effects on dropout and sample characteristics. Behav Res Methods Instrum Comput 2003 May;35(2):217-226. [Medline: 12834076]
21. Al-Asadi AM, Klein B, Meyer D. Pretreatment attrition and formal withdrawal during treatment and their predictors: an exploratory study of the anxiety online data. J Med Internet Res 2014;16(6):e152 [FREE Full text] [doi: 10.2196/jmir.2989] [Medline: 24938311]
22. Galesic M. Dropouts on the Web. Effects of interest and burden experienced during an online survey. J Off Stat 2006;22:313-328.
23. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
24. Coa K, Patrick H. Baseline motivation type as a predictor of dropout in a healthy eating text messaging program. JMIR Mhealth Uhealth 2016 Sep 29;4(3):e114 [FREE Full text] [doi: 10.2196/mhealth.5992] [Medline: 27688034]
25. Sundström C, Gajecki M, Johansson M, Blankers M, Sinadinovic K, Stenlund-Gens E, et al. Guided and unguided internet-based treatment for problematic alcohol use - a randomized controlled pilot trial. PLoS One 2016 Jul;11(7):e0157817 [FREE Full text] [doi: 10.1371/journal.pone.0157817] [Medline: 27383389]
26. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
27. Bewick BM, West R, Gill J, O'May F, Mulhern B, Barkham M, et al. Providing web-based feedback and social norms information to reduce student alcohol intake: a multisite investigation. J Med Internet Res 2010 Dec 19;12(5):e59 [FREE Full text] [doi: 10.2196/jmir.1461] [Medline: 21169171]

28. Reinwand DA, Crutzen R, Elfeddali I, Schneider F, Schulz DN, Smit ES, et al. Impact of educational level on study attrition and evaluation of web-based computer-tailored interventions: results from seven randomized controlled trials. J Med Internet Res 2015 Oct 07;17(10):e228 [FREE Full text] [doi: 10.2196/jmir.4941] [Medline: 26446779]
29. Edwards AG, Rollnick S. Outcome studies of brief alcohol intervention in general practice: the problem of lost subjects. Addiction 1997 Dec;92(12):1699-1704. [Medline: 9581002]
30. Geraghty AW, Kirby S, Essery R, Little P, Bronstein A, Turner D, et al. Internet-based vestibular rehabilitation for adults aged 50 years and over: a protocol for a randomised controlled trial. BMJ Open 2014 Jul 22;4(7):e005871 [FREE Full text] [doi: 10.1136/bmjopen-2014-005871] [Medline: 25052178]
31. Bewick BM, Trusler K, Mulhern B, Barkham M, Hill AJ. The feasibility and effectiveness of a web-based personalised feedback and social norms alcohol intervention in UK university students: a randomised control trial. Addict Behav 2008 Sep;33(9):1192-1198. [doi: 10.1016/j.addbeh.2008.05.002] [Medline: 18554819]
32. Bewick BM, West RM, Barkham M, Mulhern B, Marlow R, Traviss G, et al. The effectiveness of a Web-based personalized feedback and social norms alcohol intervention on United Kingdom university students: randomized controlled trial. J Med Internet Res 2013 Jul 24;15(7):e137 [FREE Full text] [doi: 10.2196/jmir.2581] [Medline: 23883616]
33. Bingham CR, Barretto AI, Walton MA, Bryant CM, Shope JT, Raghunathan TE. Efficacy of a web-based, tailored, alcohol prevention/intervention program for college students: initial findings. J Am Coll Health 2010;58(4):349-356. [doi: 10.1080/07448480903501178] [Medline: 20159758]
34. Shadish W, Cook T, Campbell D. Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton-Mifflin; 2002.
35. Molenberghs G, Kenward M. Missing data in clinical studies. Chichester: Wiley; 2007.
36. Crutzen R, Viechtbauer W, Spigt M, Kotz D. Differential attrition in health behaviour change trials: a systematic review and meta-analysis. Psychol Health 2015 Jan;30(1):122-134. [doi: 10.1080/08870446.2014.953526] [Medline: 25109224]
37. Hochheimer CJ, Sabo RT, Krist AH, Day T, Cyrus J, Woolf SH. Methods for evaluating respondent attrition in web-based surveys. J Med Internet Res 2016 Nov 22;18(11):e301 [FREE Full text] [doi: 10.2196/jmir.6342] [Medline: 27876687]
38. Schulz DN, Candel MJ, Kremers SP, Reinwand DA, Jander A, de VH. Effects of a Web-based tailored intervention to reduce alcohol consumption in adults: randomized controlled trial. J Med Internet Res 2013;15(9):e206 [FREE Full text] [doi: 10.2196/jmir.2568] [Medline: 24045005]
39. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
40. World MO. Declaration of Helsinki (1964). BMJ 1996;7070:1448-1449.
41. Harris PR, Napper L. Self-affirmation and the biased processing of threatening health-risk information. Pers Soc Psychol Bull 2005 Sep;31(9):1250-1263. [doi: 10.1177/0146167205274694] [Medline: 16055644]
42. Radtke T, Scholz U. Beliefs are not behavior: the distinction between compensatory health beliefs and compensatory health behavior. Zeitschrift für Gesundheitspsychologie 2016 Jul;24(3):119-129. [doi: 10.1026/0943-8149/a000161]
43. van Koningsbruggen GM, Harris PR, Smits AJ, Schu z B, Scholz U, Cooke R. Self-affirmation before exposure to health communications promotes intentions and health behavior change by increasing anticipated regret. Communic Res 2014 Oct 28;43(8):1027-1044. [doi: 10.1177/0093650214555180]
44. IARD. 2016. National drinking guidelines URL: http://www.iard.org/wp-content/uploads/2016/04/Policy-Review-National-Drinking-Guidelines.pdf [accessed 2016-09-06] [WebCite Cache ID 6kJjVuV1n]
45. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: The alcohol use disorders identification test. Guidelines for use in primary care. Switzerland: World Health Organization; 2001.
46. Kuntsche E, Kuntsche S. Development and validation of the drinking motive questionnaire revised short form (DMQ-R SF). J Clin Child Adolesc Psychol 2009 Nov;38(6):899-908. [doi: 10.1080/15374410903258967] [Medline: 20183672]
47. Festinger L. A Theory of Cognitive Dissonance. California: Stanford University Press; 1957.
48. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Jan;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
49. Kuntsche E, Gmel G. Alcohol consumption in late adolescence and early adulthood--where is the problem? Swiss Med Wkly 2013 Jul 25;143:w13826. [doi: 10.4414/smw.2013.13826] [Medline: 23888405]
50. Sobell L, Sobell M. Alcohol consumption measures. In: Allen JP, Wilson VB, editors. Assessing alcohol problems a guide for clinicians and researchers. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2003:75-101.
51. Yokoyama H. [Gamma glutamyl transpeptidase (gammaGTP) in the era of metabolic syndrome]. Nihon Arukoru Yakubutsu Igakkai Zasshi 2007 Jun;42(3):110-124. [Medline: 17665541]
52. Cohen JA, Kaplan MM. The SGOT/SGPT ratio--an indicator of alcoholic liver disease. Dig Dis Sci 1979 Nov;24(11):835-838. [Medline: 520102]
53. Atwell K, Abraham C, Duka T. A parsimonious, integrative model of key psychological correlates of UK university students' alcohol consumption. Alcohol Alcohol 2011;46(3):253-260 [FREE Full text] [doi: 10.1093/alcalc/agr016] [Medline: 21393342]

54. Németh Z, Urbán R, Kuntsche E, San Pedro EM, Roales NJ, Farkas J, et al. Drinking motives among Spanish and Hungarian young adults: a cross-national study. Alcohol Alcohol 2011;46(3):261-269 [FREE Full text] [doi: 10.1093/alcalc/agr019] [Medline: 21398656]
55. Riper H, Kramer J, Conijn B, Smit F, Schippers G, Cuijpers P. Translating effective web-based self-help for problem drinking into the real world. Alcohol Clin Exp Res 2009 Aug;33(8):1401-1408. [doi: 10.1111/j.1530-0277.2009.00970.x] [Medline: 19413646]
56. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. Am Psychol 2004 Mar;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]
57. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. Am Psychol 2004;59(2):105-117. [doi: 10.1037/0003-066X.59.2.105] [Medline: 14992637]
58. Davis CG, Thake J, Vilhena N. Social desirability biases in self-reported alcohol consumption and harms. Addict Behav 2010 Apr;35(4):302-311. [doi: 10.1016/j.addbeh.2009.11.001] [Medline: 19932936]
59. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. Addiction 2003 Dec;98(Suppl 2):1-12. [Medline: 14984237]
60. Mitka M. College binge drinking still on the rise. JAMA 2009 Aug 26;302(8):836-837. [doi: 10.1001/jama.2009.1154] [Medline: 19706853]

Abbreviations

AUDIT: Alcohol Use Disorder Identification Test
DMQ-R SF: Drinking Motive Questionnaire Revised Short Form
T1: Time point 1, one week after baseline
T2: Follow-up, one week after T1

Edited by G Eysenbach; submitted 13.12.16; peer-reviewed by K Donoghue, J Apolinário-Hagen; comments to author 21.01.17; revised version received 14.03.17; accepted 28.03.17; published 28.06.17.

Please cite as:
Radtke T, Ostergaard M, Cooke R, Scholz U
Web-Based Alcohol Intervention: Study of Systematic Attrition of Heavy Drinkers
J Med Internet Res 2017;19(6):e217
URL: http://www.jmir.org/2017/6/e217/
doi:10.2196/jmir.6780
PMID:28659251

Original Paper

Informing the Design of "Lifestyle Monitoring" Technology for the Detection of Health Deterioration in Long-Term Conditions: A Qualitative Study of People Living With Heart Failure

Sarah Hargreaves[1], BA (Hons), MSc, MA, PhD; Mark S Hawley[1], BSc (Hons), PhD; Annette Haywood[2], BSc (Hons), PhD; Pamela M Enderby[1], DSc (Hons), PhD, MSc

[1]Centre for Assistive Technology and Connected Healthcare, School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom
[2]Public Health, School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Mark S Hawley, BSc (Hons), PhD
Centre for Assistive Technology and Connected Healthcare
School of Health and Related Research
University of Sheffield
Innovation Centre
217 Portobello
Sheffield, S1 4DP
United Kingdom
Phone: 44 114 2220682
Fax: 44 114 2220682
Email: mark.hawley@sheffield.ac.uk

Abstract

Background: Health technologies are being developed to help people living at home manage long-term conditions. One such technology is "lifestyle monitoring" (LM), a telecare technology based on the idea that home activities may be monitored unobtrusively via sensors to give an indication of changes in health-state. However, questions remain about LM technology: how home activities change when participants experience differing health-states; and how sensors might capture clinically important changes to inform timely interventions.

Objective: The objective of this paper was to report the findings of a study aimed at identifying changes in activity indicative of important changes in health in people with long-term conditions, particularly changes indicative of exacerbation, by exploring the relationship between home activities and health among people with heart failure (HF). We aimed to add to the knowledge base informing the development of home monitoring technologies designed to detect health deterioration in order to facilitate early intervention and avoid hospital admissions.

Methods: This qualitative study utilized semistructured interviews to explore everyday activities undertaken during the three health-states of HF: normal days, bad days, and exacerbations. Potential recruits were identified by specialist nurses and attendees at an HF support group. The sample was purposively selected to include a range of experience of living with HF.

Results: The sample comprised a total of 20 people with HF aged 50 years and above, and 11 spouses or partners of the individuals with HF. All resided in Northern England. Participant accounts revealed that home activities are in part shaped by the degree of intrusion from HF symptoms. During an exacerbation, participants undertook activities specifically to ease symptoms, and detailed activity changes were identified. Everyday activity was also influenced by a range of factors other than health.

Conclusions: The study highlights the importance of careful development of LM technology to identify changes in activities that occur during clinically important changes in health. These detailed activity changes need to be considered by developers of LM sensors, platforms, and algorithms intended to detect early signs of deterioration. Results suggest that for LM to move forward, sensor set-up should be personalized to individual circumstances and targeted at individual health conditions. LM needs to take account of the uncertainties that arise from placing technology within the home, in order to inform sensor set-up and data interpretation. This targeted approach is likely to yield more clinically meaningful data and address some of the ethical issues of remote monitoring.

(J Med Internet Res 2017;19(6):e231) doi:10.2196/jmir.6931

KEYWORDS

independent living; human activities; heart failure; biomedical technology

Introduction

Background

Health services around the world are facing the challenge of providing care for an increasing number of people living with long-term conditions (LTC) within the community [1-3]. The greatest challenges arise from complex or difficult-to-manage conditions—for example, unstable conditions that are subject to exacerbations [4], complex multiple LTCs [5,6], or circumstances that increase vulnerability, such as old age [5] or living alone [7,8]. Health technologies are being developed to help people living at home manage LTCs. One such technology is "Lifestyle Monitoring" (LM), a telecare technology based on the idea that everyday home activities may be indirectly monitored via sensors to give an indication of the changing health state [9-13]. This remote monitoring technology is designed to unobtrusively monitor participants' home activities over time, with the aim of identifying signs of deterioration as part of the clinical management approaches within a primary or community care setting.

Activity measures are commonly used to assess medical and social needs in older people (eg, [14,15]) by indirectly testing the effects of disease on the ability to perform everyday activities, such as climbing the stairs or preparing a meal [16]. Currently such tests are mostly carried out either by health care professionals observing the performance of activities or by patients self-reporting their abilities. These tests are carried out either on a single occasion or at infrequent intervals. LM technology has been developed to infer home activities on a continuous basis using sensors sited around the home.

Reviews of LM research have highlighted deficits in knowledge [17,18]. Key to this are issues around whether and how home activities change during deteriorations in health. Gokalp et al[18] identified three key areas for further investigation in their review: (1) determining the correlation between changes in activities of daily living and well-being or health status (ie, extracting useful information relevant to subjects' well-being), (2) determining the most effective set of sensors, and (3) the lack of a decision model which explains how changes in monitored activity relate to changes in health, which in turn could inform the actions of health services. This paper focuses on the first two issues: (1) whether and how home activities change during deteriorations in health, and (2) if changes do occur, how these might be captured by sensors sited around the home. The study also informs debate on the third issue raised by the review, relating to the processes of clinical decision-making using information derived from remote monitoring technologies.

Previous LM studies have described various approaches to remotely monitoring home activities with the aim of detecting deteriorations in health. One common method has been to use sensors to create an electronic version of the activities of daily living (ADL) measure within the home, but this method has so far not provided a practical means of detecting clinically meaningful changes in home activity of patients [17,18]. This study approaches this key issue in a different way. We sought to provide a detailed qualitative exposition of whether and how activities change within the home during the various health states of a long-term condition; and through this lived experience, to inform the design of LM technology.

Aims

This paper reports the findings of a study examining everyday activities in people with heart failure (HF) in varying health states in order to identify whether participant activities change in distinctive ways across the health states, and particularly during an exacerbation, in an exemplar long-term condition. We sought to capture details of participants' activities at home at times when they experience these varying health states to inform sensor set-up for remote monitoring and to potentially capture these clinically important changes within the data stream. The study also explored other influences on everyday activity to understand the degree to which activity was influenced by the symptoms of the condition compared to other factors. The research questions were as follows:

1. How do everyday activities undertaken at home by people with HF vary according to the health state of the individual carrying them out?
2. What other factors influence everyday activity in older people?
3. How might changes in home activity be captured indirectly via sensors sited within the home?

All research questions were considered through the lens of LM, the remote monitoring technology, and from the perspective of the lived experience of the long-term condition. The focus of the study was HF, but the aims of the study were broader: to add to the to the knowledge base informing the development of home monitoring technologies designed to detect health deterioration in any long-term condition that is subject to exacerbations.

Within the United Kingdom, approximately 493,000 people are recorded on general practice registers of patients with HF [19]; although, the actual prevalence is likely to be higher. HF is a complex condition with an uncertain trajectory [20-23] and a high symptom burden [24]. The symptoms are known to impact everyday activities (eg, [25-29]) in ways that vary depending on the degree of intrusion from symptoms. For example, Edmonds et al [26] documented changes in activity in response to different degrees of breathlessness. The condition is costly both to the individual and to health services because of the risk of non-elective hospital admission and readmission [30]. It is known that people with HF may delay seeking help during an exacerbation, and therefore, symptoms may deteriorate before health services are alerted [31-33]. One aspect of this delay may arise from difficulties with symptom recognition [32-37]. Early detection of worsening symptoms may potentially halt an exacerbation, allowing the condition to be stabilized via a community health care intervention, thus avoiding the need for hospital admission. This research, therefore, particularly focused on the potential changes in activity leading up to an exacerbation and how these can be detected by sensors.

Methods

Participants and Recruitment

The study participants with HF were recruited in two ways. Participants were either identified by a heart failure specialist nurse and recruited by the research team, or were recruited from a HF support group. All participants lived in and around Barnsley, a town in the north of England.

The sample was chosen to include a diverse experience of living with HF, and in particular, participants varied in their age, severity of HF, and the length of time since diagnosis (Table 1). All participants were aged 50 years and above. Additional insights were gained from the partners of some of the participants with HF, who, at the participants' request, sat in and contributed to the interview.

Study Design

Participants were interviewed in their own homes over an 8-month period using a semistructured schedule both to ensure that key topics were discussed and to capture a broad range of experience. The key topics were identified by literature searching and included a focus on HF symptoms and activities undertaken during days classified as normal days, bad days, and exacerbations [38,39,25].

The audio-recorded interviews were typically of one hour duration, with a minimum of 40 minutes and a maximum of two hours. After each interview, field notes were taken to both reflect on the interview and gain insights into the research questions—these notes provided the starting point for data analysis. The interviews were transcribed in full.

Ethics

Ethical approval was gained from Leeds (East) Research Ethics Committee (08/H1306/46) and research governance from Barnsley Hospital NHS Foundation Trust. Participants were sent written information about the project at the initial point of recruitment. Key points were reiterated before the interview and participants were informed of their right to stop the interview at any time. Written consent was then taken.

Theoretical Framework and Analytical Approach

The project was underpinned by subtle realism: an epistemological stance based on the belief that there are multiple non-contradictory versions of reality and that although accounts differ, they all reflect reality [40]. Researchers sought to cast a critical eye over participants' accounts of life with HF to find a version "...of whose validity we (were) reasonably confident" ([41], page 50). Researchers were mindful of their own influence on this process and sought, through reflexivity, to identify any erroneous assumptions, attitudes, or beliefs in a process described by O'Cathain [42] as "situating" yourself within the research.

The interview data were analyzed using template analysis [43-45], a form of thematic analysis where the key tool is the use of a hierarchical template to summarize themes and provide a visual aid to both analysis and data interpretation. The hierarchical structure of the template is used to organize themes in a meaningful way, with key broad concepts captured in higher level themes, and lower level themes providing more detail to improve understanding of specific aspects of the broader themes.

Data Analysis

The data were coded by Sarah Hargreaves, with checks made by other researchers to validate choice of codes. A minority of initial themes were informed by the literature (a priori themes); however, these were retained only if the theme was found to be both present within the data and of relevance to the research question. If not, a priori themes were amended to reflect the data or discarded if they were found to be no longer appropriate. Other themes were identified from "features" of participant accounts (perceptions or experiences) relevant to the research questions. Initial theoretical propositions were tested at a validation exercise undertaken at a heart failure support group meeting with 27 attendees. Themes were further consolidated via the process of template development whereby different template structures were explored in order to reach a final version that both represented the themes within the interviews and directly addressed the research questions. Consider, for example, the high level theme "Deciding what to do today," which arose out of an understanding from the interviews that activities on any given day are influenced by a wide range of factors. These were detailed in lower-level codes, such as the weather, commitments, and activity assessment. The understanding of this theme was consolidated by the validation exercise, which highlighted not only the importance of symptoms as a key influence on daily activity but also individual differences. This process of template development was documented in order to create an audit trail [45].

The data were initially coded manually and at a later stage imported into NVivo 9 (QSR International) in order to utilize the text retrieval facilities to aid template development and for reporting purposes.

In this paper, all participants are referred to by pseudonyms.

Results

Participant Characteristics

Table 1 summarizes the characteristics of the interview participants with HF. Participants ranged in age, from younger participants in their 50s to older participants in their 80s, with the majority in their 70s. In addition, many participants suffered from a range of co-morbidities from mild conditions (such as mild arthritis) to potentially life-threatening ones (eg, cancer). The sample included participants who had been diagnosed with HF recently (within the last three years), and those who had been living with HF for longer. Despite efforts to recruit a gender-balanced sample, the sample was predominantly male.

The partners or spouses of 11 interviewees sat in and contributed to the interviews.

Table 1. Characteristics of participants with heart failure.

NYHA heart failure class[a]	Gender	Length of time since diagnosis[b]		Grand total
		Within the last 3 years	>3 years	
I	Male	1		1
II	Female	1		1
	Male	5	1	6
III	Female	1	1	2
	Male	1	6	7
IV	Female	1		1
	Male	1	1	2
Grand total		11	9	20

[a] Classified according to participant accounts of the impact of symptoms on everyday activity and interviewer observations [38].

[b] This information was derived from participant accounts.

During the analysis an integrative theme was identified [44]. This is a theme which King describes as "undercurrents running through participants' accounts" (page 460), and is thus pertinent to the interpretation of much of the data. In this study the integrative theme was the degree of intrusion from symptoms.

The Three Health States of Heart Failure and Impacts on Everyday Activity

The interviews confirmed that participants recognized the concept of normal and bad days and were able to describe how these two types of days differed both in terms of the degree and type of symptoms experienced, and the sort of activities that they would be likely to undertake. A normal or good day was relatively symptom-free, and this gave participants more energy to undertake everyday activities, which might be within the home. For some, this feeling of well-being inspired participants to undertake activities in the garden or to go out.

> *Well on a good day I live like a normal day virtually, don't I?...Erm if I want to go for a walk, we go for a walk. We go shopping, whatever, visit people.* [Brian]

How often these good times occurred, how long they lasted, and the degree of extra energy experienced, all varied depending on the severity of HF. For some, these good times were both infrequent and transient.

In contrast bad days were characterized by a lack of energy ("a whole system tiredness"), a lethargy, and with it an accompanying feeling of listlessness, irritation, and negativity. During such times participants preferred to stay at home and may have undertaken gentle activities or those requiring very little energy, such as watching the television or using the computer. Again the degree of intrusion from the symptoms varied, with some participants very badly affected by bad days and unable do anything, while others ignored their symptoms and kept going.

> *You can have days when you feel really down and you know it just seems like you are useless, you can't do anything and you don't have the energy to get a meal or anything* [Dawn]

Bad days were regarded as a transitional state to be endured in order to reach better times.

The third health state was exacerbation. Not all the participants had experienced an exacerbation, but for those who had, it was remembered with horror—as a time when health concerns came to the fore and everything became a struggle. The key symptoms were breathlessness, tiredness, and swelling (oedema), with breathlessness especially reaching frightening levels as participants described struggling for breath.

> *...Just if you were walking up a little incline erm that was something or nowt [nothing] really you had to stop and you would be going [participant acts out doing heavy gasping breaths, followed by a deep lurching breath] and feel as if you couldn't get enough in.* [Larry]

Exacerbations differed from bad days because of the degree of exercise intolerance and dyspnoea, and the perception that the body was out of control, trapped in a "downward spiral."

The impact on home activities built up over time, with participants describing how they spent more time at home, and even there, were limited and selective in their activities. A notable aspect of activity within this health state was slowing down, when activities were undertaken very slowly, especially more exerting activities, such as walking and climbing the stairs.

> *Erm but just everything was just so slow, so very, very slow.* [Willow]

The exacerbation also changed activities at night, with sleep broken by episodes of breathlessness which prompted the participants to undertake activities to ease their breathing, such as sitting on the edge of the bed or going downstairs to sit the night out on the sofa.

> *I come down and was watching sport two, three, four o'clock in the morning, and was dozing on the settee. You know awful, awful feeling when you waken in the night-time like that.* [Larry]

A number of participants spontaneously described how the first thing that alerted them to the onset of exacerbation was a new or increased difficulty in walking or using the stairs.

> *I had been having this increasing breathlessness and then the snow, I couldn't go further than five yards with a bag and I had to stop myself and think this is ridiculous.* [Janice]
>
> *It was walking really. I had always been a fast walker and I had to slow down.* [Willow]

During an exacerbation of symptoms, walking became an ordeal as the physical effort caused breathlessness. The length of an acceptable walk was scaled down, and the activity was carried out more slowly than normal with stops to breathe. This change was observed both inside the home and outside.

> *...from the top of the stairs into the toilet I was absolutely shattered. From there to the chair I was shattered, from that chair to the end of the settee there [a few steps] and I would have to stand [sound of panting] out of breath completely.* [Eric]
>
> *...Just walking to the bottom of the street really slowly and if other people saw me that know me, you should have seen the look on their faces to say what's the matter with you?* [Larry]

As the exacerbation progressed, participants described how activity became pared down to absolute necessities and they avoided climbing the stairs as it became too exhausting and distressing.

Other Factors Influencing Activity

The interviews explored the broad experience of living with HF, and the accounts highlighted how participants were chiefly concerned with living their lives in spite of HF.

> *Err...I don't see myself as being bad, poorly, ill. Whatever you want the classification to be. I don't. Well it is just me, I try not to see that part, I think I am up and running, I am okay.* [Alistair]

This stoical attitude presents a challenge to the concept of LM because individuals with HF may not necessarily respond to intrusive symptoms by changing activity. The most extreme example of this came from a participant who, while suffering an exacerbation, ignored his symptoms and attempted to carry on.

> *I always remember that the Doctor said to me…he said you are not working Mr Smith are you? Well I have been to the gym this morning, I have been to work all day, I just had a shower and I have run down here. He said you could have collapsed and died, this was the cardiologist.* [Larry]

Participant accounts gave an understanding of how everyday activity is influenced by both the individual situation of the participant (eg, severity of HF and co-morbidities, age, personality, gender, lifestyle, attitude to activity), and external factors (eg, housing, local geography, transport availability, weather, demand on time) A good example of this came from Janice, a participant juggling the demands of bringing up her children while living with HF.

> *What I think I should do is put an Agatha Christie on the DVD, have someone bring you afternoon tea and put your feet up…But life's not like that.* [Janice]

The interviews illustrate that in real life, the relationship between activity and health is not clear cut, thus LM systems would need to take account of this complexity and individual circumstances.

Potential Targets for Activity Monitoring to Detect Signs of Exacerbation

Participant accounts verified that an exacerbation is a distinctively different health state, and they also identified activities that were undertaken in order to ease symptoms. Participants described how symptoms built up over time during the exacerbation, and often they did not seek medical help until the symptoms were intolerable —for some in spite of their close family urging them to seek help.

The accounts of the experience of activities undertaken during different health states could provide the starting point for developing a platform of LM sensors to monitor home activities linked to an exacerbation of symptoms. Table 2 lists activities which could potentially be monitored as well as insights into how these changes might be monitored. During an exacerbation of symptoms, participants with HF, for example, described both changes to existing activities (eg, reduced speed in walking and climbing the stairs, staying at home more, and being more sedentary) and undertaking new activities in response to distressing symptoms (eg, sitting up at night to aid breathing, perhaps watching television).

The interviews also highlighted the importance of signs of ill-health relating to appearance, which would require new sensors to be developed. An important sign was swelling, with participants describing the extent of swelling to their body during an exacerbation, as described by Willow:

> *My legs were very swollen, if I sat down I needed to get them up and Edward used to have to lift them up because they were just so heavy, I couldn't lift them up.*

This swelling was reported in feet, legs, the abdomen, and even fingers. More subtle signs noted by the partners of participants with HF were changes in facial appearance including expression, appearance of eyes, blue lips, and alteration in mood and demeanor. In addition, during ill-health, the role of partners changed and they undertook more activities to support their spouse.

Table 2. Activities that change during an exacerbation and insights into how these changes might be monitored.

Home activity	Details of activity changes taking place within the home when health declines	How these changes might be monitored
Climbing the stairs	Stairs became more of a physical challenge during an exacerbation, and the climb is likely to be undertaken more slowly (the individual may also rest during the climb), and over time the activity is likely to be undertaken less frequently than usual or avoided, as symptoms spiral out of control.	A means of counting both the number and duration of stair ascents and descents should be sought—potentially with a device recording the time spent climbing from the first to the last step.
Dressing	Participants were aware of shoes or clothes becoming tighter. Dressing becomes more of a physical effort, especially putting on socks.	Some means of reporting that clothes or shoes have become tighter would be a useful indicator of increased fluid build-up. It would be important to find out the order in which parts of the body become swollen, as the fluids build up; the normal pattern in individuals should also be established.
Food preparation	Food preparation is simplified or avoided.	Information about normal food preparation methods would need to be sought in order to monitor and measure the duration of these tasks, as during periods of ill-health participants described shorter and less elaborate cooking methods. Microwave usage may potentially increase during periods of ill-health, in order to reheat previously prepared meals, or shop-bought ready meals; this could be measured using an electrical socket device.
Staying at Home	A reduction in leaving the home, and then cessation. Participants withdrew from social activities as their health declined.	A key sensor could be used to monitor when the door used to go out is locked from outside (this may differ from the door used to access the garden). The key sensor could provide an indirect means of measuring social isolation, as people may cancel social activities. A means of reporting the scale of breathlessness during speech could provide an indicator of worsening breathlessness.
Increase in sedentary activities	An increase in sedentary activities in the home, such as sitting on the sofa; as health declines, chores are likely to be limited to necessities, and activities requiring physical exertion avoided	See “sleeping on the chair/sofa” below
Activities undertaken to ease breathlessness	Some people seek air during episodes of breathlessness, opening windows or doors, using a fan, or sitting upright.	Windows and doors opening and closing could be monitored with a door sensor, with openings that occur at night potentially being of greater significance. Window monitoring would only be of value if the window was generally closed.
Sleeping in the bed	Nocturnal sleep may be broken by episodes of breathlessness, requiring the participant to sit upright, or sit on the edge of the bed to recover breath	Mattress monitors are currently used to measuring bed occupancy, although they would not be able to detect subtle signs of ill-health, such as more restless sleep. A means of monitoring the duration of time spent sitting on the edge of the bed is needed.
Sleeping on the Chair/sofa	When breathlessness is bad, some people resort to sleeping upright in a chair, or they may go down stairs in the night to sit out the night on the sofa	Information about where people sleep when they feel unwell should be sought in order to monitor use of this alternative sleeping place, as participants commonly mentioned sleeping on the sofa or chair in the lounge during episodes of nocturnal breathlessness. This measure of chair usage would be more reliable if the chair was used exclusively by the participant; and this is not unrealistic, since some participants had an upright chair that they favored.
Signs noted by partners: Symptoms and Subtle signs	Partners were aware of a deterioration of symptoms, such as increased breathlessness, and swelling, and began to help their partners undertake activities. People talk less during a decline of health, and withdraw into self. A change of appearance of the face, eyes, and demeanour was noted by partners during an exacerbation.	Some means should be provided for partners/family to report observations about the appearance of their relative with a long-term condition. Partners were very in tune with their partner’s health, and thus this valuable information should be fed into the monitoring.
Television viewing	Viewing may increase during an exacerbation, with an increase in viewing during the day, and viewing at unusual times (such as in the night) may be a particular indicator	Information about television viewing would need to be sought before monitoring, as people may always have the television on as background noise, or undertake other activities during periods of ill-health—for example, listening to the radio more. A television electrical socket device could record the duration of time the television is on, but not whether it is actually being watched. A measure of night-time viewing may be an important indicator of a change in health.
Movement around the home	Walking becomes more of a physical challenge during an exacerbation, and is undertaken more slowly and less frequently.	The number of daily steps could be recorded via a pedometer, and speed recorded via an accelerometer

Discussion

Principal Findings

The basis of LM technology is that the home activities of people at risk of an exacerbation of a long-term condition may be monitored via sensors in order to provide an indication of important changes in health. This study found that a sensor platform could potentially be tailored to identify signs of an exacerbation in HF when the setup is informed by understanding the lived experience of HF, and in particular the impact of intrusion of HF symptoms on everyday activity. We argue that this emphasis on understanding the lived experience of long-term conditions is important both for identifying which health conditions may be appropriate for LM and for tailoring the sensor setup to identify clinically important changes in health. While this study focuses on HF, the approach has a far wider application for the development of approaches to remote monitoring and data analysis in long-term conditions.

The findings suggest the need for a more personalized and targeted approach to LM to tailor sensor setup to monitor activities which are potential indicators of important changes in health. By exploring home activities during the different health states of HF, we were able to identify activities which should be monitored, and from this, choices could be made about identifying appropriate means of detecting change. This requirement to identify specific activities that are sensitive to changes in symptoms (specific to a health condition) has not been addressed by the LM literature, where descriptions of systems, methods, and the context in which they are being used, tend to be vague [17].

Although it was known that people with health conditions change their behavior during an exacerbation (eg, during an exacerbation people with HF may sleep upright), detailed description of how these changes occur within the home were lacking [46]. A key finding from this study is the detailed list of changes that occurred during exacerbations (Table 2) and this offers useful information for the developers of sensors and platforms, and to inform the development of algorithms to detect changes in the data collected. This key finding adds to insights gleaned from other studies looking at varied long-term conditions—Clarke, for example [46].

This targeted approach offers the potential for a reduction in the number of sensors targeted at specific activities, which may provide more clinically meaningful data than a standard broad-brush setup [47]. It may also lead to the development of sensors designed specifically to monitor activities of clinical significance. LM has been criticized for the failure to capture clinically significant data, and for Elwyn et al [48], this arises from the focus on data that is easy to collect rather than data that is indicative of a change in health. It is anticipated that LM data could inform clinical management of patients living within the community, offering a potential safety net, so that primary care services could be alerted to signs of decline in order to intervene early and potentially prevent hospital admission. In our study, participants delayed seeking help for an exacerbation, a finding in keeping with other HF studies [31-33]. A LM service may be of great value to patients, given that it would require no effort from the patient, and may reduce some of the uncertainty and complexity of living with unpredictable conditions.

An approach focused on collecting data of clinical significance is more ethical, as it is more likely to yield data of value to health professional decision making, leading to interventions of clinical benefit [17,48]. In addition, the creation of a bespoke LM setup for a specific health condition, which also meets the needs of users, is more likely to yield positive benefit for the user and inspire confidence in health professionals. This is a vital step in the wider adoption of LM [49].

It is likely that LM is more appropriate for some health conditions than others. This study identified some aspects of HF that make it more appropriate for LM. First, the health states within the condition were found to be distinctly different, and in particular there were specific activities found to be undertaken during an exacerbation to ease symptoms, as well as broader changes in activity that occurred (such as staying at home more) during exacerbations. In addition, the exacerbation was found to build up over time, and this would allow sufficient time to identify a change in health and intervene. Participants of the study were able to identify the first signs of an exacerbation, and this could inform the sensor setup. Although taking a bespoke approach to LM design should make it applicable to other long-term conditions, the characteristics of other conditions may make a LM approach less likely to be successful. The buildup to an exacerbation in chronic obstructive pulmonary disease (COPD) is much faster (eg, [50]). A COPD telecare study was unsuccessful in identifying the early stages of an exacerbation due to a lack of early predictors [51], and the authors speculated that the system may have been identifying signs of a bad day rather than an exacerbation.

Another potential means of enhancing the remote monitoring would be to feed in additional information. The study found that the partners of people with HF were undertaking an informal monitoring role, that is, observing their partner day and night for signs and symptoms that their health was deteriorating. This monitoring included aspects which current sensors would not be able to capture, such as observing changes in the appearance of their partner, their symptoms (observing increased breathlessness, for example), mood, and noting additional activities they were required to undertake. Other studies have noted this monitoring role as well [25]. Such observations could potentially be clinically significant, or highlight a need for closer examination by a health professional [52]. A means of feeding in this information would add an extra valuable dimension to the resources for decision making [53]. There are, however, ethical implications and potential negative impacts on family relationships arising from this approach which would need further exploration.

Remote monitoring of HF typically focuses on physiological changes [54]. Utilizing technologies that measure physiological changes in parallel with LM provides opportunity to assess the respective accuracy of both data sources and adds an additional depth of information [47]. It would also provide an additional safety net for patients. The first change in activity noted by participants of this study during an exacerbation was a slowing of their walking pace or in climbing the stairs, but they were also aware of increased breathlessness. Unlike physiological measurement technologies, LM technologies require no active engagement from those being monitored, and thus may detect exacerbations in those people unlikely to actively engage with monitoring. The continuous nature of LM may also lead to earlier detection of the beginning of an exacerbation. Future research might focus on how measures of physiological change and activity vary across the timeline of an exacerbation within the complexity of the lived experience of HF.

Complexity arises from placing technology within the home, where the monitoring interacts with the individual circumstances in which a person lives and the external factors impacting on everyday life. The health state exists not as a separate entity, which directs home activity, but within a complex environment, where factors interact and there may be conflicting demands. This presents a challenge to LM because an individual may feel ill but not attend to those feelings by resting or staying at home, as their lifestyle has other demands. It is well known that people do not respond uniformly to the challenges of ill health [55-58]; and that people with health issues may prefer to focus on living their lives, rather than the health condition [55,58,59]. This means that there will be a level of uncertainty in the data recorded by LM sensors. However we argue that by understanding more about an individual and their attitudes, priorities, circumstances, and routine, this would provide a context within which to both tailor sensor setup to fit in with individual circumstances and behaviors, and also a means of interpreting the LM data in order to reduce the margin of error [60]. We argue that this understanding is important in order to establish a link between the sensor data and the person being monitored. This would require a degree of personalization of sensor setup and interpretation, which would involve additional costs [61]; it may also involve the development of new sensors. This requirement for technologies to be designed "in the wild," taking account of how real people live in their homes is an important element in designing technologies to fit needs [62].

Limitations

The participants recruited to this study were varied in terms of age and HF severity, and therefore may not be representative of the typical elderly HF patient cared for by general practice [63]. This study focused entirely on HF, but for some the situation is complicated by co-morbidities [30], and the presence of these is likely to make LM more challenging. The small sample size is a limitation of the study, as it is unlikely that we have captured all aspects of behavior within the home in individuals with HF.

There was an imbalance within the sample, with a greater proportion of male participants and few participants who lived alone. In addition, the study participants were all from a similar geographical area within Northern England and were from the same ethnic grouping (white British), and this may have in part influenced the way that participants responded to the challenges of living with a long-term condition. Certainly participants were stoical in their approach to living with HF, which may reflect the culture of an area with a history of coal-mining and steel-working. This stoical approach to living with a long term condition has been noted by another study based in Northern England [64].

It is acknowledged that there is a risk that the presence of partners may have influenced the interviews (eg, [65-67]). However, we argue that the presence of partners was positive as they were able to provide vivid descriptions of the impact of HF on activity and to act as a counterbalance to the instinct of many of the participants with HF to underplay symptoms.

Conclusions

This study shows the importance of understanding of the lived experience of long-term conditions in order to potentially identify which conditions are suitable for LM and of detailed information about how home activities change when health deteriorates. This detailed information is the key to designing a sensor setup tailored to both specific health conditions and individual circumstances and to inform data analysis and interpretation. By this means it is anticipated that LM can provide timely, clinically meaningful data that will contribute to the management of long-term conditions of people living in the community.

Acknowledgments

The Authors wish to thank Professor Nigel King from the University of Huddersfield for his expert advice on template analysis. Dr Hargreaves was funded by Barnsley Hospital NHS Foundation Trust.

Conflicts of Interest

None declared.

References

1. Campbell D. The Guardian. 2014 Jan 03. NHS could be 'overwhelmed' by people with long-term medical conditions URL: https://www.theguardian.com/society/2014/jan/03/nhs-overwhelmed-long-term-medical-conditions [accessed 2017-06-05] [WebCite Cache ID 6qzVnpyCk]
2. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. The Lancet 2009 Oct;374(9696):1196-1208. [doi: 10.1016/S0140-6736(09)61460-4]
3. World Health Organization. WHO. Geneva: World Health Organization; 2011. Global status report on noncommunicable diseases 2010: description of the global burden of NCDs, their risk factors and determinants URL: http://www.who.int/nmh/publications/ncd_report2010/en/ [accessed 2017-06-05] [WebCite Cache ID 6qzrfeSPf]

4. Department of Health, Social Services and Public Services. Health-ni. 2012. Living with long term conditions policy framework URL: https://www.health-ni.gov.uk/publications/living-long-term-conditions-policy-framework [accessed 2017-06-05] [WebCite Cache ID 6qzs0kWbK]
5. Department of Health. Gov.uk.: Department of Health; 2012. Long term conditions compendium of information, third edition URL: https://www.gov.uk/government/publications/long-term-conditions-compendium-of-information-third-edition [accessed 2017-06-05] [WebCite Cache ID 6lG4U6qdi]
6. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet 2012 Jul;380(9836):37-43. [doi: 10.1016/S0140-6736(12)60240-2]
7. Kharicha K, Iliffe S, Harari D, Swift C, Gillmann G, Stuck AE. Health risk appraisal in older people 1: are older people living alone an “at-risk” group? Br J Gen Pract 2007;57(537):271-276.
8. Haslbeck JW, McCorkle R, Schaeffer D. Chronic illness self-management while living alone in later life: a systematic integrative review. Res Aging 2012 Jan 10;34(5):507-547. [doi: 10.1177/0164027511429808]
9. Celler BG, Hesketh T, Earnshaw W, Ilsar ED. An instrumental system for the remote monitoring of changes in functional health status of the elderly at home. 1994 Presented at: Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society - Engineering Advances: New Opportunities for Biomedical Engineers; 1994; Baltimore, MD p. 908-909.
10. Celler BG, Lovell NH, Hesketh T, Ilsar ED, Earnshaw W, Betbeder-Matibet L. Remote home monitoring of health status of the elderly. Medinfo 1995;8 Pt 1:615-619. [Medline: 8591281]
11. Kaushik AR, Celler BG. Characterization of PIR detector for monitoring occupancy patterns and functional health status of elderly people living alone at home. Technol Health Care 2007;15(4):273-288. [Medline: 17673836]
12. Kaushik AR, Lovell NH, Celler BG. Evaluation of PIR detector characteristics for monitoring occupancy patterns of elderly people living alone at home. 2007 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2007; Lyon, France p. 3802-3805.
13. Ní Scanaill SC, Carew S, Barralon P, Noury N, Lyons D, Lyons GM. A review of approaches to mobility telemonitoring of the elderly in their living environment. Ann Biomed Eng 2006;34(4):547-563. [Medline: 16550450]
14. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963;185:914-919.
15. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]
16. Bowling A. Measuring Health: A Review of Quality of Life Measurement Scales. Buckingham: Open University Press; 1997.
17. Brownsell S, Bradley D, Blackburn S, Cardinaux F, Hawley MS. A systematic review of lifestyle monitoring technologies. J Telemed Telecare 2011 Apr;17(4):185-189. [doi: 10.1258/jtt.2010.100803] [Medline: 21508080]
18. Gokalp H, Clarke M. Monitoring activities of daily living of the elderly and the potential for its use in telecare and telehealth: a review. Telemed J E Health 2013 Dec;19(12):910-923. [doi: 10.1089/tmj.2013.0109] [Medline: 24102101]
19. Townsend N, Bhatnagar P, Wilkins E, Wickramasinghe K, Rayner K, British Heart Foundation Centre on Population Approaches for Non-Communicable Disease Prevention, Nuffield Department of Population Health, University of Oxford. Cardiovascular disease statistics.: British Heart Foundation; 2015. URL: https://www.bhf.org.uk/publications/statistics/cvd-stats-2015 [accessed 2016-10-14] [WebCite Cache ID 6lG5EHPpx]
20. Jones AM, O'Connell JE, Gray CS. Living and dying with congestive heart failure: addressing the needs of older congestive heart failure patients. Age Ageing 2003;32(6):566-568. [Medline: 14599995]
21. Murray SA, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. BMJ 2005 Apr 30;330(7498):1007-1011 [FREE Full text] [doi: 10.1136/bmj.330.7498.1007] [Medline: 15860828]
22. Witham MD, Argo IS, Johnston DW, Struthers AD, McMurdo ME. Predictors of exercise capacity and everyday activity in older heart failure patients. Eur J Heart Fail 2006 Mar;8(2):203-207 [FREE Full text] [doi: 10.1016/j.ejheart.2005.03.008] [Medline: 16048742]
23. Yu DS, Lee DT, Kwong AN, Thompson DR, Woo J. Living with chronic heart failure: a review of qualitative studies of older people. J Adv Nurs 2008 Mar;61(5):474-483. [doi: 10.1111/j.1365-2648.2007.04553.x] [Medline: 18261056]
24. Zambroski CH, Moser DK, Bhat G, Ziegler C. Impact of symptom prevalence and symptom burden on quality of life in patients with heart failure. Eur J Cardiovasc Nurs 2005 Sep;4(3):198-206. [doi: 10.1016/j.ejcnurse.2005.03.010] [Medline: 15916924]
25. Clark AM, Reid ME, Morrison CE, Capewell S, Murdoch DL, McMurray JJ. The complex nature of informal care in home-based heart failure management. J Adv Nurs 2008 Feb;61(4):373-383. [doi: 10.1111/j.1365-2648.2007.04527.x] [Medline: 18234035]
26. Edmonds PM, Rogers A, Addington-Hall JM, McCoy A, Coats AJ, Gibbs JS. Patient descriptions of breathlessness in heart failure. Int J Cardiol 2005 Jan;98(1):61-66. [doi: 10.1016/j.ijcard.2003.10.039] [Medline: 15676168]
27. Evangelista LS, Dracup K, Doering LV. Treatment-seeking delays in heart failure patients. J Heart Lung Transplant 2000 Oct;19(10):932-938. [Medline: 11044686]

28. Jeon Y, Kraus SG, Jowsey T, Glasgow NJ. The experience of living with chronic heart failure: a narrative review of qualitative studies. BMC Health Serv Res 2010 Mar 24;10:77 [FREE Full text] [doi: 10.1186/1472-6963-10-77] [Medline: 20331904]
29. Pattenden JF, Roberts H, Lewin R J P. Living with heart failure; patient and carer perspectives. Eur J Cardiovasc Nurs 2007 Dec;6(4):273-279. [doi: 10.1016/j.ejcnurse.2007.01.097] [Medline: 17383236]
30. Donkor A, Cleland J, McDonagh T, Hardman S, Aminat SA, the National Heart Failure Audit Steering Group Committee. UCL. 2015. National heart failure audit: April 2014 to March 2015 URL: https://www.ucl.ac.uk/nicor/audits/heartfailure/documents/annualreports/heartfailurepublication14_15 [accessed 2017-06-05] [WebCite Cache ID 6lG5cZlbd]
31. Clark AM, Freydberg CN, McAlister FA, Tsuyuki RT, Armstrong PW, Strain LA. Patient and informal caregivers' knowledge of heart failure: necessary but insufficient for effective self-care. Eur J Heart Fail 2009 Jun;11(6):617-621 [FREE Full text] [doi: 10.1093/eurjhf/hfp058] [Medline: 19414477]
32. Horowitz CR, Rein SB, Leventhal H. A story of maladies, misconceptions and mishaps: effective management of heart failure. Soc Sci Med 2004;58(3):631-643. [Medline: 14652059]
33. Riegel B, Dickson VV, Cameron J, Johnson JC, Bunker S, Page K, et al. Symptom recognition in elders with heart failure. J Nurs Scholarsh 2010 Mar;42(1):92-100. [doi: 10.1111/j.1547-5069.2010.01333.x] [Medline: 20487191]
34. Carlson B, Riegel B, Moser DK. Self-care abilities of patients with heart failure. Heart Lung 2001;30(5):351-359. [doi: 10.1067/mhl.2001.118611] [Medline: 11604977]
35. Gallagher R. Self management, symptom monitoring and associated factors in people with heart failure living in the community. Eur J Cardiovasc Nurs 2010 Sep;9(3):153-160. [doi: 10.1016/j.ejcnurse.2009.12.006] [Medline: 20093092]
36. Giamouzis G, Kalogeropoulos A, Georgiopoulou V, Laskar S, Smith AL, Dunbar S, et al. Hospitalization epidemic in patients with heart failure: risk factors, risk prediction, knowledge gaps, and future directions. J Card Fail 2011 Jan;17(1):54-75. [doi: 10.1016/j.cardfail.2010.08.010] [Medline: 21187265]
37. National Collaborating Centre for Chronic Conditions (UK). NICE. London: Royal College of Physicians (UK); 2010. Chronic heart failure: national clinical guideline for diagnosis and management in primary and secondary care URL: https://www.nice.org.uk/guidance/cg108/evidence/full-guideline-pdf-136060525 [accessed 2017-06-05] [WebCite Cache ID 6qzxw6idW]
38. British Heart Foundation. London: British heart foundation URL: https://www.bhf.org.uk/publications/heart-conditions/an-everyday-guide-to-living-with-heart-failure [accessed 2016-10-14] [WebCite Cache ID 6lG5lzqdM]
39. Charmaz K. Good Days, Bad Days: The Self in Chronic Illness and Time. New Brunswick, NJ: Rutgers University Press; 1991.
40. Murphy E, Dingwall R. The ethics of ethnography. In: Atkinson P, Coffey A, Delamont S, Lofland J, Lofland L, editors. Handbook of ethnography. London: Sage Publications; 2007:339-351.
41. Hammersley M. What's Wrong With Ethnography. Routledge: London; 1992.
42. O'Cathain A. Reporting Mixed Methods Projects. In: Andrew S, Halcomb EJ, editors. Mixed methods for nursing and the health sciences. Oxford: Wiley-Blackwell; 2009:135-138.
43. Crabtree BJ, Miller WL. Using codes and code manuals. In: Crabtree BL, Miller WL, editors. Doing Qualitative Research. Thousand Oakes, CA: Sage Publications; 2012:163-178.
44. King N. Doing template analysis. In: Symon G, Cassell C, editors. Qualitative Organizational Research: Core Methods and Current Challenges. London: Sage; 2012:426-450.
45. King N. School of Human Health Sciences, University of Huddersfield. Template analysis URL: http://www-old.hud.ac.uk/hhs/research/template-analysis/ [WebCite Cache ID 6lG60MKUk]
46. Clarke M. The need for an integrated approach to remote monitoring of physiological data and activity data. J Telemed Telecare 2014 Apr;20(3):159-160. [doi: 10.1177/1357633X14529245] [Medline: 24643951]
47. Glascock AP, Kutzik DM. Embrace the Chaos, It's Not Noise: lessons learned from non-traditional environments. IJMHCI 2009;1(2):22-36. [doi: 10.4018/jmhci.2009040102]
48. Elwyn G, Hardisty AR, Peirce SC, May C, Evans R, Robinson DK, et al. Detecting deterioration in patients with chronic disease using telemonitoring: navigating the 'trough of disillusionment'. J Eval Clin Pract 2012 Aug;18(4):896-903. [doi: 10.1111/j.1365-2753.2011.01701.x] [Medline: 21848942]
49. Barrett D, Thorpe J, Goodwin N. Examining perspectives on telecare: factors influencing adoption, implementation, and usage. Smart Homecare Technol Telehealth 2015;3:1-8. [doi: 10.2147/SHTT.S53770]
50. Hardisty AR, Peirce SC, Preece A, Bolton CE, Conley EC, Gray WA, et al. Bridging two translation gaps: a new informatics research agenda for telemonitoring of chronic disease. Int J Med Inform 2011 Oct;80(10):734-744. [doi: 10.1016/j.ijmedinf.2011.07.002] [Medline: 21890403]
51. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. BMJ 2013 Oct 17;347:f6070. [doi: 10.1136/bmj.f6070]
52. Royal College of Nursing. RCNHCA. First steps for healthcare assistants URL: http://rcnhca.org.uk/clinical-skills/observation/ [WebCite Cache ID 6lG6Ajs0k]

53. Brownsell S, Bradley D, Cardinaux F, Hawley MS. Developing a systems and informatics based approach to lifestyle monitoring within eHealth: part II analysis and interpretation. 2011 Jul Presented at: Healthcare Informatics, Imaging and Systems Biology (HISB), First IEEE International Conference on th July; 2011; San Jose, CA p. 26-29. [doi: 10.1109/HISB.2011.25]
54. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote monitoring of patients with heart failure: an overview of systematic reviews. J Med Internet Res 2017;19(1):e18. [doi: 10.2196/jmir.6571]
55. Carel H. Illness: The Cry of the Flesh. Durham: Acumen Publishing; 2008.
56. Donovan JL, Blake DR, Fleming WG. The patient is not a blank sheet: lay beliefs and their relevance to patient education. Br J Rheumatol 1989 Feb;28(1):58-61. [Medline: 2917230]
57. Furze G, Donnison J, Lewin R. Clinician's Guide to Chronic Disease Management of Long Term Conditions: A Cognitive Behavioural Approach. Keswick: M&K Publishing; 2008.
58. Mock C, MacKenzie E, Jurkovich G, Burgess A, Cushing B, deLateur B, et al. Determinants of disability after lower extremity fracture. J Trauma 2000 Dec;49(6):1002-1011. [Medline: 11130480]
59. Charmaz K. The body, identity, and self: adapting to impairment. Sociol Q 1995;36(4):657-680.
60. Hanson J, Osipovic D, Percival J. Making sense of sensors: older people’s and professional caregivers’ attitudes towards telecare. In: Loader B, Hardy M, Keeble L, editors. Digital Welfare for the Third Age: health and social care informatics for older people. Abingdon, UK and New York, NY: Routledge; 2009:91-111.
61. Brownsell S, Blackburn S, Hawley M. User requirements for an ICT-based system to provide care, support and information access for older people in the community. J Assist Technol 2012;6(1):5-23. [doi: 10.1108/17549451211214328]
62. Wherton J, Sugarhood P, Procter R, Rouncefield M, Dewsbury G, Hinder S, et al. Designing assisted living technologies 'in the wild': preliminary experiences with cultural probe methodology. BMC Med Res Methodol 2012 Dec 20;12:188 [FREE Full text] [doi: 10.1186/1471-2288-12-188] [Medline: 23256612]
63. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. Heart 2007 Sep;93(9):1137-1146 [FREE Full text] [doi: 10.1136/hrt.2003.025270] [Medline: 17699180]
64. King N, Carroll C, Newton P, Dornan T. “You can't cure it so you have to endure it”: the experience of adaptation to diabetic renal disease. Qual Health Res 2002 Mar;12(3):329-346. [doi: 10.1177/104973202129119928] [Medline: 11918099]
65. Boeije HR. And then there was three: self-presentation styles and the presence of the partner as a third person in the Interview. Field Methods 2004;16(1):3-22. [doi: 10.1177/1525822X03259228]
66. Kendall M, Murray SA, Carduff E, Worth A, Harris F, Lloyd A, et al. Use of multiperspective qualitative interviews to understand patients' and carers' beliefs, experiences, and needs. BMJ 2009 Oct 14;339:b4122. [doi: 10.1136/bmj.b4122]
67. Reuband KH. On third persons in the interview situation and their impact on responses. Int J Public Opin Res 1992;4(3):269-274. [doi: 10.1093/ijpor/4.3.269]

Abbreviations

ADL: activities of daily living
COPD: chronic obstructive pulmonary disease
HF: heart failure
LM: lifestyle monitoring
LTC: long-term conditions

Edited by G Eysenbach; submitted 01.11.16; peer-reviewed by M Clarke, K Antypas; comments to author 29.01.17; revised version received 21.03.17; accepted 28.04.17; published 28.06.17.

Please cite as:
Hargreaves S, Hawley MS, Haywood A, Enderby PM
Informing the Design of “Lifestyle Monitoring” Technology for the Detection of Health Deterioration in Long-Term Conditions: A Qualitative Study of People Living With Heart Failure
J Med Internet Res 2017;19(6):e231
URL: http://www.jmir.org/2017/6/e231/
doi:10.2196/jmir.6931
PMID:28659253

Original Paper

Mobile Text Messaging to Improve Medication Adherence and Viral Load in a Vulnerable Canadian Population Living With Human Immunodeficiency Virus: A Repeated Measures Study

Elizabeth King[1], MD; Karen Kinvig[2], BN; Jonathan Steif[3], BSc; Annie Q Qiu[2], BSc; Evelyn J Maan[2], RN; Arianne YK Albert[4], PhD; Neora Pick[2,4,5], MD; Ariane Alimenti[2,6], MD; Mary H Kestler[2,5], MD; Deborah M Money[2,4,7], MD; Richard T Lester[5], MD; Melanie Caroline Margaret Murray[2,4,5], MD, PhD

[1]Department of Medicine, University of British Columbia, Vancouver, BC, Canada
[2]Oak Tree Clinic, British Columbia Women's Hospital, Vancouver, BC, Canada
[3]Department of Mathematics, University of British Columbia, Vancouver, BC, Canada
[4]Women's Health Research Institute, British Columbia Women's Hospital, Vancouver, BC, Canada
[5]Division of Infectious Diseases, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
[6]Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada
[7]Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Melanie Caroline Margaret Murray, MD, PhD
Division of Infectious Diseases
Department of Medicine
University of British Columbia
E600B
4500 Oak Street
Vancouver, BC, V6H 3N1
Canada
Phone: 1 604 875 2212
Fax: 1 604 875 3063
Email: melanie.murray@cw.bc.ca

Abstract

Background: Combination antiretroviral therapy (cART) as treatment for human immunodeficiency virus (HIV) infection is effective and available, but poor medication adherence limits benefits, particularly in vulnerable populations. In a Kenyan randomized controlled trial, a weekly text-messaging intervention (WelTel) improved cART adherence and HIV viral load (VL). Despite growing evidence for short message service (SMS) text-message interventions in HIV care, there is a paucity of data utilizing these interventions in marginalized or female cohorts.

Objective: This study was undertaken to assess whether the standardized WelTel SMS text-message intervention applied to a vulnerable, predominantly female, population improved cART adherence and VL.

Methods: We conducted a repeated measures study of the WelTel intervention in high-risk HIV-positive persons by measuring change in VL, CD4 count, and self-reported adherence 12 months before and 12 months after the WelTel intervention was introduced. Inclusion criteria included VL ≥200 copies/mL, indication for treatment, and meeting vulnerability criteria. Participants were given a mobile phone with unlimited texting (where required), and weekly check-in text messages were sent for one year from the WelTel computer platform. Clinical data were collected for control and intervention years. Participants were followed by a multidisciplinary team in a clinical setting. Outcomes were assessed using Wilcoxon signed ranks tests for change in CD4 and VL from control year to study end and mixed-effects logistic regressions for change in cART adherence and appointment attendance. A secondary analysis was conducted to assess the effect of response rate on the outcome by modeling final $\log_{10}$ VL by number of responses while controlling for mean $\log_{10}$ VL in the control year.

Results: Eighty-five participants enrolled in the study, but 5 withdrew (final N=80). Participants were predominantly female (90%, 72/80) with a variety of vulnerabilities. Mean VL decreased from 1098 copies/mL in the control year to 439 copies/mL at study end (P=.004). Adherence to cART significantly improved (OR 1.14, IQR 1.10-1.18; P<.001), whereas appointment attendance decreased slightly with the intervention (OR 0.81, IQR 0.67-0.99; P=.03). A response was received for 46.57% (1753/3764) of messages sent and 9.62% (362/3764) of text messages sent were replied to with a problem. An outcome analysis examining relationship between reply rate and VL did not meet statistical significance (P=.07), but may be worthy of investigating further in a larger study.

Conclusions: WelTel may be an effective tool for improving cART adherence and reducing VLs among high-risk, vulnerable HIV-positive persons.

Trial Registration: Clinicaltrials.gov NCT02603536; https://clinicaltrials.gov/ct2/show/NCT02603536 (Archived by WebCite at http://www.webcitation.org/6qK57zCwv)

(J Med Internet Res 2017;19(6):e190) doi:10.2196/jmir.6631

KEYWORDS

mHealth; HIV; medication adherence; vulnerable; female; viral load; mobile phone; engagement

Introduction

In Canada, approximately 75,500 people are living with human immunodeficiency virus (HIV), 16,800 of whom are women [1]. Combination antiretroviral therapy (cART) has led to enormous improvements in health and survival of HIV-positive individuals [2]. Moreover, by decreasing the amount of virus circulating in the plasma (viral load [VL]), cART offers the possibility of treatment as prevention by minimizing viral transmission [3,4]. Despite the well-established benefits of cART, nonadherence to medications, delayed initiation of therapy, and poor follow-up are ongoing limitations of effective HIV management leading to multidrug resistance, progression to acquired immune deficiency syndrome (AIDS), transmission to others, and mortality [2,5-7].

Engagement in longitudinal HIV care is an ongoing challenge, with studies from the United States showing long-term retention in care ranging from 37% to 65% [8-11]. Retention in care is particularly important on a public health scale because those individuals not retained in care have been estimated to account for twice as many HIV transmissions as nondiagnosed individuals [11]. In British Columbia, Canada, there is emerging evidence that cART adherence among women is lagging behind that of men (59% vs 68%, respectively) [12]. A variety of factors have been shown to decrease care retention and cART adherence in women, including intimate partner violence, medication side effects, role as care provider, and gender-specific stigma [13-19]. Other risk factors associated with poor adherence in both men and women include substance abuse, recent incarceration, psychiatric conditions, advanced HIV, and lack of social or housing supports [8,20-25]. Previous peer mentoring and behavioral interventions tested in these populations have failed to show objective improvements in HIV care [26-28]. For an illness that now has beneficial long-term treatment available, effective adherence strategies for difficult-to-engage populations will be of paramount importance from an individual and public health standpoint.

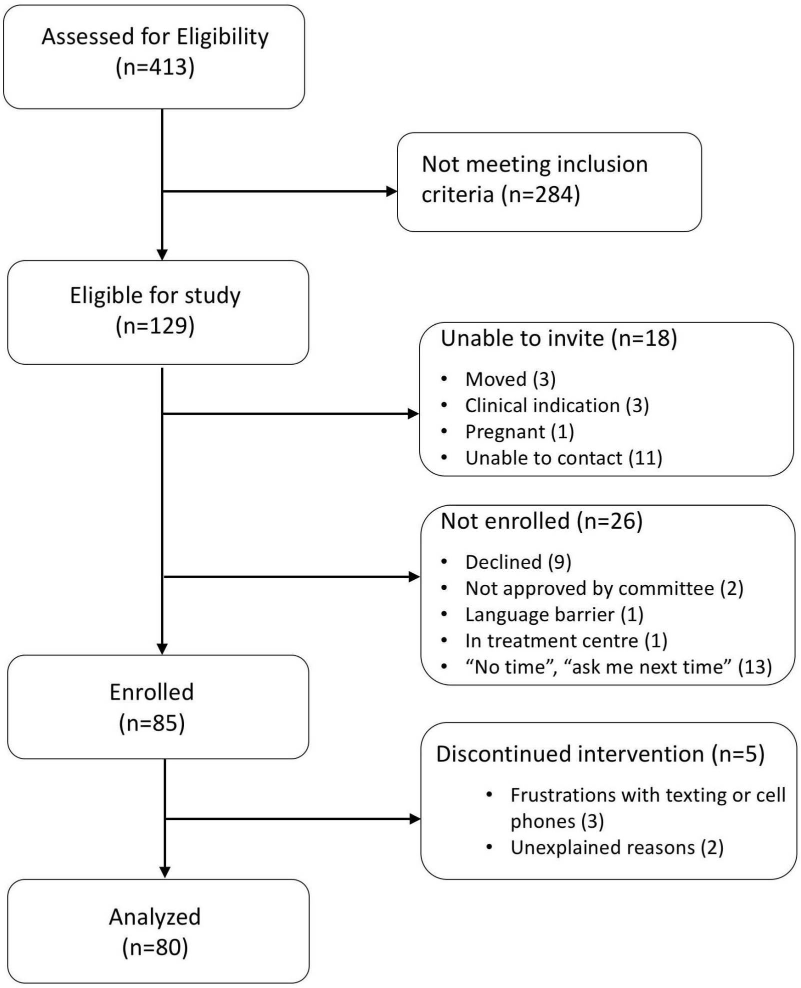
Mobile health (mHealth), the use of mobile phone technology to deliver health care, is an emerging area of disease management that can assist in patient adherence to prolonged chronic treatment regimens and monitoring of care [29-31]. Short message service (SMS) text messaging is particularly appealing as a mHealth intervention because it makes use of standard services offered by cellular providers, requires minimal additional infrastructure, and is user-friendly, ubiquitous, and flexible [29,30,32]. In a landmark randomized controlled trial, Lester et al conducted a mHealth intervention (WelTel Kenya1) whereby providing a bidirectional, weekly text-messaging intervention for one year significantly improved cART adherence and VL suppression compared with controls in an HIV-positive Kenyan population initiating cART [16]. Several other SMS text-message interventions have been developed and implemented in HIV care with promising results; however, WelTel remains the approach validated in the largest setting to date [33-35]. In addition, reviews and a meta-analysis have assessed available evidence to confirm that SMS text-message interventions can improve parameters of HIV care, especially cART adherence [17,36-38]. Despite this emerging evidence, there remains a lack of focus on vulnerable populations less likely to access care and a paucity of data focused on female populations [39-42]. Our goal in this study was to assess whether the standardized WelTel intervention applied to a vulnerable, predominantly female population improved cART adherence and VL.

Methods

Study Participants

Participants were enrolled in a repeated measures cohort study between April 2013 and May 2014 with the 12 months prior to initiation of the study used retrospectively as the control year. Participants were recruited from the Oak Tree Clinic in Vancouver, BC, Canada, which is a provincial referral center for women and families living with HIV throughout British Columbia, many of whom face multiple barriers to engagement in care. The clinic offers an interdisciplinary care team consisting of physicians, nurses, pharmacists, social workers, dieticians, and counselors addressing the holistic health needs of women and their families in a single care setting.

Figure 1. CONSORT flow diagram of trial.



As outlined in Figure 1, 413 individuals were assessed for eligibility, 129 were eligible, and 85 enrolled in the study. Five participants dropped out of the study early on and were not included in the final analysis. HIV-positive patients were eligible for study participation if they met the following inclusion criteria: attendance at the clinic for at least one year prior to study entry, an indication for cART (at time of study development, CD4 <500 cells/mm^3), detectable VL (≥200 copies/mL) in the year prior to study entry (control year), age 14 years or older, and high risk for disengagement in treatment according to a list of predetermined criteria. High-risk individuals were identified based on care team consensus that at least one of the following vulnerability criteria were present: intimate partner violence, unstable housing, advanced HIV infection/AIDS, mental health illness, cART nonadherence, difficult to contact, poor appointment attendance, active substance use, long distance from care, or recent incarceration. All candidates were reviewed by the multidisciplinary care team that decided on applicable vulnerabilities by manner of consensus. We excluded those who did not meet the preceding criteria or lived in an area with no cell phone service. Those who were unable to communicate by text messaging due to language barriers, illiteracy, or physical disabilities preventing use were also excluded. For those participants aged 14 to 18 years, parental consent for enrollment in the study was not obtained in an effort to protect confidentiality and promote safety in this marginalized cohort.

Prospective participants were introduced to the WelTel intervention during a routine clinical visit and provided written informed consent if interested in the study. Participants were given a basic cell phone with unlimited text-messaging capability if they did not have one. For those patients that did own a phone with a text-messaging plan, no payment incentive was provided. However, if an individual required a phone or phone plan at a later time, they were provided with one. When required, participants received instruction on how to use text messaging for communication. In addition to the intervention, participants continued to receive their regular care through the interdisciplinary team. In British Columbia, cART is fully funded for HIV-positive patients through the Provincial Antiretroviral Drug Treatment program and was prescribed according to published therapeutic guidelines by the BC Centre for Excellence in HIV/AIDS [43]. The study protocol was approved by the University of British Columbia Research Ethics Board (H12-03002) and registered as a clinical trial (NCT02603536).

Intervention

The SMS text-message intervention was modeled after the WelTelKenya1 study with participant-driven modifications according to a qualitative assessment conducted with patient participants and health care providers at the clinic prior to study start [16,18,44]. The WelTel model was selected because it has been validated in the largest setting to date and contains many features (ie, weekly, short, bidirectional messages) that have been subsequently shown to be effective SMS text-message strategies [16,34]. For one year, the text message “How are you?” was sent to all participants every Monday at noon from the automated WelTel platform through a number not traceable to the clinic. Participants were instructed to respond “OK” or “not OK,” and those responding “not OK” received a follow-up call by a study nurse within approximately 24 hours. If no response was received from the initial SMS text message, a second text message was sent on Wednesday stating, “Haven’t heard from you yet, how are you?” If there was still no response, this was followed by a call on Thursday by study nurses. Participants were instructed that the intervention was not an emergency service and were provided with a written reminder of the current practices for after-hours concerns and emergencies. This open-ended check-in approach was specifically designed to preserve confidentiality and provide a personal connection to clinic staff to triage all problem types. For confidentiality reasons, health care providers did not text information relating to HIV status unless asked explicitly to do so by the participant. Participants who lost their own phone or the study phone were offered a replacement on a one-time basis. Participants were allowed to reinitiate the study at any time simply by providing the team with the new phone number.

Measuring Outcomes

The primary outcome of the study was change in VL from control year (mean VL over one year prior to study entry) to final VL at study end. Secondary outcomes included change in cART adherence, CD4 count, and appointment attendance following the WelTel intervention. Parameters anticipated to respond after a lag time (VL and CD4) were compared using a mean of all measurements in the control year to final values at the study end, whereas those expected to have rapid response to the intervention (cART adherence and appointment attendance) were assessed by comparing means of control and intervention years in their entirety. To estimate the number of cART doses taken, self-reported adherence was collected at each medical visit with appointments arranged at 1- to 4-month intervals according to health status. At each visit, the patient was asked to estimate the number of ingested doses and missed doses since the past visit. Self-reported adherence was then compared to pharmacy refill records, always favoring the lowest estimate of doses taken. Finally, adherence was then calculated as the percentage of estimated doses taken from the total number of doses prescribed in the year. Appointment attendance data was gathered from the electronic booking system for the clinic.

Data were collected for one year prior to texting start date and for the duration of the study such that each participant served as their own control in the intervention. Participant self-reported ethnicity and vulnerability data were collected at time of enrollment. Chart review was used to collect the following: age, housing status, substance use, geographic location, CD4 count and percentage, VL, cART regimen (including dates of initiation or discontinuation), emergency department visits, hospital admissions, self-reported medication adherence, pharmacy refill data, and clinic appointment attendance. All text-messaging responses, follow-up calls, and team member referrals were tracked and nature of the problem noted.

Statistical Analysis

A sample size of 100 was initially targeted based on pilot study values of the primary outcome (VL) to allow for statistical assessment with a power of 80% and significance level of .05. To assess changes in VL and CD4, we computed the geometric mean ($\log_{10}$-transformed for VL) from raw data of all laboratory results from the control year for each participant. These means were then compared to the final VL or CD4 from the intervention year using nonparametric Wilcoxon signed ranks tests for paired data. Odds ratios for cART adherence and clinic attendance were analyzed based on binomial “yes/no” responses for each event with several events for each participant and then calculated by logistic regression. For each participant, the number of possible appointments and doses was tallied within each year, and the proportion of those that were attended/taken was compared between years using mixed-effect logistic regressions to control for paired measurements for each participant. Given the bidirectional nature of the text-messaging intervention, participants and care providers were not blinded throughout. In addition, outcomes were assessed according to the pre/post intervention study design rather than blinded.

To facilitate a comparison between those patients regularly using mHealth technology versus those not actively using it, we analyzed variation in the number of responses to the text messages. All participants should have received 52 messages making the number of responses a useful proxy of engagement. We used bivariate negative binomial regression to analyze the number of responses by age, gender, ethnicity, income, long-distance telephone, distance traveled to clinic ($\leq$50km, >50km), any substance use (not including smoking), and mean $\log_{10}$ VL during control year. We estimated the effect of response rate on the outcome by modeling final $\log_{10}$ VL by number of responses while controlling for mean $\log_{10}$ VL in the control year using a general linear model.

Results

Cohort Characteristics

Between April 2013 and May 2014, 85 participants were enrolled in the study. Although an initial target of 100 patients was selected based on statistical analysis, there were limitations in eligible clinic patients for enrollment. As such, study size was reanalyzed and determined to have sufficient power with a modified goal of 85 participants. Five participants withdrew from the study. Baseline demographics of the remaining participants are outlined in Table 1. The population was predominantly female (90%, 72/80) with a median age of 38 (IQR 29-47, range 15-61) years. Six (8%, 6/80) participants were younger than 18 years and received pediatric care at the clinic. The majority of the cohort had multiple vulnerabilities (76%, 62/80) of which cART nonadherence (54%, 44/80), mental health illness (48%, 39/80), active substance use (28%, 23/80), and unstable housing (28%,14/80) were the most common (Table 1). Almost half of participants did not own a cell phone at the time of enrollment (46%, 37/80). Geographic distance varied widely; however, the majority of participants lived within 50 km of the clinic. Final recorded VL measurements were taken a median of 314 days (IQR 269-348) after initiation of text messaging.

Intervention Efficacy

Viral load, the primary outcome, significantly decreased from control year to study end (*P*=.004; Table 2 and Figure 2). Overall, geometric mean VL decreased from mean 1098 (95% CI 647-1866) copies/mL in the control year to mean 439 (95% CI 217-888) copies/mL at intervention end ($\log_{10}$ VL 3.04-2.64; Cohen's *d*=0.33, 95% CI 0.02-0.65), with 38 of 80 participants having undetectable VL at study end. In contrast, CD4 did not significantly change postintervention (Table 2 and Figure 3). Of the secondary outcomes, cART adherence significantly improved from control year to intervention year with the odds of adherence increasing by 14% during the WelTel intervention (OR 1.14, IQR 1.10-1.18, *P*<.001; Table 2). Finally, there was a 19% reduction in the odds of attending clinic appointments in the intervention year (OR 0.81, IQR 0.67-0.99, *P*=.03). Other variables, including CD4 count, substance use, cART regimen, and hospital admissions, did not significantly change postintervention (Table 2).

Figure 2. Boxplot of mean $\log_{10}$ VL in the control year compared to the $\log_{10}$ VL at study end. The black bars indicate the medians; boxes indicate the interquartile range and the whiskers extend to 1.5× the interquartile range. There was a significant decrease in $\log_{10}$ VL after the intervention (P=.004).

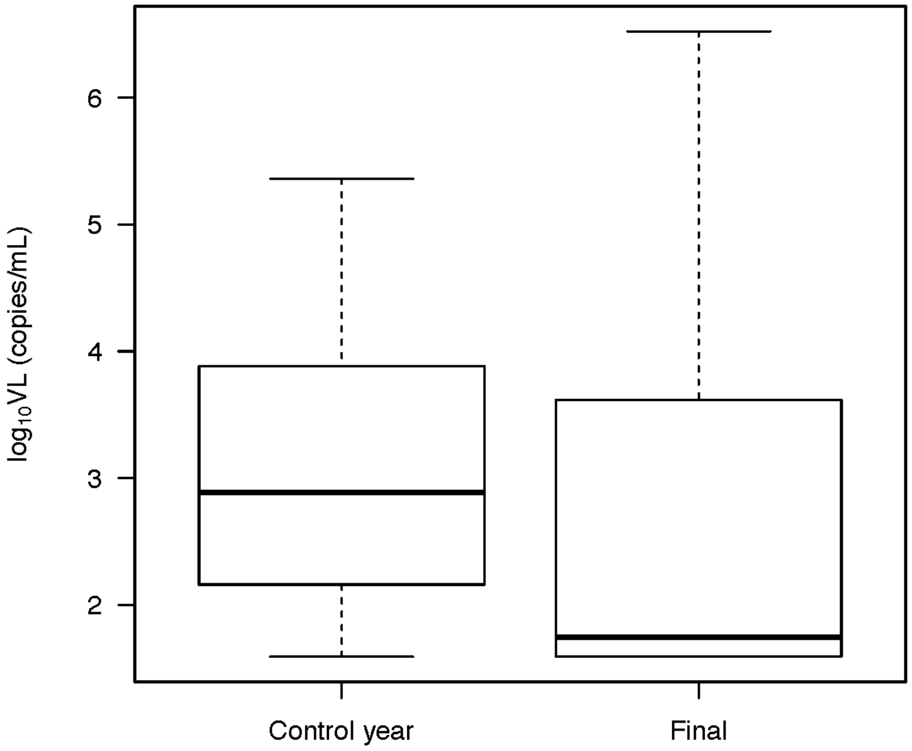


Table 1. Baseline demographics of a high-risk Canadian HIV-positive cohort (N=80).

Demographics	Participants
Gender, n (%)	
Female	72 (90)
Male	6 (8)
Transgender	2 (3)
Age (years), median (range)	38 (15-61)
Ethnicity, n (%)	
Caucasian	30 (38)
First Nations	27 (34)
African Canadian	18 (22)
South Asian	5 (6)
Income source, n (%)	
Disability	57 (71)
Welfare	6 (8)
Employed	4 (5)
Other	13 (16)
Housing status, n (%)	
Stable housing	57 (71)
Unstable housing	23 (29)
Unsheltered	1 (1)
Emergency sheltered	9 (11)
Provisionally sheltered	13 (16)
Vulnerability,[a] n (%)	
Multiple (≥2) vulnerabilities	62 (76)
Intimate partner violence	4 (5)
Unstable housing	23 (28)
Advanced HIV[b]	22 (27)
Mental health illness	39 (48)
cART nonadherence	44 (54)
Difficult to contact	12 (15)
Poor appointment attendance	22 (27)
Substance use	22 (28)
Long distance from care	6 (7)
Recent incarceration	4 (5)
Cell phone ownership, n (%)	
Yes	43 (54)
No	37 (46)
Geographic distance (km),[c] median (IQR, max)	29.6 (7.75-44.9, 1500)

[a] Vulnerabilities were expressed as percentage of population. Because multiple vulnerabilities were allowed for each individual, combined percentages will exceed 100%.

[b] Advanced HIV was defined by CD4 count ≤200 cells/mm^3.

[c] Geographic distance was calculated as the distance (in km) from the Oak Tree Clinic.

Table 2. Outcomes and health measures before and after WelTel intervention for a vulnerable cohort with HIV (N=80).

Outcomes and health measures	Control year	Intervention year	*P*
Primary outcome			
Geometric mean VL (copies/mL), mean (95% CI)	1098 (647-1866)	439 (217-888)	.004
Secondary outcomes			
CD4 (cells/mm^3),[a] median (IQR)	370 (166-550)	320 (190-600)	.24
Attendance (%), mean (95% CI)			
Clinic appointments	52 (48-55)	47 (43-50)	.03
All appointments	51 (48-55)	48 (45-51)	.12
cART adherence	60.3 (59.8-60.8)	62.2 (61.6-62.7)	<.001
Other health measures			
Substance use, n (%)			
Heroin	9 (11)	6 (8)	.45
Crack/cocaine/crystal meth	15 (19)	14 (19)	>.99
Polysubstance use[b]	13 (16)	13 (17)	>.99
Methadone	17 (21)	17 (21)	.68
cART regimen,[c] n (%)			
Protease inhibitor-based	49 (61)	39 (48)	.29
Nonnucleoside reverse transcriptase inhibitor-based	10 (12)	7 (9)	.47
Integrase-based	10 (12)	17 (21)	.18
Combination	7 (9)	11 (14)	.45
None	4 (5)	6 (7)	.53
Hospital stays[d]			
Total visits to ED for cohort, n, individual median (range)	62, 0 (0-11)	77, 0 (0-17)	.30
Total admissions to hospital, n, individual median (range)	20, 0 (0-2)	28, 0 (0-5)	.23
Days of hospital stay/admission individual median (IQR)	11.7 (5-11.5)	9.6 (1.8-15.3)	.63

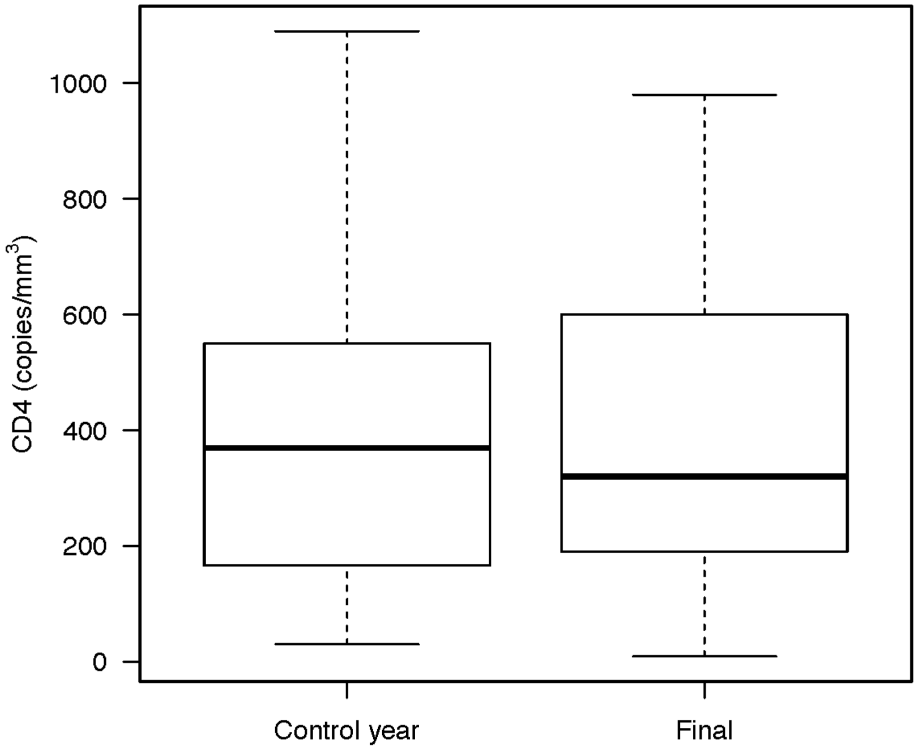
[a] Viral load and CD4 counts for intervention year were calculated based on study end data: geometric mean (95% CI) for VL and median (IQR) for CD4.

[b] Polysubstance use was defined as use of more than one of the following substances: alcohol, heroin, crack/cocaine, and methamphetamine.

[c] The reported cART regimen is the treatment combination used for the majority of study days in the control and intervention periods.

[d] Visits to the emergency department (ED) and hospital admissions were compared using Wilcoxon rank sum tests.

Figure 3. Boxplot of the mean CD4 in the control year compared to the mean CD4 at the study end. The black bars indicate the medians; boxes indicate the interquartile range and the whiskers extend to 1.5× the interquartile range. There was no significant change in CD4 after study intervention.



Intervention Responses

Of the 3674 text messages sent to the 80 remaining participants over the study period, 46.57% (1753/3764) of messages sent resulted in an initial “OK” reply, whereas 9.85% (362/3674) resulted in a reply of a “problem” and 42.43% (1559/3674) returned no reply. Although 362 “problem” replies were received in response to the Monday text message, an additional 203 “problem” replies were received later in the week, either as a second “problem” or as a new problem after an initial “OK” response. All “problem” replies were triaged by the study nurse. A total of 267 such problems were then referred to another health care team member as outlined in Figure 4. Physicians, nurse practitioners, and pharmacists were most commonly consulted to address a variety of issues that included medical advice, cART prescription refill orders, medication side effects, and medication pick-ups. Fifty participants were provided with a mobile phone at the beginning of the study, whereas 31 individuals lost a phone and required replacement. Response rates to weekly text messages initially started as high as 68% (54/80), but leveled off by the sixth week to an ongoing rate of approximately 45% (36/80) that remained constant to study end (Figure 5). On average, a fairly constant rate of 10% (8/80) of individuals reported “problems” each week.

Figure 4. Rate of health care provider involvement for all “problems” identified by SMS text message after assessment by study nurse.

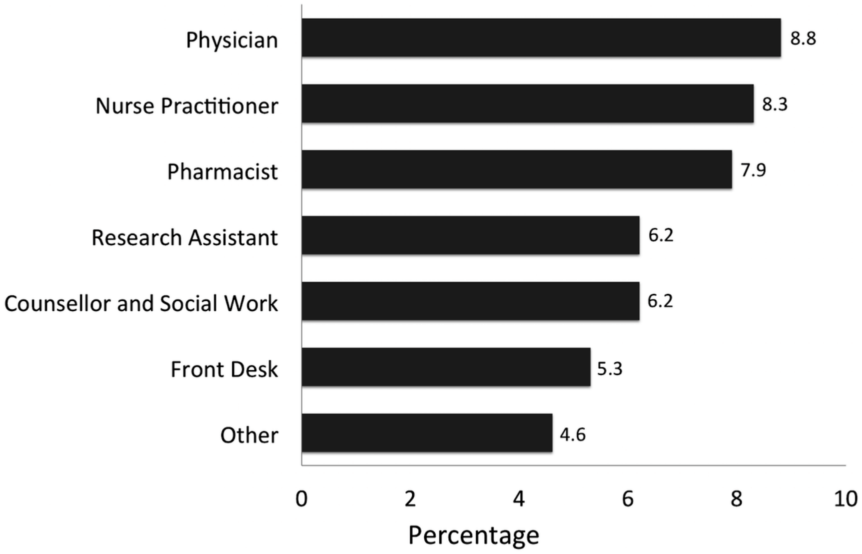
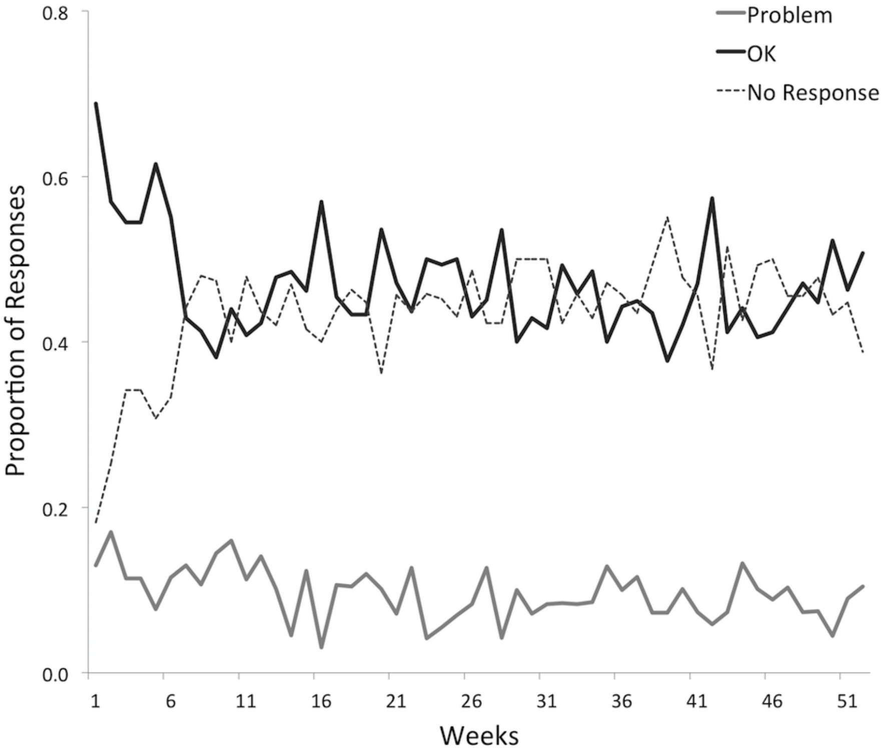


Figure 5. Proportion of weekly SMS text-message response types throughout the intervention.



Individual response rates varied greatly among participants (0%-98%) with a mean of 50% (26/52 responses). Based on these findings, an exploratory subgroup analysis was undertaken to compare demographic characteristics and VL by the number of replies to the text messages. There was no significant difference in the reply rate by any of the demographic or clinical characteristics explored (Table 3).

An outcome analysis of final $\log_{10}$ VL by number of replies revealed a trend for a relationship between decreasing VL by increasing reply rate controlling for mean $\log_{10}$ VL in the control year (*P*=.07; Figure 6), suggesting that the magnitude of the change in VL during the intervention year was larger for those who were more engaged. The model estimates that for every increase in one reply, the geometric mean VL is reduced by approximately 3% (95% CI 0%-7%).

Figure 6. Final log10 VL by number of SMS text-message responses in the intervention year. The solid line indicates predictions from the regression for median control year log10 VL (median 2.88). The dashed lines indicate the 95% CI around the regression estimate.

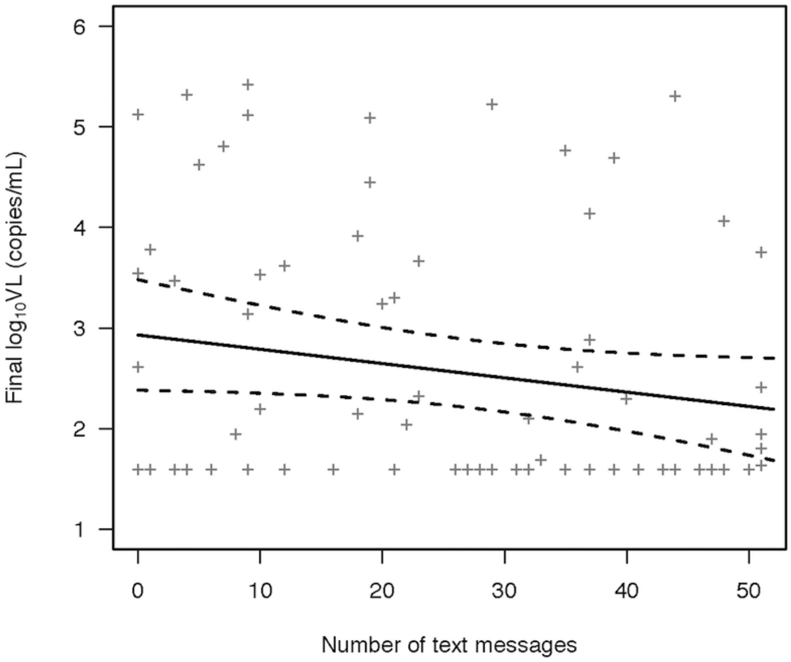


Table 3. Subgroup analysis by number of WelTel text message replies during the study year.

Outcome Variable	n	Incident rate ratio (95% CI)	*P*
Gender[a]			
Male	6	Reference	
Female	72	1.00 (0.45-1.94)	.99
Age (years)	80	1.00 (0.99-1.02)	.91
Ethnicity			
Caucasian	30	Reference	
First Nations	27	1.32 (0.84-2.07)	
African Canadian	18	1.28 (0.78-2.14)	
South Asian	5	1.50 (0.71-3.68)	.56
Income source			
Disability	57	Reference	
Other	23	1.18 (0.78-1.81)	.44
Long-distance call			
Yes	15	Reference	
No	64	0.82 (0.49-1.31)	.42
Geographic distance			
≤50 km	58	Reference	
>50 km	20	0.85 (0.56-1.32)	.47
Substance use			
None	28	Reference	
Alcohol or illicit drug use	47	0.75 (0.50-1.11)	.15
Mean control year $\log_{10}$ VL	80	0.94 (0.78-1.14)	.54

[a] The two transgender participants were excluded from this comparison.

Discussion

This study showed that a cohort of difficult-to-engage, vulnerable individuals living with HIV showed evidence of improved measures of HIV care (VL, cART adherence) when exposed to a weekly bidirectional text-messaging intervention. The study is unique in that participants were provided with mobile phones, allowing for increased engagement of marginalized populations that would otherwise be difficult to access. This intervention could potentially bring substantial health benefits to vulnerable populations in an era in which HIV care is largely limited by medication adherence.

The population demographic was a marginalized, predominantly female cohort with markers of poor HIV engagement. Participants came from a diversity of backgrounds and exhibited the wide range of vulnerabilities and barriers to care that could be encountered in any typical North American urban setting (Table 1; [24,40,45-48]). Vulnerabilities common in our population, including mental illness (48%), active substance use (28%), and unstable housing (17%), have been previously associated with poor cART adherence [24,45-50]. Cell phone ownership in this cohort was lower than average among Canadians (54% vs 82%), further confirming the inaccessible nature of this vulnerable study group [51,52]. By including a majority of women, we propose an adherence strategy applicable for the growing proportion of women living with HIV, a subpopulation that has been minimally studied and has characteristically poor cART adherence [13-15].

Our study demonstrated that effective introduction of the WelTel intervention decreased VL within our cohort. VL in our population was initially 1098 copies/mL, more than double the estimated mean of less than 500 copies/mL in British Columbia [53]. This dropped significantly following introduction of the WelTel intervention to a level that was on par with the provincial average by study end (439 copies/mL, *P*=.004; Table 2). Effective cART therapy is known to reduce viral replication [2]. Therefore, the observed VL reduction is likely a consequence of increased cART adherence (OR 1.14, IQR 1.10-1.18, *P*<.001) and minimizing treatment interruption during the WelTel intervention; we expect this relates to prompt recognition of problems and close connection with care. Due to the nature of studying a high-risk group, there is a risk of exaggerating treatment effects due to anticipated regression to the mean. However, additional analysis suggested that those who regularly used the service (ie, repliers) trended toward greater benefit to VL (Figure 4; *P*=.07). This argues that the improved parameters of HIV care may be due to exposure to the WelTel intervention rather than regression to the mean.

WelTel likely mitigated common barriers to cART adherence by (1) serving as a user-friendly, patient-directed connection tool with the care team, (2) connecting patients in an ongoing fashion with a support network to reduce the effects of isolation and stigma, and (3) aiding health care providers in identifying and resolving problems (both medical and related to the social determinants of health) as they arose, thereby preventing crises. Although the odds of adhering to cART was significantly improved during the intervention (OR 1.14, IQR 1.10-1.18, $P<.001$), the rate of adherence showed only modest improvement (60.3% vs 62.2%, Table 2). We expect that this difference appears small for a variety of reasons, including inaccuracies in self-reported adherence, failure to report treatment interruption, and a large variance in adherence rates. The strategy of early recognition and mitigation of problems has been shown to prevent treatment interruption in other settings as well, such as directly observed treatment of tuberculosis [54]. WelTel appears to improve patient-provider relationships by opening a channel of communication through which both medical concerns and those related to the social determinants of health may be addressed [18,44]. We expect that as these relationships improve and so, in turn, will medication adherence and eventually VL. Taken together, these results suggest that WelTel may be an effective patient-oriented tool to support and improve care in vulnerable HIV-positive populations.

The secondary outcome showing decreased clinic appointment attendance during our intervention year was an unexpected finding (Table 2). In other studies, participant fatigue has occurred with frequent or lengthy messaging interventions [34]. However, our observed consistency in texting response rates suggests that user fatigue does not explain our lower clinic attendance (Figure 3). Instead, we postulate that WelTel addressed common problems that would otherwise be resolved at clinic appointments to the point where participants had a decreased need for clinical appointments [18,29]. Despite having lower clinic attendance after the WelTel intervention, several other markers of health status, such as VL, emergency visits, and hospitalization rates, were the same or improved with the text-messaging intervention (Table 2). A previous hypertension study found that using two-way SMS text-message feedback and telephone consultation as a guide to plan ambulatory care management allowed for improved blood pressure management [55]. This raises the possibility that an effective two-way text-message intervention could be used to help guide HIV ambulatory follow-up frequency, in some cases safely decreasing appointment frequency. Finally, the text-messaging platform could be used to send reminders for upcoming appointments, a technique that has successfully improved clinic attendance in similar studies [34,39,56].

The text-messaging response rate was consistent throughout the study without any tapering trend (Figure 3). Previous studies involving patient participant feedback examined the optimal length and frequency of text messaging and found the best approach to maximize responses was a short weekly SMS text message, as opposed to daily texts or longer messaging interventions [33,34,39]. Our study supports this messaging strategy by showing consistent response rates after a full year of weekly text messages. The consistency in response rates that we observed also suggests that participants continue to make use of the intervention after one year, highlighting its feasibility in promoting long-term patient engagement. Previous studies have shown that mHealth interventions continue to have effect up to 18 months after introduction, but the optimal duration beyond this has not yet been studied nor have we seen whether there is a lasting effect if the intervention is withdrawn [16]. As WelTel moves toward broadening its application, future studies investigating the optimal duration of text-messaging interventions and the durability of its effect would help to guide implementation of this tool. Finally, through our analysis of participant engagement, we observed that some of those participants most poised to benefit from the intervention (ie, less adherent with cART) had poor uptake of the intervention in general. This may be because certain characteristics that result in poor adherence in HIV care also contribute to poor engagement with our intervention. Moving forward, it would be of interest to further delineate participant characteristics to anticipate who may maximally benefit from the intervention and for whom the resources will not be effectively allocated.

This study is a repeated measures study rather than a randomized controlled trial; therefore, we are only able to infer association not causation. WelTel had strong evidence of benefit in previous studies and, as such, our study was designed as a cohort study to offer maximal resources to all individuals in a high-risk population rather than random allocation [16-18]. Our study is limited by its sample size, predominantly female composition, and high-risk population studied, and it is therefore not generalizable to all populations. Medication adherence, one of our secondary outcomes, is a notoriously difficult parameter to study and patient self-reports are susceptible to recall bias and reporting inaccuracies. To mitigate this, we included pharmacy refill data in our assessment of number of doses taken, with the lower value being recorded as the correct one. Despite this, our method may still overestimate the number of pills actually ingested. Another limitation was that a large proportion of our cohort (47%) received a cell phone from us during the study, and we cannot control for all effects of cell phone ownership alone. Owning a mobile phone may impact several social determinants of health by impacting relationships, mobility, mental health, and self-perception [57,58]. Changes in these parameters may have gone on to positively influence outcomes associated with HIV care. Finally, multiple factors contribute to a patient's engagement in care, decision to initiate cART, and readiness to adhere to medication. Complex issues such as mental health and substance use cannot be solved by text messaging alone. However, it is our hope that this technology may act as a facilitator to establish a meaningful connection to care so that our most vulnerable patients may be better able to utilize and engage in the services available to them.

This study found that a simple and easily implemented mHealth intervention may significantly improve HIV care in vulnerable populations, a finding that holds great potential for real world application. Moving forward, further identification of subgroups poised to benefit most from WelTel will be important, particularly if the intervention includes the provision of a cell phone and phone plan. Finally, a cost analysis would shed light on the feasibility of introducing WelTel in a larger scale to targeted populations. We anticipate that the cost of the intervention would be mitigated by the cost-savings benefit of reduced HIV transmission and decreased use of health care resources associated with effective HIV care [2,59]. Taken as a whole, this study is pivotal in proposing an adherence strategy for at-risk HIV populations that we hope can be applied in a broad clinical context.

Acknowledgments

We would like to thank Gilead Sciences and Telus Mobility for their support of this study. The study was funded through a competitive grant competition through Gilead Sciences. Telus Mobility provided in-kind support in the form of technical support and cellular phone minutes for this study.

Conflicts of Interest

Dr Richard Lester is the founder of WelTel, the text-messaging platform used in this study. The technology platform was developed by a nonprofit organization, WelTel mHealth Society, and a private company, WelTel Incorporated. Dr Richard Lester has financial as well as professional interests in both organizations. All other authors declare no conflict of interest.

Multimedia Appendix 1

Consort E-Health Form.

[PDF File (Adobe PDF File), 850KB - jmir_v19i6e190_app1.pdf]

References

1. Public Health Agency of Canada. Summary: Estimates of HIV Incidence, Prevalence and Prorportion Undiagnosed in Canada, 2014. Ottawa, ON: Public Health Agency of Canada; 2015 Nov. URL: http://www.catie.ca/sites/default/files/2014-HIV-Estimates-in-Canada-EN.pdf [accessed 2017-02-26] [WebCite Cache ID 6oYWkC4PM]
2. Montaner JS, Hogg R, Wood E, Kerr T, Tyndall M, Levy AR, et al. The case for expanding access to highly active antiretroviral therapy to curb the growth of the HIV epidemic. Lancet 2006 Aug 5;368(9534):531-536. [doi: 10.1016/S0140-6736(06)69162-9] [Medline: 16890841]
3. 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]
4. Dube S, Boily M, Mugurungi O, Mahomva A, Chikhata F, Gregson S. Estimating vertically acquired HIV infections and the impact of the prevention of mother-to-child transmission program in Zimbabwe: insights from decision analysis models. J Acquir Immune Defic Syndr 2008 May 1;48(1):72-81. [doi: 10.1097/QAI.0b013e31816bcdbb] [Medline: 18317229]
5. Bangsberg DR, Perry S, Charlebois ED, Clark RA, Roberston M, Zolopa AR, et al. Non-adherence to highly active antiretroviral therapy predicts progression to AIDS. AIDS 2001 Jun 15;15(9):1181-1183. [Medline: 11416722]
6. Wilson DP, Law MG, Grulich AE, Cooper DA, Kaldor JM. Relation between HIV viral load and infectiousness: a model-based analysis. Lancet 2008 Jul 26;372(9635):314-320. [doi: 10.1016/S0140-6736(08)61115-0] [Medline: 18657710]
7. Lima VD, Johnston K, Hogg RS, Levy AR, Harrigan PR, Anema A, et al. Expanded access to highly active antiretroviral therapy: a potentially powerful strategy to curb the growth of the HIV epidemic. J Infect Dis 2008 Jul 1;198(1):59-67 [FREE Full text] [doi: 10.1086/588673] [Medline: 18498241]
8. Hu YW, Kinsler JJ, Sheng Z, Kang T, Bingham T, Frye DM. Using laboratory surveillance data to estimate engagement in care among persons living with HIV in Los Angeles County, 2009. AIDS Patient Care STDS 2012 Aug;26(8):471-478. [doi: 10.1089/apc.2011.0371] [Medline: 22731500]
9. Mugavero MJ, Amico KR, Horn T, Thompson MA. The state of engagement in HIV care in the United States: from cascade to continuum to control. Clin Infect Dis 2013 Oct;57(8):1164-1171 [FREE Full text] [doi: 10.1093/cid/cit420] [Medline: 23797289]
10. Yehia BR, Stephens-Shields AJ, Fleishman JA, Berry SA, Agwu AL, Metlay JP, et al. The HIV care continuum: changes over time in retention in care and viral suppression. PLoS One 2015;10(6):e0129376 [FREE Full text] [doi: 10.1371/journal.pone.0129376] [Medline: 26086089]
11. Skarbinski J, Rosenberg E, Paz-Bailey G, Hall HI, Rose CE, Viall AH, et al. Human immunodeficiency virus transmission at each step of the care continuum in the United States. JAMA Intern Med 2015 Apr;175(4):588-596. [doi: 10.1001/jamainternmed.2014.8180] [Medline: 25706928]
12. British Columbia Centre for Excellence in HIV/AIDS. HIV Monitoring Quarterly Report for British Columbia: First Quarter 2016. Vancouver, BC: British Columbia Centre for Excellence in HIV/AIDS; 2016. URL: http://www.cfenet.ubc.ca/sites/default/files/uploads/publications/centredocs/bc_monitoring_report_16q1_final_may20.pdf [WebCite Cache ID 6qUuDCWxZ]

13. Tapp C, Milloy M, Kerr T, Zhang R, Guillemi S, Hogg RS, et al. Female gender predicts lower access and adherence to antiretroviral therapy in a setting of free healthcare. BMC Infect Dis 2011;11:86 [FREE Full text] [doi: 10.1186/1471-2334-11-86] [Medline: 21466704]
14. Emery J, Pick N, Mills EJ, Cooper CL. Gender differences in clinical, immunological, and virological outcomes in highly active antiretroviral-treated HIV-HCV coinfected patients. Patient Prefer Adherence 2010;4:97-103 [FREE Full text] [Medline: 20517470]
15. Sordo del Castillo L, Ruiz-Pérez I, Olry de Labry LA. Biological, psychosocial, therapeutic and quality of life inequalities between HIV-positive men and women - a review from a gender perspective. AIDS Rev 2010;12(2):113-120. [Medline: 20571605]
16. 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]
17. 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]
18. Smillie K, Van BN, Abaki J, Pick N, Maan EJ, Friesen K, et al. A qualitative study investigating the use of a mobile phone short message service designed to improve HIV adherence and retention in care in Canada (WelTel BC1). J Assoc Nurses AIDS Care 2014;25(6):614-625 [FREE Full text] [doi: 10.1016/j.jana.2014.01.008] [Medline: 24768442]
19. Jewkes RK, Dunkle K, Nduna M, Shai N. Intimate partner violence, relationship power inequity, and incidence of HIV infection in young women in South Africa: a cohort study. Lancet 2010 Jul 03;376(9734):41-48 [FREE Full text] [doi: 10.1016/S0140-6736(10)60548-X] [Medline: 20557928]
20. Springer SA, Pesanti E, Hodges J, Macura T, Doros G, Altice FL. Effectiveness of antiretroviral therapy among HIV-infected prisoners: reincarceration and the lack of sustained benefit after release to the community. Clin Infect Dis 2004 Jun 15;38(12):1754-1760. [doi: 10.1086/421392] [Medline: 15227623]
21. Aidala AA, Wilson MG, Shubert V, Gogolishvili D, Globerman J, Rueda S, et al. Housing status, medical care, and health outcomes among people living with HIV/AIDS: a systematic review. Am J Public Health 2016 Jan;106(1):e1-e23. [doi: 10.2105/AJPH.2015.302905] [Medline: 26562123]
22. Villes V, Spire B, Lewden C, Perronne C, Besnier J, Garré M, ANRS CO-8 APROCO-COPILOTE Study Group. The effect of depressive symptoms at ART initiation on HIV clinical progression and mortality: implications in clinical practice. Antivir Ther 2007;12(7):1067-1074. [Medline: 18018765]
23. Sin NL, DiMatteo MR. Depression treatment enhances adherence to antiretroviral therapy: a meta-analysis. Ann Behav Med 2014 Jun;47(3):259-269 [FREE Full text] [doi: 10.1007/s12160-013-9559-6] [Medline: 24234601]
24. Giordano TP, Visnegarwala F, White AC, Troisi CL, Frankowski RF, Hartman CM, et al. Patients referred to an urban HIV clinic frequently fail to establish care: factors predicting failure. AIDS Care 2005 Aug;17(6):773-783. [doi: 10.1080/09540120412331336652] [Medline: 16036264]
25. Malta M, Magnanini MM, Strathdee SA, Bastos FI. Adherence to antiretroviral therapy among HIV-infected drug users: a meta-analysis. AIDS Behav 2010 Aug;14(4):731-747. [doi: 10.1007/s10461-008-9489-7] [Medline: 19020970]
26. Purcell DW, Latka MH, Metsch LR, Latkin CA, Gómez CA, Mizuno Y, et al. Results from a randomized controlled trial of a peer-mentoring intervention to reduce HIV transmission and increase access to care and adherence to HIV medications among HIV-seropositive injection drug users. J Acquir Immune Defic Syndr 2007 Nov 1;46 Suppl 2:S35-S47. [doi: 10.1097/QAI.0b013e31815767c4] [Medline: 18089983]
27. Arnsten JH, Li X, Mizuno Y, Knowlton AR, Gourevitch MN, Handley K, et al. Factors associated with antiretroviral therapy adherence and medication errors among HIV-infected injection drug users. J Acquir Immune Defic Syndr 2007 Nov 1;46 Suppl 2:S64-S71. [doi: 10.1097/QAI.0b013e31815767d6] [Medline: 18089986]
28. Jones TS, Vlahov D. What we can learn from the INSPIRE Study about improving prevention and clinical care for injection drug users living with HIV. J Acquir Immune Defic Syndr 2007 Nov 1;46 Suppl 2:S31-S34. [doi: 10.1097/QAI.0b013e318157892d] [Medline: 18089982]
29. Lester R, Karanja S. Mobile phones: exceptional tools for HIV/AIDS, health, and crisis management. Lancet Infect Dis 2008 Dec;8(12):738-739. [doi: 10.1016/S1473-3099(08)70265-2] [Medline: 19022188]
30. Wei J, Hollin I, Kachnowski S. A review of the use of mobile phone text messaging in clinical and healthy behaviour interventions. J Telemed Telecare 2011;17(1):41-48. [doi: 10.1258/jtt.2010.100322] [Medline: 21097565]
31. Kang H, Park H. A mobile app for hypertension management based on clinical practice guidelines: development and deployment. JMIR Mhealth Uhealth 2016 Feb 02;4(1):e12 [FREE Full text] [doi: 10.2196/mhealth.4966] [Medline: 26839283]
32. Buis L, Hirzel L, Dawood RM, Dawood KL, Nichols LP, Artinian NT, et al. Text messaging to improve hypertension medication adherence in african americans from primary care and emergency department settings: results from two randomized feasibility studies. JMIR Mhealth Uhealth 2017 Feb 01;5(2):e9 [FREE Full text] [doi: 10.2196/mhealth.6630] [Medline: 28148474]

33. Mbuagbaw L, van der Kop ML, Lester RT, Thirumurthy H, Pop-Eleches C, Ye C, et al. Mobile phone text messages for improving adherence to antiretroviral therapy (ART): an individual patient data meta-analysis of randomised trials. BMJ Open 2013;3(12):e003950 [FREE Full text] [doi: 10.1136/bmjopen-2013-003950] [Medline: 24345901]
34. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, 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]
35. Simoni JM, Huh D, Frick PA, Pearson CR, Andrasik MP, Dunbar PJ, et al. Peer support and pager messaging to promote antiretroviral modifying therapy in Seattle: a randomized controlled trial. J Acquir Immune Defic Syndr 2009 Dec 1;52(4):465-473 [FREE Full text] [Medline: 19911481]
36. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
37. van Velthoven MH, Brusamento S, Majeed A, Car J. Scope and effectiveness of mobile phone messaging for HIV/AIDS care: a systematic review. Psychol Health Med 2013;18(2):182-202. [doi: 10.1080/13548506.2012.701310] [Medline: 22788357]
38. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
39. Bigna JJ, Noubiap JJ, Kouanfack C, Plottel CS, Koulla-Shiro S. Effect of mobile phone reminders on follow-up medical care of children exposed to or infected with HIV in Cameroon (MORE CARE): a multicentre, single-blind, factorial, randomised controlled trial. Lancet Infect Dis 2014 Jul;14(7):600-608. [doi: 10.1016/S1473-3099(14)70741-8] [Medline: 24932893]
40. Ingersoll KS, Dillingham RA, Hettema JE, Conaway M, Freeman J, Reynolds G, et al. Pilot RCT of bidirectional text messaging for ART adherence among nonurban substance users with HIV. Health Psychol 2015 Dec;34 Suppl:1305-1315 [FREE Full text] [doi: 10.1037/hea0000295] [Medline: 26651472]
41. Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A randomized controlled trial of personalized text message reminders to promote medication adherence among HIV-positive adolescents and young adults. AIDS Behav 2016 May;20(5):1049-1059. [doi: 10.1007/s10461-015-1192-x] [Medline: 26362167]
42. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV treatment & prevention: a systematic review of the literature. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
43. Montaner JS, Guillemi S, Harris M. Therapeutic Guidelines for Antiretroviral (ARV) Treatment of Adult HIV Infection. Vancouver, BC: British Columbia Centre for Excellence in HIV/AIDS; 2015 Sep. URL: http://europepmc.org/abstract/MED/20638713 [accessed 2017-05-28] [WebCite Cache ID 6qnfgrZXs]
44. Murray MC, O'Shaughnessy S, Smillie K, Van Borek N, Graham R, Maan EJ, WelTel BC1 Study Team. Health care providers' perspectives on a weekly text-messaging intervention to engage HIV-positive persons in care (WelTel BC1). AIDS Behav 2015 Oct;19(10):1875-1887. [doi: 10.1007/s10461-015-1151-6] [Medline: 26297567]
45. Lucas GM, Chaisson RE, Moore RD. Highly active antiretroviral therapy in a large urban clinic: risk factors for virologic failure and adverse drug reactions. Ann Intern Med 1999 Jul 20;131(2):81-87. [Medline: 10419445]
46. Arnsten JH, Demas PA, Grant RW, Gourevitch MN, Farzadegan H, Howard AA, et al. Impact of active drug use on antiretroviral therapy adherence and viral suppression in HIV-infected drug users. J Gen Intern Med 2002 May;17(5):377-381 [FREE Full text] [Medline: 12047736]
47. Avants SK, Margolin A, Warburton LA, Hawkins KA, Shi J. Predictors of nonadherence to HIV-related medication regimens during methadone stabilization. Am J Addict 2001;10(1):69-78. [Medline: 11268829]
48. Chen NE, Meyer JP, Avery AK, Draine J, Flanigan TP, Lincoln T, et al. Adherence to HIV treatment and care among previously homeless jail detainees. AIDS Behav 2013 Oct;17(8):2654-2666 [FREE Full text] [doi: 10.1007/s10461-011-0080-2] [Medline: 22065234]
49. Bouhnik A, Chesney M, Carrieri P, Gallais H, Moreau J, Moatti J, MANIF 2000 Study Group. Nonadherence among HIV-infected injecting drug users: the impact of social instability. J Acquir Immune Defic Syndr 2002 Dec 15;31 Suppl 3:S149-S153. [Medline: 12562040]
50. Bouhnik A, Préau M, Vincent E, Carrieri MP, Gallais H, Lepeu G, MANIF 2000 Study Group. Depression and clinical progression in HIV-infected drug users treated with highly active antiretroviral therapy. Antivir Ther 2005;10(1):53-61. [Medline: 15751763]
51. Canadian Wireless Telecommunications Association. CWTA Subscriber Stats: Wireless Phone Subscribers in Canada in 2015. 2015. URL: https://cwta.ca/wp-content/uploads/2016/03/SubscribersStats_en_2015_Q4.pdf [accessed 2017-05-16] [WebCite Cache ID 6qUuDCWxZ]
52. PopulationPyramid.net. 2015. Population pyramids of the world from 1950 to 2100: Canada URL: http://www.populationpyramid.net/canada/2015/ [accessed 2017-05-16] [WebCite Cache ID 6qUulTRJM]
53. Montaner JS, Lima VD, Barrios R, Yip B, Wood E, Kerr T, et al. Association of highly active antiretroviral therapy coverage, population viral load, and yearly new HIV diagnoses in British Columbia, Canada: a population-based study. Lancet 2010 Aug 14;376(9740):532-539 [FREE Full text] [doi: 10.1016/S0140-6736(10)60936-1] [Medline: 20638713]

54. Kamolratanakul P, Sawert H, Lertmaharit S, Kasetjaroen Y, Akksilp S, Tulaporn C, et al. Randomized controlled trial of directly observed treatment (DOT) for patients with pulmonary tuberculosis in Thailand. Trans R Soc Trop Med Hyg 1999;93(5):552-557. [Medline: 10696421]
55. Kiselev AR, Gridnev VI, Shvartz VA, Posnenkova OM, Dovgalevsky PY. Active ambulatory care management supported by short message services and mobile phone technology in patients with arterial hypertension. J Am Soc Hypertens 2012;6(5):346-355. [doi: 10.1016/j.jash.2012.08.001] [Medline: 22995803]
56. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013;12:CD007458. [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
57. McInnes DK, Li AE, Hogan TP. Opportunities for engaging low-income, vulnerable populations in health care: a systematic review of homeless persons' access to and use of information technologies. Am J Public Health 2013 Dec;103 Suppl 2:e11-e24. [doi: 10.2105/AJPH.2013.301623] [Medline: 24148036]
58. Eyrich-Garg KM. Mobile phone technology: a new paradigm for the prevention, treatment, and research of the non-sheltered “street” homeless? J Urban Health 2010 May;87(3):365-380 [FREE Full text] [doi: 10.1007/s11524-010-9456-2] [Medline: 20397058]
59. Mocroft A, Phillips AN, Gatell J, Ledergerber B, Fisher M, Clumeck N, et al. Normalisation of CD4 counts in patients with HIV-1 infection and maximum virological suppression who are taking combination antiretroviral therapy: an observational cohort study. Lancet 2007 Aug 4;370(9585):407-413. [doi: 10.1016/S0140-6736(07)60948-9] [Medline: 17659333]

Abbreviations

AIDS: acquired immune deficiency syndrome
cART: combination antiretroviral therapy
HIV: human immunodeficiency virus
SMS: short message service
VL: viral load

Edited by G Eysenbach; submitted 08.09.16; peer-reviewed by P Cook, L Kuhns; comments to author 29.12.16; revised version received 02.03.17; accepted 11.04.17; published 01.06.17.

Please cite as:
King E, Kinvig K, Steif J, Qiu AQ, Maan EJ, Albert AYK, Pick N, Alimenti A, Kestler MH, Money DM, Lester RT, Murray MCM
Mobile Text Messaging to Improve Medication Adherence and Viral Load in a Vulnerable Canadian Population Living With Human Immunodeficiency Virus: A Repeated Measures Study
J Med Internet Res 2017;19(6):e190
URL: http://www.jmir.org/2017/6/e190/
doi:10.2196/jmir.6631
PMID:28572079

Original Paper

Tailored Communication Within Mobile Apps for Diabetes Self-Management: A Systematic Review

Heidi Holmen[1,2], RN, MSc Public Health; Astrid Klopstad Wahl[2], PhD; Milada Cvancarova Småstuen[1], PhD; Lis Ribu[1], PhD

[1]Department of Nursing and Health Promotion, Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Oslo, Norway
[2]Department of Health Sciences, Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway

Corresponding Author:
Heidi Holmen, RN, MSc Public Health
Department of Nursing and Health Promotion
Faculty of Health Sciences
Oslo and Akershus University College of Applied Sciences
PB. 4, St. Olavs plass
Oslo,
Norway
Phone: 47 67 23 62 41
Fax: 47 22 45 21 15
Email: heidi.holmen@hioa.no

Abstract

Background: The prevalence of diabetes is increasing and with the requirements for self-management and risk of late complications, it remains a challenge for the individual and society. Patients can benefit from support from health care personnel in their self-management, and the traditional communication between patients and health care personnel is changing. Smartphones and apps offer a unique platform for communication, but apps with integrated health care personnel communication based on patient data are yet to be investigated to provide evidence of possible effects.

Objective: Our goal was to systematically review studies that aimed to evaluate integrated communication within mobile apps for tailored feedback between patients with diabetes and health care personnel in terms of (1) study characteristics, (2) functions, (3) study outcomes, (4) effects, and (5) methodological quality.

Methods: A systematic literature search was conducted following our International Prospective Register of Systematic Reviews (PROSPERO) protocol, searching for apps with integrated communication for persons with diabetes tested in a controlled trial in the period 2008 to 2016. We searched the databases PubMed, Medical Literature Analysis and Retrieval System Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Central, Excerpta Medica database (EMBASE), ClinicalTrials.gov, and the World Health Organization (WHO) International Clinical Trials Registry Platform. The search was closed in September 2016. Reference lists of primary articles and review papers were assessed. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed, and we applied the Cochrane risk of bias tool to assess methodological quality.

Results: We identified 2822 citations and after duplicate removal, we assessed 1128 citations. A total of 6 papers were included in this systematic review, reporting on data from 431 persons participating in small trials of short duration. The integrated communication features were mostly individualized as written non–real-time feedback. The number of functions varied from 2 to 9, and blood glucose tracking was the most common. HbA_{1c} was the most common primary outcome, but the remaining reported outcomes were not standardized and comparable. Because of both the heterogeneity of the included trials and the poor methodological quality of the studies, a meta-analysis was not possible. A statistically significant improvement in the primary measure of outcome was found in 3 of the 6 included studies, of which 2 were HbA_{1c} and 1 was mean daytime ambulatory blood pressure. Participants in the included trials reported positive usability or feasibility postintervention in 5 out of 6 trials. The overall methodological quality of the trials was, however, scored as an uncertain risk of bias.

Conclusions: This systematic review highlights the need for more trials of higher methodological quality. Few studies offer an integrated function for communication and feedback from health care personnel, and the research field represents an area of heterogeneity with few studies of highly rigorous methodological quality. This, in combination with a low number of participants and a short follow-up, is making it difficult to provide reliable evidence of effects for stakeholders.

(J Med Internet Res 2017;19(6):e227) doi:10.2196/jmir.7045

KEYWORDS

diabetes mellitus (MeSH); communication (MeSH); mobile apps; self-management; systematic review; mHealth

Introduction

About 415 million people have diabetes globally, and management of diabetes and its complications remains a global health emergency that already accounts for 12% of global health expenditure [1,2]. Diabetes' impact is related to micro- and macrovascular complications [3,4] as well as deteriorated quality of life and increased rates of depression and anxiety [5,6]. The mobile health (mHealth) literature indicates that individuals using mobile apps for self-management achieve positive health outcomes [7]. Within the diabetes literature, both beneficial and adverse effects of mHealth solutions for self-management have been discussed; in summary, apps may be feasible and convenient for many but not all because of preferences, economy, and health literacy [8-11]. Possible functions in mobile apps include interaction functions such as messages and chatting with health care personnel (HCP); health-monitoring functions such as tracking blood glucose, weight, blood pressure, and medication; lifestyle-monitoring functions like physical activity and dietary habits; and educational functions supplying information. In addition, tracking of psychosocial status using patient-reported outcomes (PROs) is recognized as important in improving the understanding of living with a chronic disease and quality of care [12,13]; however, this function remains rare in apps for diabetes [14].

We argue that a key limitation of previous reviews is their lack of specific focus on communication, despite the emphasis Chomutare and colleagues [15] have placed on personalized education and feedback. The possibility for patients and HCP to review patient data within an app has been discussed previously [9,11,16,17]; however, it has not been thoroughly investigated. This is similar to a discussion we have had in our previous research after testing a mobile diabetes diary app with or without telephone contact with a diabetes specialist nurse, where the diabetes specialist nurse did not review any patient-related data within the app [18-20]. Despite encouragement, the participants did not discuss their data during their HCP consultations. Subsequently, some of our patients emphasized that if HCP had monitored, reviewed, or given feedback on their data through the app, the positive contributions of their data tracking and health counseling might have been greater (personal communication by Astrid Torbjørnsen, November 18, 2016). Further, our participants had a high disease burden and an undebatable need for change [21], so their needs were not met in our low-intensity intervention, and feedback based on the individual patient data might have changed this. In addition, Chomutare and colleagues [15] revealed a lack of personalized feedback in the apps they reviewed in 2011, and argued that this might be the missing link in diabetes self-management supported by apps. In general, earlier reviews of mobile apps for diabetes self-management include both reviews of apps available commercially evaluating mainly content and user ratings [10,14,15,22] and reviews of research and controlled trials to investigate possible effects of apps [7,23]. To date, there seems to be limited but encouraging evidence for the effectiveness of such apps compared with usual care, but the lack of rigorous research methodology is a weakness [9].

Within the research on technology-supported self-management, the effects of HCP communication via short message service (SMS), either alone or in combination with apps, have been investigated and have demonstrated promising results in the reviewed literature [24-26]. These effects might increase when the communication function is integrated within the app. Communication between patients and HCP based on individual health data to support the self-management of diabetes may produce improved health outcomes [27] such as increased self-management skills [18,20], increased self-monitoring of blood glucose and foot inspections, and decreased hemoglobin A_{1c}(HbA_{1c}) [28], as well as increased self-management and satisfaction with care with decreased diabetes distress and body mass index (BMI) [29]. Further, this tailored communication has been suggested to be a key preference among patients and providers [9,15]. Several professions might be involved: primary care physicians, diabetes specialist nurses, podiatrists, endocrinologists, clinical nutritionists, and others. Earlier research has suggested that alarmingly few patients attend self-management programs [30], and travel distances, rural areas, costs, and more might compromise the frequency of face-to-face HCP consultations, where technology might be an efficient alternative [8,23]. Receiving feedback on how to self-manage could represent a better solution for the patient than gathering data and reviewing them alone, and this would make the app more valuable than a paper-based diary [8]. Patients are increasingly becoming consumers of health, and if persons with diabetes prefer to communicate with their HCP through an app, it remains to be investigated whether apps with tailored communication can support diabetes self-management.

Self-management interventions have traditionally been based on theoretical frameworks, which are necessary to understand change [31]. Further, there has been proposed a linearity between applied theory and effect [32]. However, as mHealth becomes more frequent, a lack of theoretical foundation has been pointed out [8,9]. The goal of several apps is to help promote behavior change, which supports the argument for theory-based interventions. A recent review describes the need for integrating cognitive behavioral therapy into apps for diabetes, where the authors also propose a framework to reach this goal [33], which is an important step forward in understanding behavior change supported by mHealth and further increasing the quality of the apps.

Research on mobile apps with an integrated, tailored communication function is scarce, as the app interventions often include additional phone calls [34], SMS [35], face-to-face meetings [36], group meetings, or some combination of these [37,38] in addition to the mobile app itself. To the best of our knowledge, results based on apps with integrated and tailored communication alone have not been systematically summarized. This review aims to address this knowledge gap by systematically reviewing studies that aimed to evaluate integrated communication within mobile apps for tailored feedback between patients with diabetes and HCP in terms of (1) study characteristics, (2) functions, (3) study outcomes, (4) effects, and (5) methodological quality.

Methods

Protocol and Registration

The review protocol [39] was registered in the International Prospective Register of Systematic Reviews (PROSPERO) [CRD42016038640] and was presented and discussed by the first author in an oral conference presentation [40] in accordance with the PROSPERO [41] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) [42,43].

Information Sources and Search

A systematic literature search was conducted according to the PRISMA guidelines [44]. Medical literature published from January 2008 was searched in January 2016, with an updated search closed on September 23, 2016, using Medical Literature Analysis and Retrieval System Online (MEDLINE) , PubMed, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Excerpta Medica database (EMBASE), ClinicalTrials.gov, and the World Health Organization (WHO) International Clinical Trials Registry Platform. We reviewed reference lists of relevant reviews and studies, and we also conducted hand searches in relevant journals of the field in addition to studies based on tips from colleagues in the field.

In collaboration with a librarian at the Oslo and Akershus University College of Applied Sciences and a librarian at the University in Oslo, we organized a search strategy consisting of the terms “mobile applications,” “cell phones,” “mobile phones,” “smartphones,” “portable applications,” “mobile technology,” “portable technology,” or “app.” These were then combined with “diabetes mellitus” and/or “diabetes mellitus type 1” and/or “diabetes mellitus type 2” and/or “diabetic ketoacidosis.” The search strategy was tailored to each database for optimal results (Textbox 1). The specific search strategy for each database can be provided by the first author upon request. We did not set a language limitation; however, we did set a limitation on publication year to studies published from 2008, as we decided technologies prior to 2008 were unlikely to be mobile apps.

Eligibility Criteria

To be eligible, a study had to test a mobile app (software in a smartphone) with an integrated communication function for communication and/or feedback between patients and providers based on individual patient data. In this review, communication is conceptualized as medically trained personnel providing any kind of feedback based on patient data, being real time, chatting, individualized algorithms, or individualized trend analyses. The patient group had to have diabetes and be over the age of 16 years. The trials had to have a control group, either randomized, quasirandomized, or controlled clinical trial. We excluded trials that were for the primary prevention of diabetes, those regarding gestational diabetes, and those pertaining to a closed-loop or artificial pancreas system, as we regard those individuals to be unique in the way they perceive change and interact with HCP.

Textbox 1. Search strategy applied in MEDLINE.

Search strategy:

1. Mobile applications/ (697)
2. Cell phones/ (5888)
3. (Smartphone* or smart phone* or mobile phone* or cell phone* or cellphone*).tw.kf (7888)
4. (Mobile adj3 application*).tw.kf (1077)
5. (Portable adj3 application*).tw.kf (276)
6. (Mobile adj3 technolog*).tw.kf (1322)
7. (Portable adj3 technolog*).tw.kf (161)
8. (App or apps).tw.kf (15895)
9. Or/1-8 (26696)
10. Diabetes mellitus/ or exp diabetes mellitus, type 1/ or exp diabetes mellitus, type 2/ or diabetic ketoacidosis/ (246647)
11. Diabetes.tw.kf (386565)
12. 10 or 11 (448207)
13. 9 and 12 (643)
14. Limit 13 to yr=“2008-current” (565)

Study Selection

Two reviewers (HH and LR) independently reviewed all the titles and/or abstracts from the search. We applied our inclusion and exclusion criteria set a priori. For possibly eligible studies, a full text copy was retrieved and reviewed independently by HH and LR. Discrepancies were resolved by discussion or with the involvement of a third reviewer (AKW). Authors were contacted consecutively to clarify study design and determine whether the intervention was an app with integrated and tailored communication and no additional communication components. We sent one reminder to the nonresponders.

Data Extraction

Data were extracted for all eligible studies using a structured form that included descriptive information, type of design, outcomes, and follow-up with results and dropouts, as well as any data regarding a theoretical framework or a guideline-based approach in the app development or feedback process of the intervention. One reviewer (HH) performed the extraction, while a second reviewer (LR) performed quality assurance and checked that correct information was collected.

The baseline characteristics of the included trials are reported as means from the original papers and as weighted means to summarize overall sample characteristics of this systematic review. When a weighted mean is given, a median is not reported as there were small discrepancies between weighted means and medians.

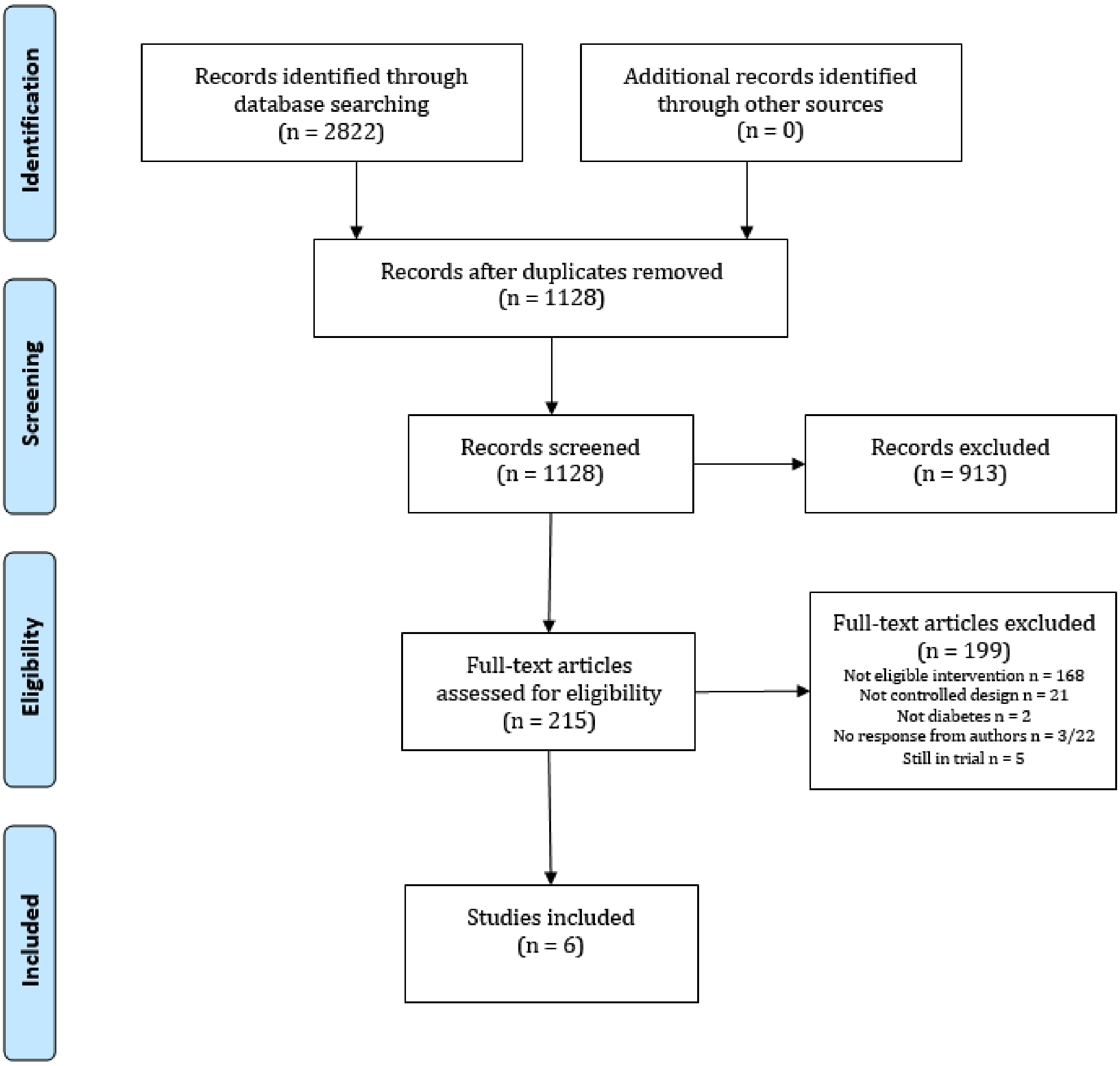
Outcomes

In this systematic review, we reviewed diverse health outcomes (physical and psychosocial) used as both primary and secondary outcomes.

Quality Assessment

The information reported in each article was used to assess the methodological quality of each study using the Cochrane methodology for risk of systematic bias (ROB) [45]. The ROB scoring was performed individually by 3 researchers (HH, AKW, and LR) and discussed to achieve consensus. To systematize the risk scores, Review Manager (Cochrane Community) was used.

Figure 1. Flowchart.



Results

Summary

A total of 2822 papers were identified during the search (Figure 1). After the removal of 1694 duplicates, the remaining 1128 citations were screened through title and/or abstract, and we removed 913 citations because they clearly did not meet our inclusion criteria. The full text of the remaining 215 citations was then obtained to clarify their study details, and we contacted 22 authors to clarify that their intervention consisted of an app with integrated and tailored communication and no additional contacts. Of the 22 authors we contacted, 18 responded immediately, 1 responded after a reminder, and 3 requests remained unanswered after 1 reminder. The corresponding citations were excluded from the review. After the termination of the search, 6 citations [46-51] were included in this review. The main reasons for exclusion were research related to the prevention of diabetes, mobile apps without communication, and other media used for communication including email, phone calls, and SMS (texting). Several studies were identified that had an intervention consisting of a mobile app with communication, while some of these had additional contacts by telephone, Web, or face to face and were therefore excluded [34-38].

Study Characteristics

The included studies were heterogeneous in study procedures and design (Tables 1 and 2); 4 used a randomized controlled trial (RCT) design, of which 1 was a pilot study. One study used a matched, controlled design, and 1 study randomly selected participants before assigning them into 2 groups. A total of 3 of the studies were conducted in Asia, (ie, Japan [50], China [51], and Korea [47]), 1 in the Democratic Republic of Congo [49], 1 in Canada [48], and 1 was a multicenter study conducted in 3 European countries [46]: Italy, Spain, and the Czech Republic. The papers were published between the years 2012 and 2016. Usual care was not described in detail in any of the included studies.

Table 1. Study characteristics.

Study	Year	Country	Randomization	Allocation	Dropouts n (%)
Fioravanti [46]	2015	Czech Republic, Italy, Spain	Not described	Unclear	3 (5.6)
Kim [47]	2014	Korea	Matched control design, not randomized	Unconcealed	3 (7.9)[a]
Logan [48]	2012	Canada	Block randomization using blocks of 4 and 6	Unclear	6 (5.5)
Takenga [49]	2014	Congo, Germany	Not described	Unclear	9 (22.5)
Waki [50]	2014	Japan	Not described	Unclear	5 (9.3)
Zhou [51]	2016	China	Random number table	Unclear	NA[b]

[a] Dropout only from the intervention group.

[b] NA: not available.

Table 2. Trial design.

Study	Patients included	Intervention group	Control group	Duration	Measurement times
Fioravanti [46]	51[a]	METABO app, chatting and message function with HCP[b]	Usual care	1 month	Baseline and 1 month
Kim [47]	70[c]	Mobile smartcare app, weekly feedback from HCP, warnings when hypos registered, and reminders	Matched controls from electronic medical records	3 months	Baseline and 3 months
Logan [48]	110	Real-time self-management messages based on care paths of averages of transmitted blood pressure readings	Tele monitoring without messages	12 months	Baseline and 12 months
Takenga [49]	40	MobilDiab app, feedback and messages from HCP	Usual care	2 months	Baseline and 2 months
Waki [50]	54	Dialbetics app, feedback based on patient data and guidelines	Usual care	3 months	Baseline and 3 months
Zhou [51]	100	Welltang app, weekly feedback, and upon patient-request	Usual care, monthly	3 months	Baseline and 3 months

[a] Included N=54, numbers given for N=51 completers.

[b] HCP: health care personnel.

[c] Included N=73, numbers given for N=70 completers.

Table 3. Participant characteristics at baseline of included trials, N=6.

	Age (mean years)	Gender (male/female)	Type 1 diabetes	Type 2 diabetes	Duration of diabetes (mean years)	HbA_{1c}[a] (mean %)	BMI[b] (mean kg^2)
Fioravanti [46] N=51[c]	48.0[d]	36/15[d]	29[d]	22[d]	20.0[d]	7.9[d]	25.7[d]
Kim [47] N=70[e]	52.8	40/30	0	70	11.8	7.7	25.0
Logan [48] N=110	62.9	61/49	NA[f]	NA	NA	7.4	30.9
Takenga [49] N=40	53.3	29/11	NA	NA	NA	8.6	NA
Waki [50] N=54	57.2	41/13	0	54	9.0	7.0	26.7
Zhou [51] N=100	54.2	57/43	18	82	6.6	9.8	23.0

[a] HbA_{1c}: hemoglobin A_{1c}.

[b] BMI: body mass index.

[c] Included N=54, numbers given for N=51 completers.

[d] Provided upon request.

[e] Included N=73, numbers given for N=70 completers.

[f] NA: not available.

Participants

Overall, the 6 trials reported data from 431 participants as shown in Table 3, with a sample size varying from N=40 to N=110 and a median of 64 participants. One trial did not give any demographic data in the original article [46]; however, the author provided this information upon request. A total of 2 trials did not report any data regarding their total of n=6 dropouts [46,47]. Overall, 47 participants were specified to have type 1 diabetes and 228 type 2 diabetes, while 2 studies [48,49] did not specify type of diabetes for their combined total of 150 participants. Weighted mean age was 55.8 years, including 264 males and 160 females. Duration of diabetes was provided upon request from one trial [46], and reported in 3 papers, giving a weighted mean of 11 years [47,50,51]. HbA_{1c} was collected in all trials [46-51] with a weighted mean of 8.1%, and BMI was reported in 4 papers [46,47,50,51] and provided by 1 author in an email, giving a weighted mean BMI of 26.5 kg^2.

Functions of the Mobile Apps

The mobile apps used in the included studies varied in their form and functions (Table 4), and a theoretical foundation was largely lacking.

The feedback used was either automatic or manual feedback, both tailored, and 4 apps also offered direct messages from the patient in free text. A total of 3 studies had automated individualized feedback consisting of text tailored to the participant baseline data and their current readings [46,48,50]. One of these had the participant data evaluated according to diabetes treatment guidelines [50], a second study had an additional message function [46], while a third had no additional feedback or messages [48]. The other 3 studies had individualized feedback given directly by the physician [49], medical staff [47], or the study team [51], and 2 of these had an additional message function for questions in free text [49,51].

The MobilDiab study had therapy plans, instructions, and recommendations sent by the physician in the app [49]. The Welltang app offered answers to questions within the day in addition to weekly or fortnightly feedback [51]. The METABO app [46] had both the app and the content of messages tailored to the type of diabetes: those with type 2 diabetes had a less complex app and received more persuasive messages; patients could also turn off alerts they did not want to receive and tailor the timing of the messages. DialBetics was an extensive app, consisting of automatic transfer of data and feedback based on blood glucose readings, diet, blood pressure, physical activity, and weight, where the users received immediate feedback based on every registration in the app, evaluated following the Japanese Diabetes Society guidelines [50].

A total of 2 apps had critical alerts sent to the patients if their entered readings were outside preset thresholds [46,51]. In MobilDiab, the physicians received an alert if emergency values were recorded and they instructed the patient [49], while in DialBetics, any readings outside preset thresholds triggered an alert sent to the study team [50]. A total of 3 apps alerted patients when they missed readings [48,50,51] and a fourth had automatic alerts regarding hyperglycemia; the medical team called the patient if they recorded a hypoglycemic value or if they missed several readings [47].

Table 4. Functions of the mobile apps.

Study	Communication	Blood glucose	Diet	Blood pressure	Medication	Physical activity	Weight
Fioravanti [46]	Chat with HCP[a], messages and individualized automated feedback according to the TTM[b]	Manual input	Manual input	Manual input	Manual input	Manual input	Manual input
Kim [47]	Messages and individualized feedback	Manual input	—	Manual input	—	—	—
Logan [48]	Individualized automated feedback	—	—	Bluetooth	—	—	—
Takenga [49]	Messages and individualized feedback	Automatic transfer and manual input	Manual input	Manual input	Manual input	Manual input	Manual input
Waki [50]	Individualized automated feedback according to Japan Diabetes Society guidelines	Automatic transfer	Voice, text, or photo of meal	Automatic transfer	—	Automatic transfer of pedometer or voice or text	Automatic transfer
Zhou [51]	Messages and individualized feedback	Manual input	Manual input	—	Manual input	—	—

[a] HCP: health care personnel.

[b] TTM: transtheoretical model stages of change.

The most frequent function besides communication was registration of blood glucose; this was found in 5 apps. One of the 5 offered automatic transfer [50] of blood glucose readings from the meter to the app, while 3 had manual input of the measured blood glucose [46,47,51]. One app had both: automatic transfer from a specific glucose meter and manual input if the patients used a different meter [49]. Blood pressure measurement was offered in 5 of 6 apps, while diet and graphical trends of measures were offered in 4 out of 6 apps. Tracking and imputation of medication, levels of physical activity, and weight were functions in 3 out of 6 apps, in addition to their diabetes information functions. A total of 2 apps offered individual goal setting: 1 offered a connection to continuous glucose monitors and 1 had laboratory data in the app. None of the apps in the included studies had psychosocial measures as a function. The number of functions in addition to communication ranged from 2 to 9 with a median of 6 functions. In 3 of the included studies, the intervention also consisted of a digital solution like an app or a Web page for the involved HCP [46,49,51].

Outcomes and Effects

Primary outcomes were specified in 5 trials, whereas 1 study [49] did not specify the order of the outcomes (Table 5), and various outcomes were used to evaluate the interventions in the individual trials.

HbA_{1c} was reported in 4 of the 6 included trials, and stated as the primary outcome in 3 papers. A total of 2 papers [50,51] reported a significant decrease in the intervention groups compared with the control groups, namely –0.4% and –1.95%, while 2 trials reported no change between groups [47,49] and the remaining 2 papers did not report change in HbA_{1c} as an outcome [46,48].

Change in blood pressure as an outcome was reported inconsistently using both systolic blood pressure (SBP), diastolic blood pressure (DBP), and mean blood pressure among the papers reporting blood pressure [47,48,50,51]. One paper reported changes in mean daytime ambulatory SBP as the primary outcome and found a significant decrease in the intervention group compared with the control group [48]. A total of 3 papers reported no significant change in either SBP or DBP between the intervention and control groups [47,50,51], while 2 papers reported no measures of blood pressure [46,49].

Regarding diabetes knowledge, there were no significant differences between the intervention group and the control group in 2 trials using this as an outcome [46,51], although neither used validated measures in their data collection.

Various assessments of usability and satisfaction were reported [46,47,49-51] but common for all was the use of nonvalidated and comparable questionnaires for this evaluation.

One paper reported a significant increase in depressive symptoms using the Hospital Anxiety and Depression Scale (HADS) in the intervention group [48].

Table 5. Outcomes and effects of included studies.

Study	Outcome measures[a]	Effects			
		HbA_{1c}[b]	Blood pressure	App-related evaluations	Other evaluations
Fioravanti [46]	Feasibility (primary), acceptance, adherence, usage, knowledge, glycemic control, quality of life	NA[c]	NA	Feasible	Increased medication adherence and diabetes knowledge in intervention group
Kim [47]	HbA_{1c} (primary), anthropometrics, satisfaction, comfort, convenience, functionality	No change	SBP[d]increased in intervention group; not significant between groups	Increased satisfaction	NA
Logan [48]	Mean daytime ambulatory SBP (primary), antihypertensive medication, HADS[e], comfort with home BP[f]measurement	NA	Significant decrease in mean daytime ambulatory SBP	NA	Worsened HADS in intervention group
Takenga [49]	HbA_{1c}, mean blood glucose, usability, acceptance, efficiency, therapy satisfaction	Decreased in intervention group	NA	Positive usability	NA
Waki [50]	HbA_{1c} (primary), fasting blood glucose, BP, BMI[g], LDL[h], HDL[i], triglycerides, medication, self-management, usability	Significant decrease in intervention group	NA	Positive usability	NA
Zhou [51]	HbA_{1c} (primary), blood glucose, LDL, weight, BP, hypoglycemia, satisfaction with diabetes care, usability of app, diabetes knowledge, self-care	Significant decrease in intervention group	No change	Positive usability (dichotomous)	Significant increase in diabetes knowledge and self-care in the intervention group

[a] Questionnaire not standardized unless stated otherwise.

[b] HbA_{1c}: hemoglobin A_{1c}.

[c] NA: not available.

[d] SBP: systolic blood pressure.

[e] HADS: Hospital Anxiety and Depression Scale.

[f] BP: blood pressure.

[g] BMI: body mass index.

[h] LDL: low density lipoprotein.

[i] HDL: high density lipoprotein.

Methodological Quality Assessment

Overall, the methodological quality as assessed by the ROB [45] was dominated by uncertainty risk because of lack of information in the included articles, as visualized in Figures 2 and 3. A lack of information in the publications was scored as “uncertain,” while we rated articles with sufficient information according to the Cochrane ROB guidelines [45]. As Figure 2 shows, the overall risk of bias is greatest regarding performance bias and selective reporting because of lack of blinding of the intervention and lack of reporting outcomes a priori in databases such as ClinicalTrials.gov or publishing the research protocols. “Other bias” is the domain with lowest risk, but several points can be highlighted, including economic interests, patent interests, and other factors influencing free research. We have, however, no indication that such issues are present in our included studies and have rated them low.

When we applied the ROB tool to our 6 papers, “unclear” was given 18 times, “high” was given 11 times, and “low” was given 11 times, supporting an overall unclear ROB among the included studies. Randomization procedures were reported insufficiently in 3 papers [46,49,50], resulting in an unclear ROB, while Kim and colleagues [47] had a matched control design and hence had a high risk because the participants were fully aware of their group. Blinding of participants and personnel was not performed in 4 papers [47,48,50,51], leading to a high risk according to guidelines, and not mentioned in 2 papers [46,49], giving an unclear risk. Blinding of the outcome was performed in 1 study giving a low risk of bias [48]. The 2 trials reporting not having performed blinding of outcomes [50,51] were scored as high risk, and the 3 papers [46,47,49] not mentioning this were scored as unclear risks. The completeness of outcome data was unclear in 4 of the 6 papers as there were inconsistencies in reporting rates and reasons for attrition; 2 papers, however, reported sufficient information and were given a low risk of attrition bias. Selective reporting was assessed as high in all 5 trials not registered in a clinical trials database and uncertain when this information was lacking. One study was registered in a WHO-approved register for clinical trials and hence scored low on reporting bias [51].

Figure 2. Risk of bias: review authors' judgements about each risk of bias item presented as percentages across all included studies.

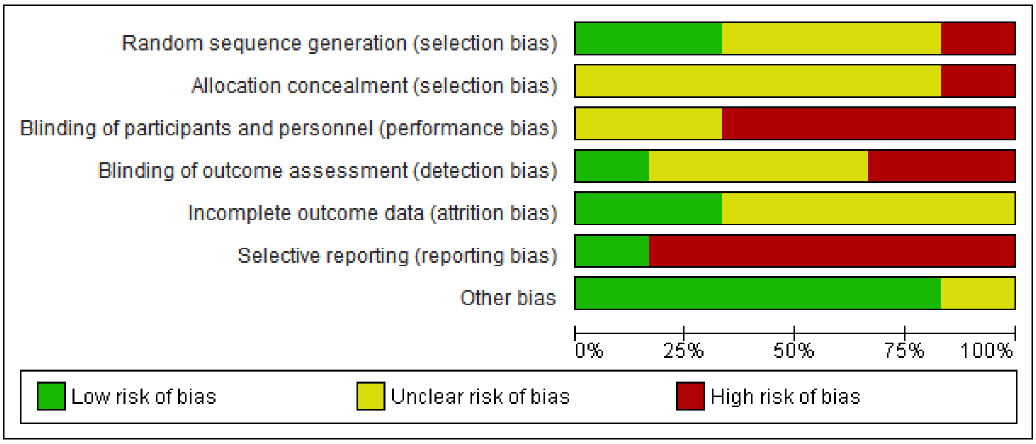
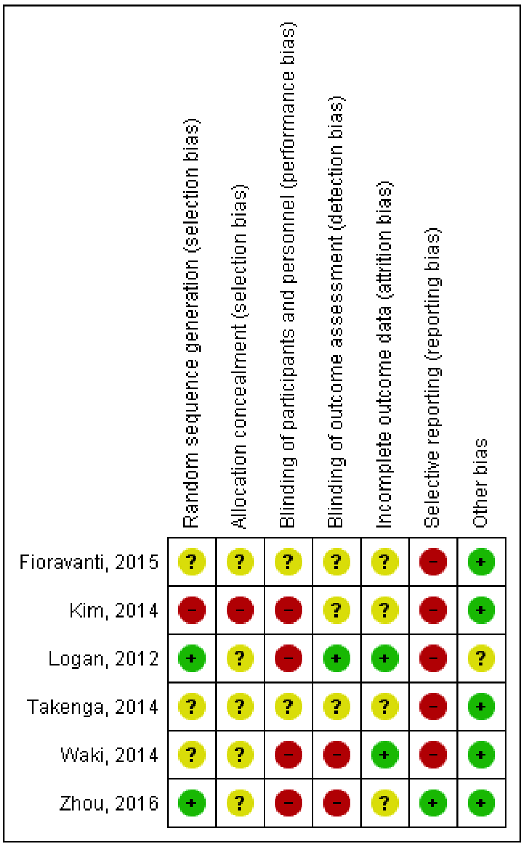


Figure 3. Risk of bias summary: review authors' judgments about each risk of bias item for each included study.



	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Fioravanti, 2015	?	?	?	?	?	−	+
Kim, 2014	−	−	−	?	?	−	+
Logan, 2012	+	?	−	+	+	−	?
Takenga, 2014	?	?	?	?	?	−	+
Waki, 2014	?	?	−	−	+	−	+
Zhou, 2016	+	?	−	−	?	+	+

Discussion

Principal Findings

This systematic review describes the study characteristics, functions, outcomes, effects, and methodological quality of intervention trials studying apps for diabetes self-management with a tailored and integrated HCP-patient communication function. To summarize, the studies included in this review represented a heterogeneous research area. The mobile app communication functions integrated in the studied apps were largely automated feedback from HCP, and the number of additional functions varied from 2 to 9, of which blood glucose registration was the most frequent. Statistically significant effects were found in 3 of the 6 trials: 2 reported a decrease in HbA_{1c} [50,51] and 1 reported a decrease in SBP [48]. The unclear methodological quality of these few studies has implications for the evidence from this systematic review. We argue, however, that our paper has an important message regarding the state of the research field, and it highlights the need for more controlled trials of higher methodological quality.

We found only 6 controlled trials with apps offering integrated communication functions, and SMS, phone calls, and face-to-face consultations are still common in the mHealth research field. From this rigorous yet wide systematic search, we had anticipated a larger number of controlled trials investigating mobile apps with individualized and integrated feedback from HCP, as a 2011 review called for such research [15] and the availability of hundreds of apps is frequently highlighted [9,23]. Previous research has also discussed the possibility of communication through a mobile app for health purposes (eg, for collecting and analyzing health data) related to the idea of one platform to serve all patient needs [7,11]. We identified several apps through our search, but their interventions were not in the scope of this review, with additional contacts through either email, SMS, or phone calls [34-38]. Further, we might have had a higher number of eligible trials if the search had included other chronic conditions. However, as diabetes self-management might be uniquely complex, including clinical variables, we suggest that the results derived from such research would have been of less value to the diabetes field.

Others have also discussed whether apps are less scientifically tested than other medical solutions [9], possibly explaining the low number of relevant scientific papers for this review. It might be difficult to commercialize an app involving HCP because of the practicalities and high costs. Further, it is increasingly recognized that apps should be regulated in terms of their effect, security, and privacy [9,23,52]. Currently, the US Food and Drug Administration offers unbinding guidelines for their regulation, approval, and clearance of apps, without the control authority [53]. CE marking is frequently applied in the European countries, however, this is based on self-certification and accounts for the health, safety, and environment protection related to the app. A third regulation is the Health Insurance Portability and Accountability Act (HIPAA), which pertains to the US national standards for electronic health, especially regarding devices that collect, store, or share identifiable data with HCP. None of the current reviewed papers discussed these regulations. Navigating these regulations, in addition to cost and practicalities, might make research in this area less attractive because the real world demands cost effectiveness [23,47], an outcome not covered by this review.

The studies included in this review include research from Northern America, Europe, Asia, and Africa, and all have small samples and short follow-ups. The longest follow-up was 12 months [48], whereas the rest had 1 to 3 months, possibly too short to prove an effect as one author suggested [47]. As is common in the area, the majority of the studies did not distinctly separate type 1 and type 2 diabetes, except for one [46], describing a less complex intervention for those with type 2 diabetes. We argue that although persons with diabetes experience many of the same symptoms and must take the same measures, it would have been of interest to investigate the 2 diabetes groups separately, as the psychological aspect and the person's interest in change and self-management may differ.

The most common function besides communication appeared to be self-monitoring of blood glucose, a function in 5 of the 6 apps. This is not surprising, as self-monitoring of blood glucose is found to lower HbA_{1c} [54]. However, it is still debated whether persons with type 2 diabetes benefit from blood glucose measures [55]. Additional functions are crucial, as an app should offer more than the traditional paper diary [8]. It is alarming that few developers of apps discuss behavior change theory or treatment guidelines, even though the majority of available apps, including those in this review, aim to change behavior or habits [15,22]. Treatment guidelines or behavior change theory should guide intervention development as both can increase the quality of the app [7,8,9], and possibly this is best achieved if researchers from several fields work together (eg, health care researchers and technological engineers). Further, a linearity between behavior change theory and effects has been suggested [32], and its use would strengthen all arguments regarding the practical use of the app. One example may be the input of values, where greater personal reflection is gained through manual input [35]. However, manual input can be time consuming and there is a larger risk of faulty input than with automatic input. As pointed out in 2 of the current trials [48,50], tracking of blood pressure might reveal those with an out-of-range blood pressure in need of medication that might not be identified in a general practitioner office visit. The same argument is valid for monitoring blood glucose values: thus, these 2 functions of blood glucose and blood pressure remain important to reach the treatment goals for diabetes.

Another point of interest is that medication tracking was a function in just 3 of the 6 apps [46,49,51], meaning that the individualized feedback in the remaining 3 apps [47,48,50] does not evaluate usage of medication that might be critical for the patients. Nor was there tracking of psychological measures in the 6 trials we included. Measures of this kind are not much used in apps for diabetes; however, they might provide useful information for the patient and the provider [33]. For example, depressive symptoms are a significant risk in those with diabetes [6]. There is a known association between anxiety and self-focus on bodily symptoms [48] that may be triggered by self-monitoring of blood glucose, and this may support the need for measures to reveal such symptoms as they can degrade self-management and glycemic control.

Several outcomes were used to evaluate the apps' ease of use among the studies in this review: patient and/or HCP satisfaction, degree of technical issues like delay of data transfer, use of time, acceptance of feedback, and usefulness. However, as none of these used validated measures, the evidence regarding app-related evaluations such as satisfaction, acceptability, usability, and feasibility as outcomes from this systematic review is weak. These concepts will, however, remain important to ensure that the apps are used and 1 paper highlighted the association between use or satisfaction and effect [47]. As Kim and colleagues [47] argue in their paper, a well-functioning tool must be provided to increase use and satisfaction and to decrease the risk of deteriorating glycemic control.

As a more standardized outcome, HbA_{1c} was reported in 4 of the 6 trials in our systematic review, and this seems to be the most common outcome in diabetes trials regardless of intervention. Whether HbA_{1c} is an appropriate outcome in trials aiming for lifestyle change is a relevant question that we have debated previously [18], as has Garabedian [23].

The current included trials did not report lifestyle measures such as physical activity or dietary habits, making an evaluation of their effect on lifestyle difficult. None of the studies reported adverse events or safety as an outcome, except that Zhou et al [51] reported that in their trial they were infrequent in both groups. We regard the lack of focus on adverse events as an important weakness as there is a risk of hypoglycemia attached to the use of apps because of possible changes in medication or behavior.

A total of 3 studies found significant effects in their primary outcomes: 1 in SBP [48] and 2 in HbA_{1c} [50,51]. The remaining 3 studies remained inconclusive. One possible explanation might relate to the patients' interest in data tracking and the recurrent reminders of having a chronic illness. This leads back to the identification of the individual's interest in mHealth and also that it might be useful for some, but not all. Attitudes and intentions should be clarified for an app to be useful, regardless of functions [56]. Both researchers and clinicians must remember that patients often have limited interest in tracking their health, and for the app to be useful, there should be some clarification of the patients' expectancies of the app, its usefulness, and possible adverse events [57].

The overall methodological quality of the included trials was low, with small samples and weak designs, which threatens the generalizability and reliability of the results. The lack of detailed description of the comparison group is a limitation among all of the included studies, and although the national guidelines are often used to define usual care, we cannot evaluate the content of the comparison group. Further, poor reporting of study details made efficient evaluation difficult and important details regarding group allocation, blinding, and preregistration in trial databases were unclear. Use of a standardized guideline for reporting, such as eHealth Consolidated Standards of Reporting Trials (eHealth CONSORT) [58], would have improved the reporting significantly, supported by earlier research confirming that adherence to such guidelines is low in medical informatics [59]. A pooled analysis was not possible because of the high ROB and heterogeneous outcomes. The ROB domains regarding blinding have been discussed among the authors, as blinding of such interventions is often difficult [60]; however, blinding of the outcome studied should be possible. The authors have discussed the domain "other" in ROB, and possibly reporting according to the eHealth CONSORT [58] could improve the reporting and evaluation of intervention trials. However, this is only applicable for randomized trials. A strength among the studies was the low dropout rates, which must be considered unusual compared with earlier research in this field [61].

Limitations

This review has some limitations. We performed a systematic search using rigorous methods; however, as MeSH terms are still new in the field of technology in health research, the use of keywords might have contaminated our search. A 2014 consensus paper [62] stated that the term "app" should be used before "application," which might positively influence the field in the years to come but to date this is still not frequently applied. Our search had a high N, closely related to such contamination. This review only assessed published trials, and we cannot rule out any publication bias. The inclusion of trials in this review was demanding as the interventions often had additional communication outside the scope of this review, but we cannot provide accurate numbers on how many studies this applies to, largely because of the heterogeneity among the excluded trials and lack of resources to handle this information systematically. However, we argue that the heterogeneity of the interventions represents the field and that there is still no consensus regarding preferred communication with HCP. Further, we did not assess the quality of the apps included in the current trials, as this was not in the scope of our work, and this might represent a limitation.

The application of the Cochrane ROB tool [45] may represent a limitation, as this tool might not be applicable in pragmatic technology trials, related to the previous discussion on blinding. Another possible threat in interventions and trials evaluating use of apps might be the less frequent usage of the app over time. Because of the short follow-ups, we cannot confirm whether this is a decline in use or whether it represents a more dynamic use of an app in periods where the persons with diabetes want to or should use the app more. However, if the app use changes substantially during the study period, assumptions might be drawn on the wrong basis, and none of the current trials included app use as an outcome, either in terms of which functions were mostly used or app use frequency through the study period (eg, in terms of number of times that apps were accessed). The health literacy aspect can also contribute, as participants' use may decline if the intervention/app is too difficult to understand [10].

Implications

This systematic review has not produced specific evidence for stakeholders regarding future decisions. We believe that the next generation of patients with diabetes has different needs and requests for the health care system and technology development and use than what is available today. Another important point might be the conflicting interests among the health care researchers and technology researchers regarding patent or economic interests in the device or app they are testing, and possibly, their drive for positive results could bias the published material. Therefore, it is even more important in this field to perform a thorough and unbiased evaluation and report the results from controlled trials regardless of their outcomes, which would form new evidence and provide benefit and information for policymakers. We suggest using a health technology assessment framework like the Model for the Assessment of Telemedicine (MAST) [63] for a scientific evaluation of important domains when testing technology and also the eHealth CONSORT for reporting scientific trials [58]. Further, the availability of hundreds of apps makes it difficult to find clinically relevant apps, and the need for updated reviews will continue to be large in the future [23]. There is, however, a need for higher methodological quality trials to improve the field and inform future reviews. The studies in this review were mostly pilot studies with small sample sizes and interventions that might be too extensive to be implemented in real-life contexts.

Conclusion

The conclusions from this systematic review are limited. The unclear and poor methodological quality of this emerging research field is of major concern, and although 3 studies found that apps with integrated feedback significantly improve the primary outcome, the evidence has limitations because of its poor methodological quality. Mobile apps will be a part of the health care system in the future; therefore, we require robust research in this area to make the right choices for the patient, for the health care system, and for society.

Acknowledgments

HH prepared, conducted, and sorted the systematic search, analyzed and applied the ROB tool, and wrote and edited the paper. LR commented on the search and sorted the results, applied the ROB tool, and edited the paper. AKW applied the ROB tool, commented on the results, and edited the paper. MCS edited the paper.

We would like to express our gratitude to the Learning Centre and Library section head Trine Remvik and the academic librarian Malene Wøhlk Gundersen at the Learning Centre and Library at the Institute of Nursing and Health Promotion at Oslo and Akershus University College for their help with the systematic search. In addition, senior librarian Hilde I Flaatten at the Medical Library at the University of Oslo contributed great expertise in tailoring the searches. We would also like to express our gratitude to Astrid Torbjørnsen at the Oslo and Akershus University College of Applied Sciences for valuable insights and discussion throughout the search process and when writing the paper.

The Norwegian Research Council and the Oslo and Akershus University College of Applied Sciences funded this project, and the authors express their gratitude.

We express our gratitude to the 3 anonymous reviewers of JMIR for their insightful and critical comments on the manuscript of this paper.

Conflicts of Interest

None declared.

References

1. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract 2014 Feb;103(2):137-149. [doi: 10.1016/j.diabres.2013.11.002] [Medline: 24630390]
2. da Rocha FJ, Ogurtsova K, Linnenkamp U, Guariguata L, Seuring T, Zhang P, et al. IDF Diabetes Atlas estimates of 2014 global health expenditures on diabetes. Diabetes Res Clin Pract 2016 Jul;117:48-54. [doi: 10.1016/j.diabres.2016.04.016] [Medline: 27329022]
3. International Diabetes Foundation. IDF Diabetes Atlas, 7th edition. Brussels: International Diabetes Federation; 2015.
4. Gregg EW, Li Y, Wang J, Burrows NR, Ali MK, Rolka D, et al. Changes in diabetes-related complications in the United States, 1990-2010. N Engl J Med 2014 Apr 17;370(16):1514-1523. [doi: 10.1056/NEJMoa1310799] [Medline: 24738668]
5. Schram MT, Baan CA, Pouwer F. Depression and quality of life in patients with diabetes: a systematic review from the European depression in diabetes (EDID) research consortium. Curr Diabetes Rev 2009 May;5(2):112-119 [FREE Full text] [Medline: 19442096]
6. Lin EH, Rutter CM, Katon W, Heckbert SR, Ciechanowski P, Oliver MM, et al. Depression and advanced complications of diabetes: a prospective cohort study. Diabetes Care 2010 Feb;33(2):264-269 [FREE Full text] [doi: 10.2337/dc09-1068] [Medline: 19933989]
7. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
8. Goyal S, Cafazzo JA. Mobile phone health apps for diabetes management: current evidence and future developments. QJM 2013 Dec;106(12):1067-1069 [FREE Full text] [doi: 10.1093/qjmed/hct203] [Medline: 24106313]
9. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
10. Caburnay CA, Graff K, Harris JK, McQueen A, Smith M, Fairchild M, et al. Evaluating diabetes mobile applications for health literate designs and functionality, 2014. Prev Chronic Dis 2015;12:E61 [FREE Full text] [doi: 10.5888/pcd12.140433] [Medline: 25950568]
11. Klonoff DC, Kerr D. Digital diabetes communication: there's an app for that. J Diabetes Sci Technol 2016 Sep;10(5):1003-1005. [doi: 10.1177/1932296816660210] [Medline: 27464752]
12. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013 Jan 28;346:f167 [FREE Full text] [Medline: 23358487]

13. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Med Care 2013 Aug;51(8 Suppl 3):S73-S79 [FREE Full text] [doi: 10.1097/MLR.0b013e31829b1d84] [Medline: 23774513]
14. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016 Sep;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]
15. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
16. Heintzman ND. A digital ecosystem of diabetes data and technology: services, systems, and tools enabled by wearables, sensors, and apps. J Diabetes Sci Technol 2015 Dec 20;10(1):35-41 [FREE Full text] [doi: 10.1177/1932296815622453] [Medline: 26685994]
17. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: A qualitative inquiry. J Telemed Telecare 2016 Apr;22(3):172-178. [doi: 10.1177/1357633X15595178] [Medline: 26199275]
18. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian Randomized Controlled Trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
19. Ribu L, Holmen H, Torbjørnsen A, Wahl AK, Grøttland A, Småstuen MC, et al. Low-intensity self-management intervention for persons with type 2 diabetes using a mobile phone-based diabetes diary, with and without health counseling and motivational interviewing: protocol for a randomized controlled trial. JMIR Res Protoc 2013 Aug 26;2(2):e34 [FREE Full text] [doi: 10.2196/resprot.2768] [Medline: 23978690]
20. Torbjørnsen A, Jenum AK, Småstuen MC, Arsand E, Holmen H, Wahl AK, et al. A low-intensity mobile health intervention with and without health counseling for persons with type 2 diabetes, part 1: baseline and short-term results from a randomized controlled trial in the Norwegian part of RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]
21. Holmen H, Wahl A, Torbjørnsen A, Jenum AK, Småstuen MC, Ribu L. Stages of change for physical activity and dietary habits in persons with type 2 diabetes included in a mobile health intervention: the Norwegian study in RENEWING HEALTH. BMJ Open Diabetes Res Care 2016 May;4(1):e000193 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000193] [Medline: 27239317]
22. Breland JY, Yeh VM, Yu J. Adherence to evidence-based guidelines among diabetes self-management apps. Transl Behav Med 2013 Sep;3(3):277-286 [FREE Full text] [doi: 10.1007/s13142-013-0205-4] [Medline: 24073179]
23. Garabedian LF, Ross-Degnan D, Wharam JF. Mobile phone and smartphone technologies for diabetes care and self-management. Curr Diab Rep 2015 Oct 12;15(12):1-9. [doi: 10.1007/s11892-015-0680-8] [Medline: 26458380]
24. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
25. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
26. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
27. Thorpe CT, Fahey LE, Johnson H, Deshpande M, Thorpe JM, Fisher EB. Facilitating healthy coping in patients with diabetes: a systematic review. Diabetes Educ 2013 Oct;39(1):33-52 [FREE Full text] [doi: 10.1177/0145721712464400] [Medline: 23073967]
28. Piette JD, Weinberger M, Kraemer FB, McPhee SJ. Impact of automated calls with nurse follow-up on diabetes treatment outcomes in a Department of Veterans Affairs Health Care System: a randomized controlled trial. Diabetes Care 2001 Feb;24(2):202-208. [Medline: 11213866]
29. Whittemore R, Melkus GD, Sullivan A, Grey M. A nurse-coaching intervention for women with type 2 diabetes. Diabetes Educ 2004;30(5):795-804. [Medline: 15510531]
30. Nicolucci A, Kovacs BK, Holt RI, Comaschi M, Hermanns N, Ishii H, Dawn2 Study Group. Diabetes attitudes, wishes and needs second study (DAWN2): cross-national benchmarking of diabetes-related psychosocial outcomes for people with diabetes. Diabet Med 2013 Jul;30(7):767-777. [doi: 10.1111/dme.12245] [Medline: 23711019]
31. Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. Diabetes Care 2001 Mar;24(3):561-587. [Medline: 11289485]
32. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

33. Alanzi TM, Istepanian RS, Philip N. An integrated model for cognitive behavioural therapy for mobile diabetes self-management system. In: Conf Proc IEEE Eng Med Biol Soc. 2014 Presented at: Conf Proc IEEE Eng Med Biol Soc; 2014; Chicago p. 5393-5396. [doi: 10.1109/EMBC.2014.6944845]
34. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Jul 25;34(9):1934-1942. [doi: 10.2337/dc11-0366] [Medline: 21788632]
35. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
36. Orsama A, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013 Aug;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]
37. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 05;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
38. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's life! Mobile and Web-based monitoring and feedback tool embedded in primary care increases physical activity: a cluster randomized controlled trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
39. Holmen H, Wahl AK, Småstuen M, Ribu L. Functions and effectiveness of mobile apps with integrated health care personnel communication for patients with diabetes mellitus: a systematic review of controlled trials. PROSPERO, 2016. 2016. URL: http://www.crd.york.ac.uk/prospero/display_record.asp?src=trip&ID=CRD42016038640 [accessed 2017-06-19] [WebCite Cache ID 6rKjqI7YU]
40. Holmen H. Mobile apps with integrated health care personnel communication: protocol for a systematic review. URL: http://www.hib.no/contentassets/994235d9122e44c98a99bbb1ff00a518/program-solstrand-2016.pdf [accessed 2016-11-18] [WebCite Cache ID 6m6yqBs0j]
41. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012 Feb 09;1:2 [FREE Full text] [doi: 10.1186/2046-4053-1-2] [Medline: 22587842]
42. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
43. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;349:g7647 [FREE Full text] [Medline: 25555855]
44. 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]
45. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken: Wiley; 2008.
46. Fioravanti A, Fico G, Salvi D, García-Betances RI, Arredondo MT. Automatic messaging for improving patients engagement in diabetes management: an exploratory study. Med Biol Eng Comput 2015 Dec;53(12):1285-1294. [doi: 10.1007/s11517-014-1237-8] [Medline: 25564181]
47. Kim H, Choi W, Baek EK, Kim YA, Yang SJ, Choi IY, et al. Efficacy of the smartphone-based glucose management application stratified by user satisfaction. Diabetes Metab J 2014 Jun;38(3):204-210 [FREE Full text] [doi: 10.4093/dmj.2014.38.3.204] [Medline: 25003074]
48. Logan AG, Irvine MJ, McIsaac WJ, Tisler A, Rossos PG, Easty A, et al. Effect of home blood pressure telemonitoring with self-care support on uncontrolled systolic hypertension in diabetics. Hypertension 2012 Jul;60(1):51-57 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.111.188409] [Medline: 22615116]
49. Takenga C, Berndt R, Musongya O, Kitero J, Katoke R, Molo K, et al. An ICT-based diabetes management system tested for health care delivery in the African context. Int J Telemed Appl 2014;2014:437307 [FREE Full text] [doi: 10.1155/2014/437307] [Medline: 25136358]
50. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215. [doi: 10.1177/1932296814526495] [Medline: 24876569]
51. Zhou W, Chen M, Yuan J, Sun Y. Welltang—a smart phone-based diabetes management application—improves blood glucose control in Chinese people with diabetes. Diabetes Res Clin Pract 2016 Jun;116:105-110. [doi: 10.1016/j.diabres.2016.03.018] [Medline: 27321324]
52. Carter A, Liddle J, Hall W, Chenery H. Mobile phones in research and treatment: ethical guidelines and future directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]

53. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med 2014 Jul 24;371(4):372-379. [doi: 10.1056/NEJMhle1403384] [Medline: 25054722]
54. Clar C, Barnard K, Cummins E, Royle P, Waugh N, Aberdeen Health Technology Assessment Group. Self-monitoring of blood glucose in type 2 diabetes: systematic review. Health Technol Assess 2010 Mar;14(12):1-140 [FREE Full text] [doi: 10.3310/hta14120] [Medline: 20226138]
55. Malanda UL, Welschen LM, Riphagen II, Dekker JM, Nijpels G, Bot SD. Self-monitoring of blood glucose in patients with type 2 diabetes mellitus who are not using insulin. Cochrane Database Syst Rev 2012 Jan 18;1:CD005060. [doi: 10.1002/14651858.CD005060.pub3] [Medline: 22258959]
56. Dobson KG, Hall P. A pilot study examining patient attitudes and intentions to adopt assistive technologies into type 2 diabetes self-management. J Diabetes Sci Technol 2015 Mar;9(2):309-315 [FREE Full text] [doi: 10.1177/1932296814560395] [Medline: 25427965]
57. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. You get reminded you're a sick person: personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015 Aug;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
58. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
59. Augestad KM, Berntsen G, Lassen K, Bellika JG, Wootton R, Lindsetmo RO, Study Group of Research Quality in Medical Informatics and Decision Support (SQUID). Standards for reporting randomized controlled trials in medical informatics: a systematic review of CONSORT adherence in RCTs on clinical decision support. J Am Med Inform Assoc 2012 Jan;19(1):13-21 [FREE Full text] [doi: 10.1136/amiajnl-2011-000411] [Medline: 21803926]
60. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet 2002 Feb 23;359(9307):696-700. [doi: 10.1016/S0140-6736(02)07816-9] [Medline: 11879884]
61. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
62. Lewis TL, Boissaud-Cooke MA, Aungst TD, Eysenbach G. Consensus on use of the term “app” versus “application” for reporting of mHealth research. J Med Internet Res 2014 Jul;16(7):e174 [FREE Full text] [doi: 10.2196/jmir.3460] [Medline: 25033233]
63. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: MAST. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]

Abbreviations

BMI: body mass index
BP: blood pressure
CINAHL: Cumulative Index to Nursing and Allied Health Literature
CONSORT: Consolidated Standards of Reporting Trials
DBP: diastolic blood pressure
EMBASE: Excerpta Medica database
HADS: Hospital Anxiety and Depression Scale
HCP: health care personnel
HbA$_{1c}$: hemoglobin A1c
HDL: high-density lipoprotein
LDL: low-density lipoprotein
MAST: Model for the Assessment of Telemedicine
MEDLINE: Medical Literature Analysis and Retrieval System Online
NA: not applicable
PRO: patient-reported outcome
PROSPERO: International Prospective Register of Systematic Reviews
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
ROB: risk of bias
SBP: systolic blood pressure
SMS: short message service
TTM: transtheoretical model stages of change
WHO: World Health Organization

Edited by A Keepanasseril; submitted 24.11.16; peer-reviewed by L Quinlan, B Holtz; comments to author 04.01.17; revised version received 17.01.17; accepted 28.04.17; published 23.06.17.

Please cite as:
Holmen H, Wahl AK, Cvancarova Småstuen M, Ribu L
Tailored Communication Within Mobile Apps for Diabetes Self-Management: A Systematic Review
J Med Internet Res 2017;19(6):e227
URL: http://www.jmir.org/2017/6/e227/
doi:10.2196/jmir.7045
PMID:28645890

Original Paper

Tracing the Potential Flow of Consumer Data: A Network Analysis of Prominent Health and Fitness Apps

Quinn Grundy[1], RN, PhD; Fabian P Held[2], PhD; Lisa A Bero[1], PhD

[1]Charles Perkins Centre, Faculty of Pharmacy, The University of Sydney, Sydney, Australia
[2]Charles Perkins Centre, Faculty of Science, The University of Sydney, Sydney, Australia

Corresponding Author:
Quinn Grundy, RN, PhD
Charles Perkins Centre
Faculty of Pharmacy
The University of Sydney
D17, Level 6 The Hub
Charles Perkins Centre
Sydney, 2006
Australia
Phone: 61 286271566
Fax: 61 286271601
Email: quinn.grundy@sydney.edu.au

Abstract

Background: A great deal of consumer data, collected actively through consumer reporting or passively through sensors, is shared among apps. Developers increasingly allow their programs to communicate with other apps, sensors, and Web-based services, which are promoted as features to potential users. However, health apps also routinely pose risks related to information leaks, information manipulation, and loss of information. There has been less investigation into the kinds of user data that developers are likely to collect, and who might have access to it.

Objective: We sought to describe how consumer data generated from mobile health apps might be distributed and reused. We also aimed to outline risks to individual privacy and security presented by this potential for aggregating and combining user data across apps.

Methods: We purposively sampled prominent health and fitness apps available in the United States, Canada, and Australia Google Play and iTunes app stores in November 2015. Two independent coders extracted data from app promotional materials on app and developer characteristics, and the developer-reported collection and sharing of user data. We conducted a descriptive analysis of app, developer, and user data collection characteristics. Using structural equivalence analysis, we conducted a network analysis of sampled apps' self-reported sharing of user-generated data.

Results: We included 297 unique apps published by 231 individual developers, which requested 58 different permissions (mean 7.95, SD 6.57). We grouped apps into 222 app families on the basis of shared ownership. Analysis of self-reported data sharing revealed a network of 359 app family nodes, with one connected central component of 210 app families (58.5%). Most (143/222, 64.4%) of the sampled app families did not report sharing any data and were therefore isolated from each other and from the core network. Fifteen app families assumed more central network positions as gatekeepers on the shortest paths that data would have to travel between other app families.

Conclusions: This cross-sectional analysis highlights the possibilities for user data collection and potential paths that data is able to travel among a sample of prominent health and fitness apps. While individual apps may not collect personally identifiable information, app families and the partners with which they share data may be able to aggregate consumer data, thus achieving a much more comprehensive picture of the individual consumer. The organizations behind the centrally connected app families represent diverse industries, including apparel manufacturers and social media platforms that are not traditionally involved in health or fitness. This analysis highlights the potential for anticipated and voluntary but also possibly unanticipated and involuntary sharing of user data, validating privacy and security concerns in mobile health.

(J Med Internet Res 2017;19(6):e233) doi:10.2196/jmir.7347

KEYWORDS

mobile health; smartphone; privacy

Introduction

Mobile health is an exploding market, having doubled in just 2.5 years and reaching over 100,000 apps with market revenues projected to grow to US $26bn by 2017[1]. Every major technology company, including Apple, Google, and even Uber, has signaled an intention to enter the “digital health” market [2]. In mobile health, diverse stakeholders including companies, health professionals, and consumers perceive the collection, analysis, and sharing of user data to be of especially high value. The promise of big health data heralds shifts in health care toward consumer self-management through wearables and mobile apps with a focus on prevention and optimization [3]. For consumers, health and fitness apps also promise crowd-sourced information and social networks of support [4]. Researchers are also designing and testing methods to generate insights from the wealth of information shared over social media, such as Twitter, by users of mobile health and fitness apps [5].

A great deal of consumer data, collected actively through consumer reporting or passively through sensors, is shared among apps. App developers increasingly allow their programs to communicate with other apps, sensors, and Web-based services, which are promoted as features to potential users. This is typically done through shared application programming interfaces (API). A survey of 5000 app developers, representing almost 11,000 mobile health apps found that fitness and nutrition apps are among the most advanced in sharing via APIs [6]. This sharing most commonly occurs with data aggregators (eg, Apple’s HealthKit, a software platform that pools data from multiple health apps), sensors such as wearables, and directly between apps [6]. However, app developers may also choose to share consumer data with a variety of tools including analytic tools, social media APIs, performance tools and digital advertisers [6].

Apps serve an increasingly broad range of functions, some of which involve access to and the creation of vast amounts of highly personal data including a user’s location, text messages, or access to the mobile phone’s camera or photos [7]. Much of this information may be essential to the app’s functionality; however, the widespread collection, retention, and sharing of user data through apps has also created concern related to consumer privacy and the security of health-related data [8-10]. Because health apps by design often have access to personal health information, they pose a higher risk to consumers’ privacy; this information is often highly valuable to third parties, heightening the risk [10]. However, health apps also routinely pose risks related to information leaks, information manipulation, and loss of information [10]. There has been less investigation into the kinds of user data developers are likely to collect, and who might have access to it [7]. For example, Li developed a privacy threat model based on a health-related social networking site that accounts for multiple actors in the data usage and sharing network [11]. One principle threat accounted for in this model is user profiling across multiple sites: third parties can link multiple user accounts across apps to create aggregated user profiles and a more complete picture of a consumer’s social network. These aggregate profiles are then monetized and used for marketing, screening prospective tenants or employees, or maliciously for identity fraud [11].

Generating insights from mobile health data that can be translated into public health benefit will largely depend on the actions of those who own these data and the decisions they make regarding what to collect, how it will be aggregated and analyzed, and whether to share. Thus, it is essential to identify key actors and make their relationships transparent. Unfortunately, there is little transparency around these practices. The objective of this study was to describe how consumer data generated from mobile health apps might be distributed and reused with the aim of identifying potential user privacy and security risks. We investigated the self-reported collection and sharing of user data, while recognizing that this likely underreports the extent of both data collection and distribution. We investigated the nature of user data collection in the form of “permissions” that developers requested. We also traced the network of self-reported data sharing among a sample of prominent mobile health and fitness apps available in the United States, Canada, and Australia to better understand the potential for user data distribution within these networks.

Methods

Study Design

We conducted a structured content analysis of a purposive sample of prominent health and fitness apps available in the United States, Canada, and Australia. We then conducted a social network analysis of apps’ self-reported data sharing possibilities to understand how consumer data collected through the mobile platform might travel through this network.

Sampling

We generated a purposive sample of prominent apps available in the United States, Canada, or Australia Google Play and iTunes app stores during November 2015. We employed purposive sampling to identify apps that were most likely to have data sharing ties, meaning they were likely to impact a large number of consumers. Due to the localized and personalized nature of app store search algorithms and rapidly changing population of apps, we triangulated two sampling strategies [12]: (1) a crawling program that systematically sampled the top-ranked 100 apps from the iTunes and Google Play app stores in the United States, Canada, and Australia and (2) purposive sampling of high-profile apps. These strategies were complementary by allowing exploration of data sharing relationships for developers that are both “well established” or “up and coming” so as to capture the dynamic nature of app development.

We first systematically sampled apps on a weekly basis that were ranked by the app stores as “top 100” using a crawling program that interacted directly with the stores’ API and automatically extracted the apps' metadata. We identified 441 apps that were ranked in the top 100 in at least one country store during November 2015.

Beginning on November 1, 2015, we screened mainstream media (BBC, The New York Times, and The Guardian) and industry newsletters (MobiHealth News and RockHealth weekly) on a daily basis. We extracted the metadata for any health-related app receiving coverage. We continued screening until we generated a sample of 50 apps, which we determined would complement our systematic sample, allowing some representation for apps new to the market. We reached a sample of 50 on January 21, 2016.

Two researchers independently screened the sample of 491 apps for inclusion and excluded obvious duplicates. The inclusion criteria were that the app: explicitly pertained to a medical (eg, diabetes) or health condition (eg, obesity), health risk factor (eg, smoking), or health behavior (eg, walking), and provided guidance or a recommendation (eg, a workout program), tracked or recorded personal data, or made a health claim (eg, "improve heart health," "lose weight," or "reduce your anxiety"). Discrepancies were resolved through discussion until consensus was reached with input from a third researcher when necessary.

Data Collection

We created an a priori coding instrument, which has been published in REDCap [13], a secure, Web-based application used to collect and manage data hosted at The University of Sydney (Multimedia Appendix 1). The instrument, based on a systematic review of methods for app content analysis [12], covered 4 domains: (1) app characteristics, (2) partnerships and affiliations, (3) developer and funding characteristics, and (4) permissions. We piloted the instrument on a sample of 70 randomly selected apps from our sample to ensure comprehensiveness of survey items and a high degree of inter-rater reliability.

We extracted data between December 12, 2015 and April 1, 2016 from app store descriptions, websites linked from the app store description and Google searches, with Google Play as a default content source when apps were distributed in both app stores. Two researchers independently extracted data related to partnerships and affiliations, defined as "any mention of a branded product, service or company, especially noting partnerships, collaborations, sponsors (funders), or brands," and resolved any discrepancies through discussion until consensus was reached.

For apps available in Google Play, developers disclose how their apps will interact with the user's device and personal information through reporting permissions [7]; Apple does not require developers to report permissions for apps distributed through iTunes. Permissions data were extracted for apps from the Google Play store description as reported by developers. Google encourages developers to request the minimum number of permissions required for an app's functionality [14]; we did not, however, judge whether this was the case.

Data Analysis

We inductively categorized relationships within the category "partnerships and affiliations," one of which was data sharing. We defined instances of data sharing as any mention of a digital app, website, platform, sensor, wearable, or other smart device that was reported by the developer as a "partner" or as having the ability to share data. Examples of these promotional-type messages included: "Integration with Google Fit & MyFitnessPal — your running apps [sic] perfect companion"; or, "Easy, automatic exercise calorie counter, syncing with Fitbit, Withings, Jawbone and Garmin trackers and weight scales."

The spelling of names for a large number of apps varied between platforms, countries, and store descriptions. These included variations in spelling ("plus", instead of "+"), free and paid versions of the same app, and different naming practices between iTunes and Google Play. We grouped different instances of the same app through a two-stage process, first identifying similar names of apps automatically using approximate string matching, with a second author cross checking results [15].

We conducted descriptive analyses of app and developer characteristics in Microsoft Excel. We conducted an analysis of the network of sharing of user data among apps to understand the potential for data distribution in R using analysis packages igraph (1.0.1) and tnet (3.0.14). For this analysis, we grouped sampled apps together into app "families" that were offered by the same developer or owned by the same entity, again using approximate string matching, with a second author cross checking and expanding groups by joint ownership [15]. This approach assumed that the app families shared a common owner who had access to the data that were collected by their family of apps and the power to formulate the apps' terms and conditions that grant legal access and in most cases, ownership, of these data.

The apps in our sample reported sharing user data with other included apps as well as apps that were not originally sampled in the methods described above. Both the initially sampled and these secondary apps were included as nodes in the network analysis. We constructed the network by connecting these app "families" with data sharing links whenever data sharing was reported. As our data does not give us any insights into the content of shared data or the direction of the flow, we considered these ties to be undirected and binary.

Descriptive measures of social network analysis provided a summary of the observed network. We calculated network density as the ratio of observed versus theoretically possible connections in a fully connected network and degree as the number of neighboring nodes per app.

We quantified the position of an app (x) within the network by calculating its closeness centrality index. Closeness measures the potential for an app to access a piece of user data from anywhere in the network based on its network connections. Closeness to a single app (y) is defined as the inverse of the shortest path to y, or 0 if y cannot be reached from x. To calculate an index that compares all apps' network positions, we summed up these values across all other apps to describe the position [16] (Figure 1).

Finally, random walk community detection was used to further specify how groups of apps relate to each other based on the structure of the network. This method identifies clusters of apps that are more closely connected either directly or through their neighbors [17]. This is achieved by simulating a process of random communication between neighboring apps, where a “message” is passed on from neighbor to neighbor for a certain number of steps. Essentially this process measures the probability that a message sent from x ends up at y.

Figure 1. Closeness centrality index formula.

$$C(x) = \sum_{y!=x} d(x,y)^{-1}$$

Results

The sample included 297 unique apps published by 231 unique developers. The majority had been sampled using the crawling program 265 (265/297, 89.2%) and 32 (32/297, 10.8%) from media sources. The majority of apps were available in both Google Play and iTunes (202/297, 68.0%), were free to download (172/297, 57.9%) and provided a link to a privacy policy in store or on the linked website (217/297, 73.1%).

User Data Collection

We obtained data on permissions requested for the 241 apps that were available in Google Play, as iTunes does not report this data. Apps requested a total of 58 different permissions (mean 7.95, SD 6.57), ranging from 0 to 32 different permissions requested per app. The most common types of permissions requested related to Internet access including “full network access” (228/241, 94.6%) and “view network connections” (218/241, 90.5%). Seven apps did not request any permission. Two apps each requested 32 different permissions: Huawei Wear and Under Armour Record. All of the 26 apps that requested 20 or more different permissions were inductively categorized as activity monitors, which make use of GPS or an accelerometer such as UP by Jawbone, Map My Ride, or Nike+ Training Club, or as multi-focus health apps, which provided tailored insights related to diet, physical activity and sleep, such as Noom Coach and Microsoft Health.

Google Play classifies app permissions as “normal” or “dangerous” [14]. A “normal” permission requires access to data or resources outside of the app, but there is little risk to a user’s privacy or the operation of other apps; for example, control over the phone’s vibration. “Dangerous” permissions request data or resources that involve the user’s private information, stored data, or affect the operation of other apps; for example, granting the app access to a user’s contacts, confidential calendar information, or unique device ID. Of the 58 permissions requested by sampled apps, 26 are classified as “dangerous.” The most commonly requested “dangerous” permissions were “read storage” (195/241, 80.9% of sampled apps) and “modify storage” (189/241, 78.4% of sampled apps), which allow the app to read and write to the device’s internal storage.

User Data Sharing

We grouped the 297 apps into 222 app “families” (ie, those which shared a developer or an acquiring company). These app families reported sharing data with 137 secondary app families, leading to a network of 359 app-family-nodes, with one connected central component of 210 app families (210/359, 58.5%). This meant that there were undirected paths of data sharing that connected 210 app families directly or indirectly. Of these 210 connected app families, 75 app families (75/210, 35.7%) were health app families in our initial sample. Most (143/222, 64.4%) of the sampled app families did not report sharing any data and were therefore isolated from each other and from the core network. Thirty app families (30/222, 13.5%) from our sample reported only a single tie to another app family; 119 secondary app families were only identified in a data sharing relationship once. Within the connected component, any 2 app families were connected by a maximum of 6 steps, and with an average of 3.28 connections between any two apps. The density of the entire network was 0.005, meaning that only very few of the theoretically possible connections between all the app families were in fact realized.

The distribution of closeness centrality scores clearly separated the isolated app families and revealed a skewed distribution of app families’ positions in the core. We highlight the 15 highest scoring app families in Figure 2. These families take on particularly important roles in the network: they may be distributing or receiving user-generated data with many other apps. Table 1 describes characteristics of these fifteen central app families, 12 of which included apps that were in our original sample. The majority of these central app families were publicly traded corporations (10/15, 67%) and represented the technology (wearables, mobile health, and social media) and fashion (sports apparel) sectors.

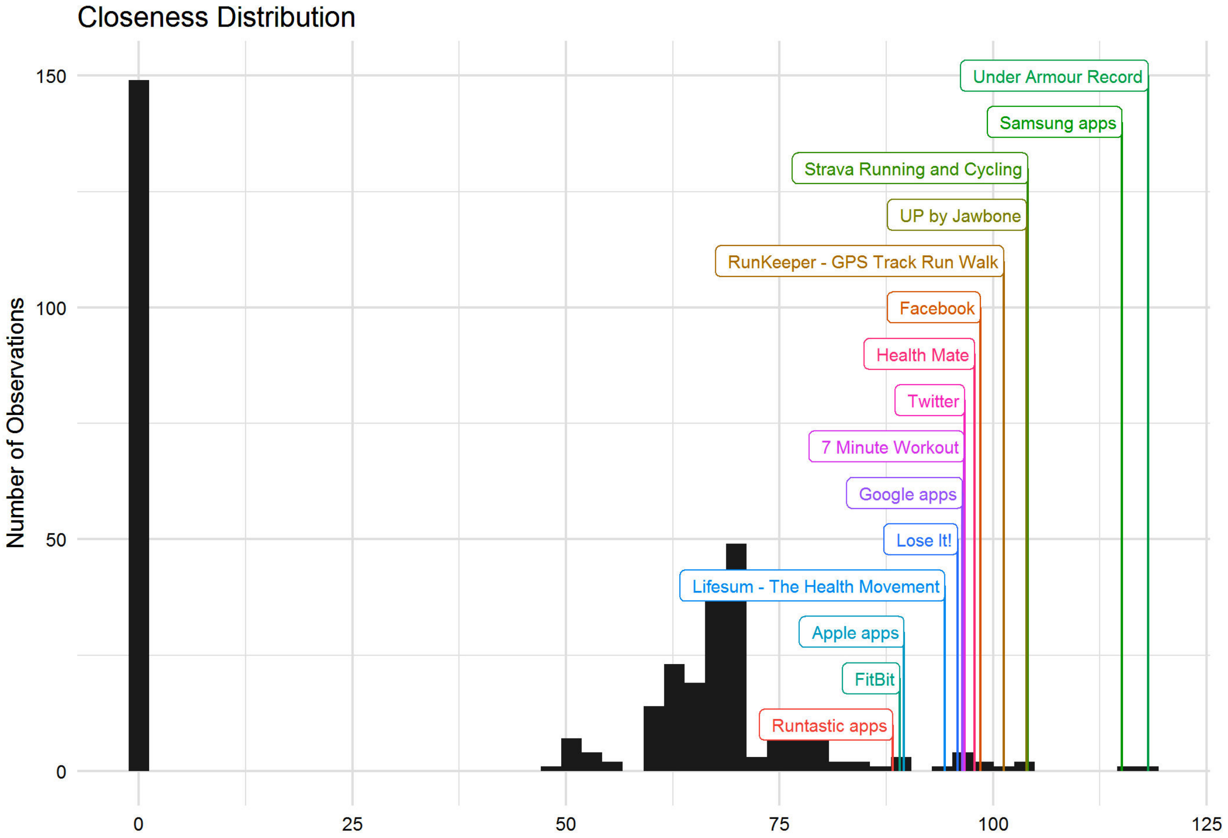
Table 1. Characteristics of centrally-positioned app families.

App family	Developer	Developer type	Sector
7 Minute Workout	Wahoo Fitness	Privately held	Wearables
Under Armour apps (under Armour Record, MyFitnessPal, MapMyFitness, Endomondo)	Under Armour Inc	Publicly traded	Sports apparel
RunKeeper - GPS Track Run Walk	FitnessKeeper, Inc. (acquired by ASICS)	Publicly traded	Mobile health
Lose It!	FitNow, Inc.	Privately held	Mobile health
Google apps (eg, Google Tracks, Gmail, Google Drive, Google Maps, Google Wallet, Google+, YouTube)	Alphabet Inc.	Publicly traded	Technology
Health Mate	Withings (acquired by Nokia)	Publicly traded	Technology
UP by Jawbone apps	Jawbone	Privately held	Wearables
Lifesum - The Health Movement	Lifesum	“Start-up”	Mobile health
Samsung apps (eg, S Health, SmartThings)	Samsung Group	Publicly traded	Technology
Strava Running and Cycling	Strava	Privately held	Mobile health
Facebook	Facebook, Inc.	Publicly traded	Social media
Twitter	Twitter Inc.	Publicly traded	Social media
Apple apps (eg, HealthKit, iTunes, Apple Watch, Apple TV)	Apple Inc.	Publicly traded	Technology
Fitbit	Fitbit Inc	Publicly traded	Wearables
Runtastic apps (eg, Runtastic Running & Fitness, Runtastic Me, Runtastic Road Bike)	Runtastic (acquired by Adidas)	Publicly traded	Mobile health

Finally, we simulated a random communication process between these app families, using the data-sharing network as the channel of information transmission. This allowed us to identify groups of app families that may be more likely to share user data with each other, as they are more tightly connected with each other in the network. The resulting network is shown in an interactive visualization in the Multimedia Appendix 1. The size of a dot represents how it was sampled: app families in our initial sample are bigger. Colors indicate the membership to communities identified through the walktrap clustering algorithm. We chose to simulate a random communication process of 4 steps because more steps did not noticeably increase the results’ modularity.

The result shows one large group of health and fitness app families that tended to link with each other (light blue). This group also included the major manufacturers of mobile devices and their respective health and fitness data aggregators (Google Tracks (now defunct)/Google Inc., HealthKit/Apple Inc. and S Health/Samsung Group). Furthermore, there were several communities that fanned out from individual apps. These included Samsung Apps, 7 Minute workout, and UP by Jawbone. Similar patterns were seen for apps that were not in our initial sample, including Apple apps, Facebook, and Twitter. In fact, the two social networking sites connected to the same community of app families.

Figure 2. Distribution of the closeness centrality index highlighting the top fifteen entries.



Discussion

Overview

This analysis provides a cross-sectional view of the possibilities for user data collection and potential paths that these data are able to travel among a sample of prominent health and fitness apps. The implication for consumers is that while individual apps may not collect personally identifiable information, app families and the partners with which they share data, may be able to aggregate consumer data, achieving a much more comprehensive picture of the individual consumer. We used information that was self-reported by developers in promotional materials such as app store descriptions as a proxy for forms of technological communication such as shared APIs. Thus, it could be inferred that the connections between apps in this study are those that developers choose to advertise, which users may value. This also means that this analysis likely underrepresents the extent of sharing of user data, as it cannot be presumed that user-selected networks are the lone third parties accessing consumer data [18]. For example, the US Federal Trade Commission analyzed 12 health and fitness apps and found they shared data such as usernames, names, email addresses, postal codes, geo-location, and exercise and diet habits with 76 third parties [19]. The extent of user data collection and sharing described here may only be the tip of an iceberg.

Principal Findings

Despite capturing only a fraction of the extent of user data sharing, we found that a core group of app families are sparsely connected and may be exchanging user data, though this analysis is not able to determine the directionality of exchange. This opens up new and important avenues for research: both into the ways that user-generated data is actually being shared among mobile health apps and third parties, but also and perhaps more importantly, into the ways that this data is combined and used.

The market of mobile health and fitness apps displays a remarkable dichotomy: most app families are stand-alone offerings that do not report any integration with the rest of the market. However within the connected core of app families, there are a number of hypothetical pathways for the sharing of users' data between app families and third parties, such as social media sites. Fifteen app families held crucial positions that may enable them to act as gatekeepers in the flow of data and to gather user information from diverse sources. Notably, the organizations behind these app families represent diverse industries, including apparel manufacturers and social media platforms that are not traditionally involved in health or fitness. Three entities—Apple, Facebook, Inc. and Twitter Inc.—were not even present in our original sample, yet assumed central positions in the network. These central network positions may also amplify an entity's ability to aggregate consumer-generated data, which has not only privacy, but market implications. These central network positions may be akin to monopolies given the value of consumer data, and anti-trust regulators are working to ensure data competition [20]. App developers are likely incentivized to build Facebook, Twitter or Apple Watch connectivity through their APIs due to their large consumer bases, which may put developers that are more circumspect about data sharing at a competitive disadvantage.

On an average, our sample of Android health and fitness apps requested a higher number of permissions than apps in general [7]. Pew Research Center analyzed over 1 million apps available in Google Play in 2014 and found that apps requested on average 5 permissions, in comparison to the nearly 8 permissions requested per health app in our sample [7]. Similar to apps in general, sampled Android health apps most commonly requested permissions relating to Internet access, which allows for the delivery of targeted advertising and the ability to access the phone's storage, which allows apps to not only save content to the device, but also to access personal data stored by other apps [7]. Sampled Android health and fitness apps also tended to request "dangerous" permissions more frequently than apps in general [7].

These findings validate privacy and security concerns in mobile health due to the sensitivity of health-related data as well as the collection of more kinds of personally identifiable data [8,9]. Particularly, the third party use of personal health-related information could result in employment, housing, or education-related discrimination, loss of insurance coverage or higher premiums, predatory advertising, or medical identity theft [21-23]. However, the implications of a privacy violation or security breach related to user-generated data via mobile health apps are largely speculative, emergent or not well understood. For example, recent news coverage highlighted the value of even a mobile phone number as an unregulated and unique digital identifier that can be used to link a user's personal data held by multiple companies from social networking sites to health apps to credit agencies [24]. This information is made available to developers and third parties when a user grants the permission "read phone status and identity."

This analysis highlights the potential for anticipated and voluntary, but also possibly unanticipated and involuntary sharing of user data, though we could not distinguish between the two. Though users may willingly engage in sharing their data, such as posting to social media, they may not fully appreciate the nature of data collection or the extent to which sharing occurs. A study with Facebook users found that users were under-informed about the nature of data collection within the app, and that even after receiving explicit information, many still did not fully understand the extent to which apps could access personal data [25]. Though Google Play allows developers to disclose the permissions requested, these disclosures are not accompanied by non-technical, lay descriptions of a given permission, nor what the likely and possible implications of granting such permission entails. Similarly, the sharing and protection of user data is far from transparent. In a study of privacy policies of the 300 most frequently rated health apps in the Apple and Google Play app stores, researchers found only 30% had a privacy policy with an average college-level reading grade level, and 2 out of 3 did not specifically address the app itself [26]. Thus, as mobile app companies gain unprecedented knowledge about app users, consumers know very little or nothing about how this knowledge is used [22].

This network analysis similarly highlights the potential for the unanticipated travel of user data through networks of health and fitness apps and their supporting entities. In fact, regulators are beginning to take note of the way that user data travels within app networks. A German privacy regulator recently ordered Facebook to stop collecting and storing the data of WhatsApp users, a social messaging app it had acquired [27]. The regulators argued that Facebook infringed data protection law when it announced a policy change that would allow unprecedented use of WhatsApp users' data, including phone numbers, without effective approval from users. Facebook is currently appealing this order [27].

Limitations

Our analysis of a purposive sample of apps is based upon data collected over a 4-month period from December 2015 through April 2016. Due to the highly dynamic nature of the app marketplace, it is likely that some of the apps are no longer available and that data are out of date. As we relied upon developer self-report, we could not verify whether the permissions requested are actually being used by the sampled apps or how user data is then analyzed or shared, thus, they serve as a proxy for the type of transmitted information. As these permissions are self-reported and are only reported by Android developers, it is likely that this represents an underestimate of permissions requested. This analysis highlights the possibilities for the sharing of consumer-generated data, but it does not differentiate between the potential and actual flow of data or the direction in which data travels. For example, there may be practical barriers to data sharing or this exchange may be subject to user control. Future research should consider performing traffic analyses, such as a man-in-the-middle approach, to verify the nature and extent of consumer data sharing and protection [8]. Furthermore, our analysis does not distinguish between user data that is shared consciously by the consumer, such as posting a workout to social media, and the user data that is collected by the app passively through the permissions requested, such as reading a user's contacts. However, this analysis does provide a unique snapshot in time that highlights the value of taking a network approach to studying health and fitness apps.

Conclusions

The possibilities afforded by the collection of consumer-generated, health-related data through mobile health and fitness apps are vast. However, the ways that consumer-generated, and sometimes sensitive personal information, is shared and passed on within the mobile ecosystem is far from transparent. The connections among apps in this study are only those that developers choose to advertise, and thus are likely only the tip of the iceberg. Thus, concerns regarding privacy and security should apply to this mobile ecosystem and not apps alone. Several major players in mobile health and fitness hold key gatekeeping positions in the sharing of consumer data, which amplifies their ability to amass consumer-generated data. Policymakers should particularly account for these actors in ensuring consumer privacy, security, and market competition.

Acknowledgments

The authors would like to acknowledge Lauren Kirby, BSN, RN, for her assistance with data collection, and Chris Klocheck, MSc, for development of the crawling programs. This project was supported by a Charles Perkins Centre summer student scholarship at the University of Sydney. The funder had no role in the review of or decision to publish the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Network of reported data sharing, highlighting communities of densely connected apps that have a higher potential of sharing users' data among each other, due to their direct and indirect connections. Bigger nodes were included in our initial sample, smaller nodes were mentioned by them as partners. Networks are presented using a force directed layout algorithm from the D3 library (http://d3js.org/).

[HTML File, 344KB - jmir_v19i6e233_app1.html]

References

1. Research2guidance. Research2guidance. Berlin, Germany; 2014. mHealth app developer economics 2014 URL: http://research2guidance.com/r2g/mHealth-App-Developer-Economics-2014.pdf [accessed 2017-06-16] [WebCite Cache ID 6rFuYyetL]
2. Wilbanks JT, Topol EJ. Stop the privatization of health data. Nature 2016 Dec 21;535(7612):345-348. [doi: 10.1038/535345a] [Medline: 27443724]
3. Swan M. Health 2050: the realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. J Pers Med 2012;2(3):93-118 [FREE Full text] [doi: 10.3390/jpm2030093] [Medline: 25562203]
4. Millington B. Smartphone apps and the mobile privatization of health and fitness. Crit Stud Media Comm 2014 Oct 29;31(5):479-493. [doi: 10.1080/15295036.2014.973429]
5. Vickey TA, Ginis KM, Dabrowski M. Twitter classification model: the ABC of two million fitness tweets. Transl Behav Med 2013 Sep;3(3):304-311 [FREE Full text] [doi: 10.1007/s13142-013-0209-0] [Medline: 24073182]
6. Research2guidance. Research2guidance. Berlin, Germany; 2015. mHealth App Developer Economics 2015 URL: http://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2015.pdf [accessed 2017-06-16] [WebCite Cache ID 6rFueEZl9]
7. Olmstead K, Atkinson M. Pew Research Center. Washington, DC: Pew Research Center; 2015 Nov 10. Apps permissions in the Google play store URL: http://www.pewinternet.org/2015/11/10/apps-permissions-in-the-google-play-store/ [accessed 2017-06-16] [WebCite Cache ID 6nb2Lvrhl]
8. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
9. Blenner SR, Köllmer M, Rouse AJ, Daneshvar N, Williams C, Andrews LB. Privacy policies of android diabetes apps and sharing of health information. JAMA 2016 Mar 08;315(10):1051-1052. [doi: 10.1001/jama.2015.19426] [Medline: 26954415]
10. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and android. JMIR Mhealth Uhealth 2015;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
11. Li J. A privacy preservation model for health-related social networking sites. J Med Internet Res 2015 Jul 08;17(7):e168 [FREE Full text] [doi: 10.2196/jmir.3973] [Medline: 26155953]
12. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
13. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
14. Google, Inc. Android. Mountain View, CA: Google, Inc; 2017. Android developers: system permissions URL: https://developer.android.com/guide/topics/security/permissions.html#normal-dangerous [accessed 2017-06-16] [WebCite Cache ID 6nb2UwaJO]
15. Navarro G. A guided tour to approximate string matching. ACM Comput Surv 2001;33(1):31-88. [doi: 10.1145/375360.375365]
16. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths. Social Networks 2010 Jul;32(3):245-251. [doi: 10.1016/j.socnet.2010.03.006]

17. Pons P, Latapy M. Computing communities in large networks using random walks. In: Yolum , Güngör T, Gürgen F, Özturan C, editors. Lecture Notes in Comput Sci. Berlin, Heidelberg: Springer; 2005:284-293.
18. Till C. Exercise as labour: quantified self and the transformation of exercise into labour. Societies 2014 Aug 28;4(3):446-462. [doi: 10.3390/soc4030446]
19. Kaye K. Adage. New York, NY: Advertising Age; 2014 May 07. FTC: fitness apps can help you shred calories - and privacy URL: http://adage.com/article/privacy-and-regulation/ftc-signals-focus-health-fitness-data-privacy/293080/ [accessed 2017-01-18] [WebCite Cache ID 6nb2dfgAM]
20. Lohr S. The New York Times. New York, NY: The New York Times Company; 2017 Jan 08. Data could be the next tech hot button for regulators URL: http://https:/www.nytimes.com/2017/01/08/technology/data-regulators-google-facebook-monopoly.html?_r=0 [accessed 2017-01-18] [WebCite Cache ID 6nb2nKTlo]
21. Rowan M, Dehlinger J. A privacy policy comparison of health and fitness related mobile applications. In: Procedia Comput Sci. 2014 Presented at: 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks; 2014; Halifax, Nova Scotia p. 348-355. [doi: 10.1016/j.procs.2014.08.051]
22. Pasquale F. The Black Box Society: The Secret Algorithms that Control Money and Information. Cambridge, MA: Harvard University Press; 2015.
23. Humer C, Finkle J. Reuters. London, UK: Thomson Reuters; 2014 Sep 24. Your medical record is worth more to hackers than your credit card URL: http://www.reuters.com/article/us-cybersecurity-hospitals-idUSKCN0HJ21I20140924 [accessed 2017-01-18] [WebCite Cache ID 6nb2tKTjS]
24. Lohr S. The New York Times. New York, NY: The New York Times Company; 2016 Nov 12. A 10-digit key code to your private life: your cellphone number URL: https://www.nytimes.com/2016/11/13/business/cellphone-number-social-security-number-10-digit-key-code-to-private-life.html [accessed 2017-06-16] [WebCite Cache ID 6nb2zQCuL]
25. Golbeck J, Mauriello M. User perception of facebook app data access: a comparison of methods and privacy concerns. Future Internet 2016 Mar 25;8(2):9. [doi: 10.3390/fi8020009]
26. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2014 Aug 21;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
27. Ashok I. International Business Times. New York, NY: IBT Media; 2016 Sep 28. Facebook to appeal against Germany's order to stop WhatsApp user data collection URL: http://www.ibtimes.co.uk/facebook-appeal-against-germanys-order-stop-whatsapp-user-data-collection-1583644 [accessed 2017-01-18] [WebCite Cache ID 6nb38kqFg]

Abbreviations

API: application programming interface

Edited by G Eysenbach; submitted 18.01.17; peer-reviewed by T Dehling, JT Prieto, G Weber; comments to author 26.03.17; revised version received 09.05.17; accepted 26.05.17; published 28.06.17.

Please cite as:
Grundy Q, Held FP, Bero LA
Tracing the Potential Flow of Consumer Data: A Network Analysis of Prominent Health and Fitness Apps
J Med Internet Res 2017;19(6):e233
URL: http://www.jmir.org/2017/6/e233/
doi:10.2196/jmir.7347
PMID:28659254

Original Paper

MyGeneFriends: A Social Network Linking Genes, Genetic Diseases, and Researchers

Alexis Allot[1], PhD; Kirsley Chennen[1], PhD; Yannis Nevers[1], MSc; Laetitia Poidevin[1], MSc; Arnaud Kress[1], MSc; Raymond Ripp[1], PhD; Julie Dawn Thompson[1], PhD; Olivier Poch[1], PhD; Odile Lecompte[1], PhD

ICUBE UMR 7357, Complex Systems and Translational Bioinformatics, Université de Strasbourg - CNRS - FMTS, Strasbourg, France

Corresponding Author:
Odile Lecompte, PhD
ICUBE UMR 7357
Complex Systems and Translational Bioinformatics
Université de Strasbourg - CNRS - FMTS
CSTB – ICUBE UMR7357
4 rue de Kirschleger
Strasbourg, 67085
France
Phone: 33 3 68 85 32 96
Fax: 33 3 68 85 35 18
Email: odile.lecompte@unistra.fr

Abstract

Background: The constant and massive increase of biological data offers unprecedented opportunities to decipher the function and evolution of genes and their roles in human diseases. However, the multiplicity of sources and flow of data mean that efficient access to useful information and knowledge production has become a major challenge. This challenge can be addressed by taking inspiration from Web 2.0 and particularly social networks, which are at the forefront of big data exploration and human-data interaction.

Objective: MyGeneFriends is a Web platform inspired by social networks, devoted to genetic disease analysis, and organized around three types of proactive agents: genes, humans, and genetic diseases. The aim of this study was to improve exploration and exploitation of biological, postgenomic era big data.

Methods: MyGeneFriends leverages conventions popularized by top social networks (Facebook, LinkedIn, etc), such as networks of friends, profile pages, friendship recommendations, affinity scores, news feeds, content recommendation, and data visualization.

Results: MyGeneFriends provides simple and intuitive interactions with data through evaluation and visualization of connections (friendships) between genes, humans, and diseases. The platform suggests new friends and publications and allows agents to follow the activity of their friends. It dynamically personalizes information depending on the user's specific interests and provides an efficient way to share information with collaborators. Furthermore, the user's behavior itself generates new information that constitutes an added value integrated in the network, which can be used to discover new connections between biological agents.

Conclusions: We have developed MyGeneFriends, a Web platform leveraging conventions from popular social networks to redefine the relationship between humans and biological big data and improve human processing of biomedical data. MyGeneFriends is available at lbgi.fr/mygenefriends.

(J Med Internet Res 2017;19(6):e212) doi:10.2196/jmir.6676

KEYWORDS

health care; social media; genetic variation; hereditary disease

Introduction

Social and Scientific Contexts

Web 2.0 and, particularly, social networks (Facebook, Google+, and LinkedIn), interconnect billions of users and manage terabytes of dynamic data flow [1]. They are at the forefront of the interactions and cooperation between humans and big data, and as such, they have established or popularized new conventions. A central concept in these innovations is the notion of an agent, representing an autonomous and active network member with various prerogatives. Notably, an agent can (1) add new information, via micro-blogging for example; (2) spread information through the network via sharing [2]; (3) evaluate information with like, dislike, or vote reactions; (4) partition information using privacy settings; or (5) annotate information with comments. Agents play an active role in the evolution of the network structure by creating nodes (agent profile pages) and bidirectional (friendship) or unidirectional (follower) links between agents. They also partition agents into groups and connect agents to unstructured information (tagging). These actions are processed by specialized tools embedded in the network to create valuable feedback in the form of filtered and personalized information such as friendship suggestions, affinity scores between people, news feeds, targeted advertisements [3], merchandise suggestions [4], or real-time world observations [5].

The field of biology is evolving and adapting at a tremendous rate in response to the widespread use of high throughput methods and the rise of personal genomics [6]. For the end user of biological data, the paradigm shift initiated by the emergence of this big data [7] has led to important changes in the research landscape [8]. To keep up with the huge volumes of data and information, users need to easily and intuitively access, communicate, and network with useful information of personal interest. Therefore, data storage platforms and workflow infrastructures must evolve to integrate Web 2.0 and social network conventions.

Bioinformatics in the Web 2.0 Era

In this context, several major bioinformatics resources have introduced tools for personalized data flow management. The Online Mendelian Inheritance in Man (OMIM) [9] resource now proposes MIMmatch [10], a service allowing users to receive email notifications when entries for their favorite genes or diseases have changed. MyNCBI [11] retains user preferences to provide customized services for NCBI databases, whereas the Uniprot [12] website has been updated to allow users to select only categories of information they are interested in, to mask large-scale publications, and to use a basket to store proteins of interest. Similar efforts toward more efficient and personalized information management are also emerging in the exploitation of the increasing publication flow. Bibsonomy [13] allows a researcher to collect and manage publications and collaborate with colleagues, whereas PubChase and ReadCube recommend new publications depending on the content of an existing library. BioTextQuest+ [14] provides an interactive exploration platform for PubMed [15] and OMIM, and facilitates knowledge extraction by document clustering and bioentity recognition. GoPubMed [16] proposes pertinent publication searches by using background knowledge in the form of ontologies (gene ontology [GO], Medical Subject Headings [MeSH], etc) that take into account the user's keywords, but also synonyms and child concepts, whereas DeepQA4PA [17] returns GO concepts associated with publications related to a specific question. After identifying a gene or list of genes of interest, GeneMania [18] and GenesLikeMe [19] identify and score related genes that may also interest the user based on ontologies, disorders, compounds, phenotypes, expression levels, domains, sequences, and other data.

Important efforts have also been devoted to contextualizing entities by connecting them to their network. For instance, FACTA+ [20], Pubtator [21], or PAML-IST [22] return publications and their links to various biological entities such as compounds, drugs, enzymes, genes, diseases, symptoms, mutations, species, and others. EuropePMC [23] adds connections to GO and experimental factor ontology (EFO), and iHOP [24] highlights the most recent publications linked to a protein. Interaction with this complex data has been facilitated by the progressive democratization of visualization techniques. For instance, Javascript libraries like BioJS [25] provide reusable components for visualization of biological information (3D structures, phylogenetic trees, proteomes, pathways, and multiple sequence alignments), contributed by users and stored in a registry. Visualization techniques facilitate understanding of information updates, clarify links between entities and groups of entities, and highlight metadata information such as data sources, confidence estimates, and so on. For example, the ExAC browser [26] provides clear visualization of variations in a gene, the Semantic Body Browser [27] shows gene expression in a human and a mouse with a heat map on a schematized body, and NetGestalt [28] introduces 1-dimensional visualization of network modules to facilitate network comparisons.

Conversely, other tools aim to extract relationships between entities. For example, Chilibot [29] searches interaction (stimulation, inhibition, etc) or parallel (studied together, coexistence, homology, etc) relationships between user-submitted genes or proteins. EvexDB [30] extracts specific events: regulatory control, coregulation, or binding to a given gene.

Finally, some bioinformatics resources have introduced specific collaborative and social components, with wiki-inspired approaches like Proteopedia [31] or WikiGene [32], collaborative sequence annotations such as WebApollo [33], or voting for medical relevance and scientific evidence of variations with GeneTalk [34]. Recent initiatives such as Coremine or MAGI [35] combine these trends. Coremine allows exploration of various biomedical concepts and connections between them, addition of private or public comments, alerts on new articles or connections, and bookmarking. MAGI combines public and private cancer genomics datasets with sharing and collaborative annotation features as well as with interactive visualizations of variants, gene expression, and protein-protein interactions.

MyGeneFriends

Building on these advances, we have developed MyGeneFriends, a Web platform inspired by social networks, to redefine and enhance the relationship between humans and biological big data. By leveraging and combining conventions and practices arising from popular social networks, it provides more intuitive interactions with biological data and simplifies access to complex information by organizing it around three agents: genetic diseases, genes, and humans. This allows MyGeneFriends to be used not only by researchers and clinicians but also by the public, including empowered patients.

We focused on human genetic diseases (closely connected to genes and human users), as they represent major clinical challenges and provide a simplified context to shed light on major common diseases. MyGeneFriends allows retrieval, management, contextualization, and annotation of information related to genes (expression, localization, and so on), genetic diseases (phenotypes, variations, and so on), and humans (interests, publications, and so on). The platform leverages user behavior and networking to personalize data visualization and the flood of information for each human user's needs, and allows project-oriented collaborations. Publication and friendship suggestions facilitate the identification of new relevant genes and diseases. Finally, we capitalize on the global social network to extract additional knowledge. MyGeneFriends was used during its development by members of our laboratory that provided continuous feedback. Additional feedback was collected from clinicians and researchers of the Medical Genetics Laboratory of Strasbourg and from colleagues from other laboratories that was particularly useful for improving the visualization of variations linked to a disease.

The aim of this paper was to introduce readers to MyGeneFriends and describe how practices from social networks can be applied to improve access to biological data.

Methods

Platform Architecture

The MyGeneFriends platform integrates multiple services (Multimedia Appendix 1) to extract and integrate large amounts of heterogeneous data. The data are stored and managed in a Postgres database, with a backup copy produced daily and stored on an external server. Elasticsearch [36] is used for powerful, complex, and fast plain text queries of publications and is synchronized daily with the MyGeneFriends database. The website is based on a stateless framework (Play framework) that includes many useful features such as error handling, build-in support for Json, WebServices, WebSockets, CoffeeScript, EBean object-relational mapper (ORM), localization, logging, and WebJars. The Play framework ensures easy horizontal scaling and scalability for increasing website traffic.

To execute local scripts and programs, a Web service has been developed using the Flask framework, which is called by MyGeneFriends using REST requests to run analysis or integration tasks. Data integration scripts are written in python, using peewee as the ORM.

Data Sources

Gene-related data including gene symbol, short description, type, and protein sequence are mainly obtained from the Ensembl [37] database. UCSC provides simple access to RefSeq [38] annotations for transcripts. To map gene identifiers between Ensembl and NCBI, we combine mappings performed by Ensembl and NCBI, together with gene symbol mapping, and extract one-to-one relationships. Gene expression data are obtained from the Human Genome Atlas microarray data [39] available in the gene expression omnibus (GEO) [40] database and validated using in-house statistical methods. In addition, relative signal intensities are calculated for heat map visualization using log signal intensities normalized in the range (0-1). Cellular localization of gene products is based on cellular component terms from GO [41]. Phylogenetic distributions for human genes and 100 eukaryotic species are retrieved from the OrthoInspector database [42] and used to categorize genes according to their evolutionary profile.

The relationships between genes and publications are defined using the gene2pubmed file from the NCBI. Publication abstracts are downloaded from Pubmed and integrated in the MyGeneFriends database. The python natural language toolkit (NLTK) [43] library is used to extract keywords from textual data linked to genes and diseases. It tokenizes the text into phrases and words, stems words in order to retrieve a canonical form, and filters words on the basis of the NCBI list of stop words (words that occur frequently in texts but are not informative) and in-house filters for word size, numbers, and special characters. Then, we take advantage of the gensim [44] library to calculate the Inverse Document Frequency (IDF) of the keywords and the TF*IDF (Term Frequency * Inverse Document Frequency). The IDF is used as a specificity score, and the TF*IDF is used to weight the relationship between a keyword and a gene or disease.

The main disease-related data are obtained from OMIM and Orphanet. In order to take into account differences in disease definitions from different data sources and propose a unified view of the current disease knowledge, an integration process was developed with two simple rules (Multimedia Appendix 2). After integration, diseases are linked to phenotypes using human phenotype ontology (HPO) [45] data files (hp.obo and phenotype_annotations.tab) containing phenotypes and phenotype-disease relationships. Variations and variation-disease relationships are extracted from the curated set provided by ClinVar [46] in the variant call format (VCF) file (limited to records with an rs# identifier). Each line is parsed and a variant entry is integrated into MyGeneFriends as a couple of genomic position and allele, allowing precise definition of the relationships between diseases and mutations. Variant effect predictor (VEP) [47] is used to link variations retrieved from ClinVar to Ensembl transcripts and to estimate their effect. The effects are then automatically classified into more general categories using the sequence ontology [48] data.

Data Flow Management

The data flow management involves the integration of data from diverse sources (databases, FTP servers, and local files) into the MyGeneFriends database. After cleaning and parsing mined data, additional analyses are automatically processed, such as keyword extraction from biological text or generation of links between variants and transcripts (mentioned previously). Then, MyGeneFriends compares remote and local data to generate news events. One or more fields from each item is used as a unique identifier. If a remote item has an identifier (one or several selected fields from an item) that is absent from the local database, it is considered to be a "new" event. If a local item has an identifier that is not present in the remote source, it is considered to be a "delete" event. If the identifier is present in both remote and local sources, the items are compared field by field to generate “update” events. Once these events are generated, the local database is synchronized with the remote source.

Finally, the way the news item is presented to the user depends on the biological context of the considered element. When an agent is linked to a publication that is not available in the MyGeneFriends database, the publication is downloaded and made accessible directly from the news panel. When a sequence is updated, a sequence alignment is generated using ClustalW [49]. When a textual information changes, such as the description of a disease, the google-diff [50] python library is used to compare both versions of the text and highlight the differences.

Data Display as Word Clouds

Word cloud representations are used in the visualization panel of an agent to display the cellular localization of the protein encoded by a gene and the phenotypes associated with a disease. Specific terms are considered as more informative and emphasized in the word cloud. The specificity of a term (cellular component in GO and phenotype in HPO) describing an agent is estimated using the information content (IC) metric [51]. The IC is defined as the negative natural logarithm of the probability of a term t:

$IC\ (t)=-\log P\ (t)$,

where P (t) is based on the frequency of the term in the considered ontology.

The specificity is then defined as the IC normalized in the range (0-1), where 0 corresponds to the minimal font size and 1 to the maximal font size during word cloud rendering.

Friendships

The MyGeneFriends network is based on friendships between agents. Human friendships are defined by users, whereas gene-gene, disease-disease, and gene-disease links are automatically built from external sources (search tool for recurring instances of neighboring genes, STRING [52]; and HPO) or inferred from the MyGeneFriends network (Error: Reference source not found). STRING global scores (higher than 0.7, corresponding to “high confidence” in STRING) are used as a metric of friendship between genes based on protein-protein interaction data. Causative genes mined by HPO from OMIM and Orphanet are exploited to link genes and diseases.

In addition to these external sources, MyGeneFriends establish links based on common properties. Diseases sharing phenotypes are related to each other with a score defined as the sum of specificity scores of phenotypes common to both diseases, divided by the sum of specificity scores of all phenotypes related to both diseases. Similarly, genes sharing GO [41] terms are connected according to two different metrics. The first metric (“GO simple”) is based on the number of shared GO terms between 2 genes, whereas the second corresponds to the functional semantic similarity (FSS) [51]. Genes and diseases related to the same variant(s) are also linked. Moreover, genes are evolutionarily linked when applicable, on the basis of the Jaccard distance calculated between in-house phylogenetic profiles produced by OrthoInspector [42].

Finally, nonhuman agents can become friends based on social connections emerging from the network itself: genes sharing human or disease friends are connected, as well as diseases with common human or gene friends.

Suggestions and Affinity Score

To suggest new gene or disease friends or new publications to a user, MyGeneFriends relies on the content of the user’s active Topic. For friendship suggestions, each nonhuman agent from MyGeneFriends (a_c) is scored relative to the user’s active Topic and the top 10 candidates are suggested as new friends. The score of an agent (gene or disease) given a Topic is the sum of scores (S) between this agent and all agents of the same type in the active Topic (a_t):

$\text{score}(a_c) = {}^{N}_{t=0}\ S(a_t, a_c)$To score genes, we use the global STRING score, whereas the score between two diseases d1 and d2 is calculated using the Information Content (IC) of the related phenotypes (P) as:

$\text{Score}(d1, d2) = (\ IC\ (P \in d1 \cap d2))/(\ IC\ (P \in d1 \cup d2))$

In addition, we provide an affinity score (a_{aff}) reflecting the proximity between an agent and the content of the user Topic and thus, the relevance of befriending this agent. It is displayed on the gene and disease profile pages when the agent can be related to the content of the Topic. The affinity score is defined as:

$a_{aff} = a_c\ /\max ac\ X\ 100$

To suggest pertinent publications, MyGeneFriends uses keywords associated with the active Topic. These keywords have been either added manually by the user or automatically inferred (see formula below). The keywords are weighted and used to query the Elasticsearch server to retrieve pertinent publications. For Elasticsearch, weights between 0 and 1 reduce the relevance of a term, and weights higher than 1 increase it. Therefore, the weight for each keyword given the content of the Topic (k_t) is defined as:

$\text{weight}(k_t) = 1 + {}^{N}_{t=0}\ ka_t + k_h$,

where ka_t is the score describing the relationship between the keyword and an agent from the Topic, and k_h is a factor applied if the user has explicitly added this keyword to the Topic.

Results

Overview of the Platform

MyGeneFriends is a new social network leveraging conventions from Web 2.0 and interconnecting three kinds of autonomous and active agents: human genes, humans, and genetic diseases. All genetic disorders including malformations, groups of phenotypes, etc are included in the network, as well as all types of human genes (coding and noncoding) in agreement with the growing evidence concerning the importance of noncoding genes in biological processes and diseases [53-55].

All agent-related data is accessible via standardized profile pages. Daily data mining and integration processes have been developed to maintain the "nonhuman" agents (more than 63,000 human genes and 14,000 genetic diseases) up to date and generate a news flow (more than 1 million news items were created in the last year) by exploiting public (Ensembl [37], NCBI, Uniprot [12], HPO [45], OMIM [10], Orphanet [56], OrthoInspector [42], etc) and in-house data resources. All data retrieved or processed by MyGeneFriends and related to genes and diseases are "public," whereas data submitted by humans are "private" (visible only by the owner) by default, unless the human decides to make it "protected" (visible by owner and selected collaborators) or "public" (visible to anyone).

The MyGeneFriends network arises from several millions of connections (called "friendships") between agents, resulting from automated dynamic data mining processes combined with human actions (Figure 1). Assessment of gene-gene, gene-disease, and disease-disease connections (nonhuman friendships) are based on automated mining of bibliographic, evolutionary, functional, phenotypic, or social data. Human friendships with genes, diseases, and other humans are defined by the user through gene targeting, definition of research interests (Topics), or user targeting (groups). Human friendships with genes or diseases can be private, protected, or public, although they are public by default to encourage networking. This data privacy management [57] is crucial to keep essential data private, while being open enough to "attract" new information and collaborators.

By exploiting human actions, MyGeneFriends can automatically (1) personalize information and visualization by highlighting and filtering pertinent data, (2) suggest new publications and friends (gene or disease), and (3) provide subnetworks for collaborations on defined research interests.

Agent Profiles

Each agent in MyGeneFriends has a profile (Figure 2) that provides a unified architecture and organization to ensure intuitive navigation through the network and access to relevant and personalized information about agents. These profiles contain 4 major sections: "header," "basic information," "friends," and "news."

Figure 1. Ontology of friendships between agents in MyGeneFriends. Agents are linked by numerous friendships (corresponding to green boxes) of different kinds (blue boxes). First, we separate decision-driven friendships (agent actions) from naturally occurring friendships (mined). Then, we split natural friendships into those due to direct contact between agents, and those influenced by an external factor. This external factor mimics the human tendency of befriending people with the same interests (represented here as phenotypes, annotations, variants, and phylogenetic distributions) or common friends (genes, humans, and diseases).

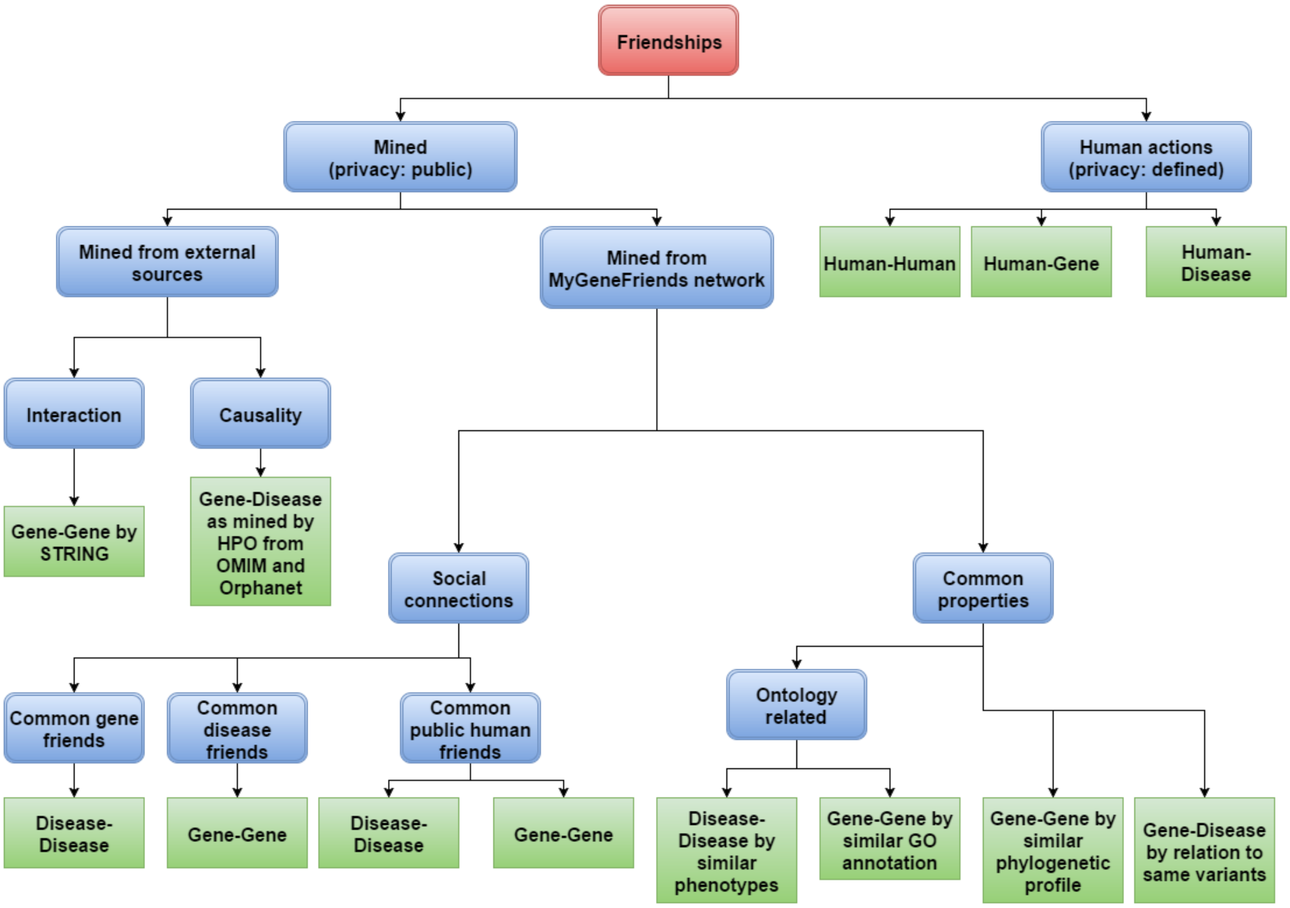
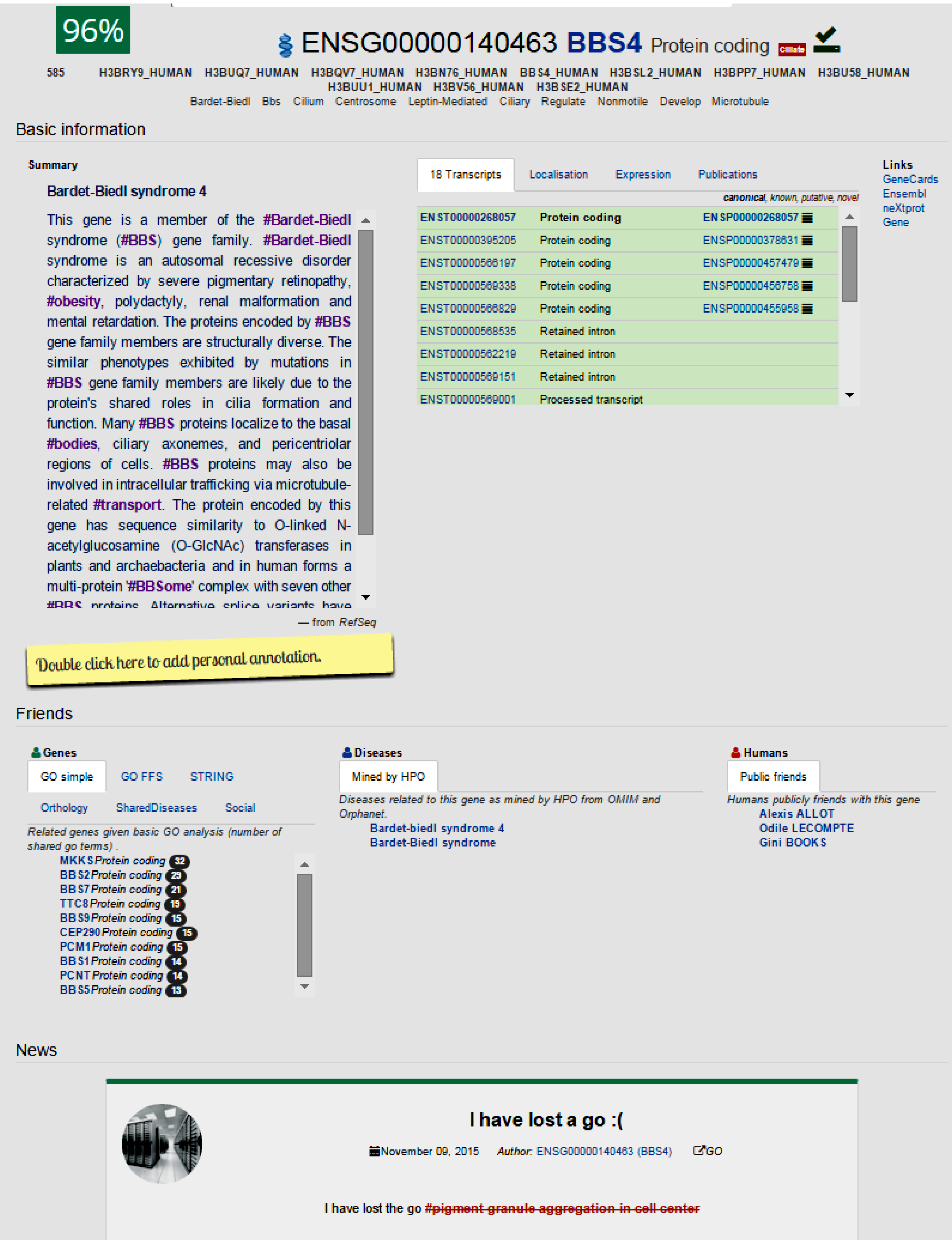


Figure 2. Representative profile page of a MyGeneFriends agent (here the gene BBS4). Four distinct sections are shown. The “header” section briefly introduces the agent, displays a list of synonyms, and allows friendship management. It shows the affinity score (here 96%) estimating how interesting this agent could be for the user. The “basic Information” section shows more detailed information about the agent: a description, different visualizations describing the agent, links to external sources, and a personal annotation from the user. The “friends” section allows navigation through the “friends of friends” network by displaying public friends of the agent, grouped according to their type. Finally, the “news” section displays all the news related to the agent.



ENSG00000140463 **BBS4** Protein coding

585 H3BRY9_HUMAN H3BUQ7_HUMAN H3BQV7_HUMAN H3BN76_HUMAN BBS4_HUMAN H3BSL2_HUMAN H3BPP7_HUMAN H3BU58_HUMAN H3BUU1_HUMAN H3BV56_HUMAN H3BSE2_HUMAN

Bardet-Biedl Bbs Cilium Centrosome Leptin-Mediated Ciliary Regulate Nonmotile Develop Microtubule

Basic information

Summary

Bardet-Biedl syndrome 4

This gene is a member of the #Bardet-Biedl syndrome (#BBS) gene family. #Bardet-Biedl syndrome is an autosomal recessive disorder characterized by severe pigmentary retinopathy, #obesity, polydactyly, renal malformation and mental retardation. The proteins encoded by #BBS gene family members are structurally diverse. The similar phenotypes exhibited by mutations in #BBS gene family members are likely due to the protein's shared roles in cilia formation and function. Many #BBS proteins localize to the basal #bodies, ciliary axonemes, and pericentriolar regions of cells. #BBS proteins may also be involved in intracellular trafficking via microtubule-related #transport. The protein encoded by this gene has sequence similarity to O-linked N-acetylglucosamine (O-GlcNAc) transferases in plants and archaebacteria and in human forms a multi-protein '#BBSome' complex with seven other #BBS proteins. Alternative splice variants have

— from *RefSeq*

Double click here to add personal annotation.

canonical, *known*, *putative*, *novel*

ENST00000268057	**Protein coding**	ENSP00000268057
ENST00000395205	Protein coding	ENSP00000378631
ENST00000566197	Protein coding	ENSP00000457479
ENST00000569338	Protein coding	ENSP00000456758
ENST00000566829	Protein coding	ENSP00000455958
ENST00000568535	Retained intron	
ENST00000562219	Retained intron	
ENST00000569151	Retained intron	
ENST00000569001	Processed transcript	

Links
GeneCards
Ensembl
neXtprot
Gene

Friends

Genes

Related genes given basic GO analysis (number of shared go terms) .

MKKS	*Protein coding*	32
BBS2	*Protein coding*	23
BBS7	*Protein coding*	21
TTC8	*Protein coding*	19
BBS9	*Protein coding*	15
CEP290	*Protein coding*	15
PCM1	*Protein coding*	15
BBS1	*Protein coding*	14
PCNT	*Protein coding*	14
BBS5	*Protein coding*	13

Diseases

Diseases related to this gene as mined by HPO from OMIM and Orphanet.

- Bardet-biedl syndrome 4
- Bardet-Biedl syndrome

Humans

Humans publicly friends with this gene

- Alexis ALLOT
- Odile LECOMPTE
- Gini BOOKS

News

I have lost a go :(

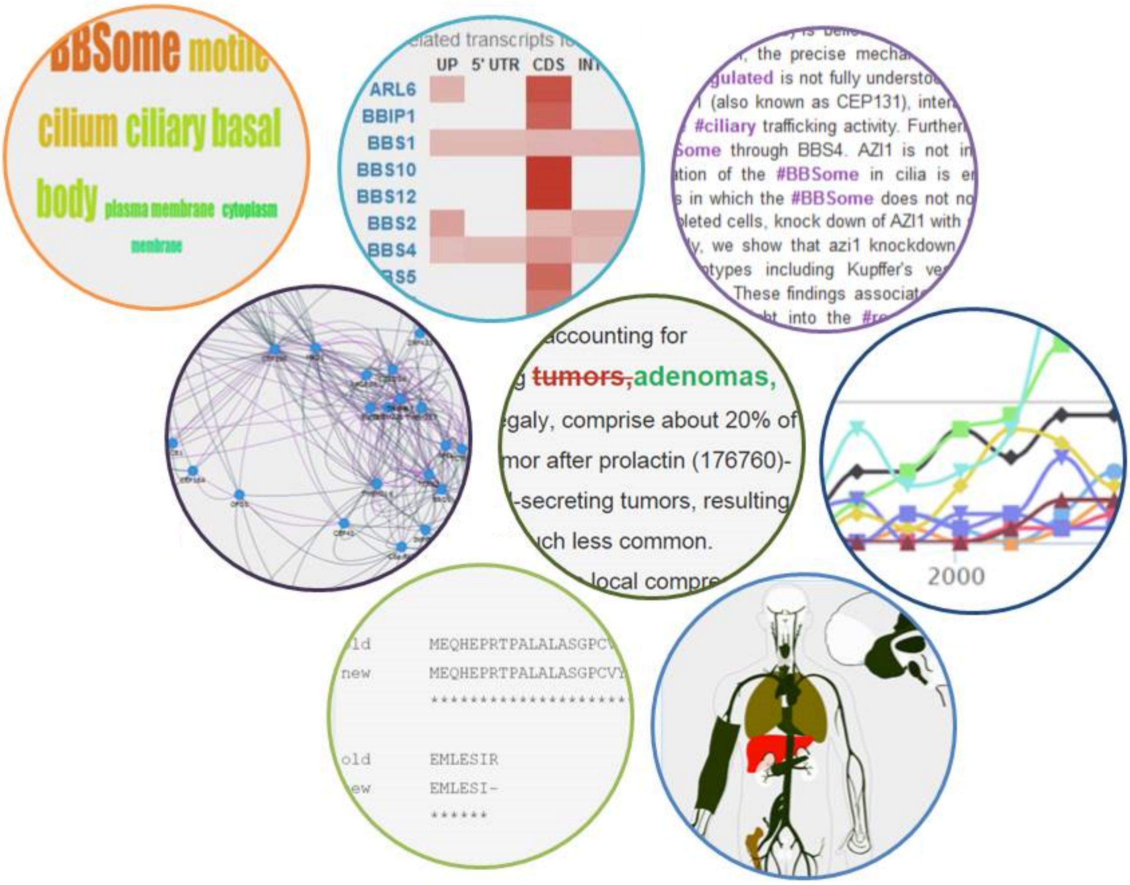
November 09, 2015 *Author:* ENSG00000140463 (BBS4) GO

I have lost the go ~~#pigment granule aggregation in cell center~~

The top agent-related keywords are displayed in the header to briefly introduce the agent, whereas the summary in the “basic information” section provides a more detailed description. Humans can expand the official description of a nonhuman agent by adding personal annotations or unpublished results that can then be accessed at any time and shared with collaborators (Figure 2). Exploration panels give access to the most important information using visualization techniques (see Figure 3) to highlight specific information for genes and diseases as described below. The “friends” section of the profile displays links to public friends (genes, diseases, or humans) of the agent, allowing further networking with potentially interesting agents. Finally, the news feed is an intuitive way to track changes in information related to an agent.

To personalize the profile view, the keywords inferred to be important for the viewer are highlighted in the agent description. For example, if the user is friends with cilia-related genes, the word cilium is highlighted in the description of the other agents (human, gene, or disease). Moreover, if a nonhuman agent is related to the user’s current collaborators, an affinity score is shown, inviting the user to befriend this agent. If the agent already collaborates with the user, the score reflects how close it is to other collaborators.

Figure 3. MyGeneFriends uses various visualization techniques to optimize the display of biological information: (1) word clouds highlight the most specific ontology terms, (2) barcodes offer a synaptic and interactive view of the density of variations related to regions of a gene or effect on a protein, (3) highlighting words in text identifies the most pertinent paragraphs for a given human user, (4) networks of friends help to understand the connections between agents and identify groups of highly connected agents, (5) colors highlight modifications in textual information related to agents, (6) timelines show the evolution of the popularity of gene collaborators in a Topic, (7) pairwise alignments identify the differences between two versions of protein sequences, and (8) heat maps on schemas of the human body, brain, and fetus allow easy analysis of the expression pattern of a given gene.



Gene Profile

Gene profiles use RefSeq summaries to describe agents and connect these agents to external resources via links to Ensembl, GeneCards, NCBI, and neXtprot websites.

Exploration panels display the most important aspects of the gene. The first panel presents gene transcripts with their properties: sequence, type (protein coding, miRNA, etc), reliability (known, putative, and novel), and corresponding protein sequence, if any. The second panel shows the subcellular localization(s) of the encoded proteins, defined by the GO cellular component ontology, as a word cloud (Figure 3). The third panel shows the gene expression for protein coding genes as a heat map in more than 40 tissues, through an interactive schematic view of the human body (male and female), brain, and foetus (Figure 3). Pan and zoom capabilities (jquery.panzoom.js) allow users to navigate through the schematic view and visualize even the smallest tissues. Additional information such as tissue description and probe set signal intensities are available. In addition to the visualization of gene expression, the “Expression filter” tool allows users to find genes of interest based on their expression or absence of expression in a defined set of tissues. Publications associated with the gene are displayed with their abstract and can be liked, disliked, or marked as valuable. The number of all genes related to the publication, as well as the count of likes and dislikes, help to estimate the relevance of the publication for the considered gene. Moreover, genes related to a publication can be visualized as an interactive graph, allowing further networking and identification of additional genes of interest.

Genetic Disease Profile

Diseases are extracted from the OMIM (all entries except explicit genes) and Orphanet (all entries, including groups of phenotypes) databases. The preference of exhaustivity over specificity is motivated by the inherent difficulty in defining a disease. We use expert created links between Orphanet and OMIM entries (displayed on Orphanet entries) as the main data source to merge diseases. When a disease is not linked to any other, or when a clear one-to-one mapping can be made between an Orphanet and an OMIM entry, the entries from both databases are fused into a single one (see Multimedia Appendix 2). Once this process is complete, we use the remaining one-to-many connections (eg, one entry for “Bardet-Biedl syndrome” in Orphanet corresponds to multiple entries in OMIM for each Bardet Biedl syndrome subtype) to create groups of highly connected diseases, which we call “metadiseases.”

Two main features have been selected to characterize a disease on the disease profile panel: (1) variations explaining the causes of a disease, and (2) phenotypes describing its consequences. Phenotypes are represented by a word cloud highlighting rare HPO phenotypes associated with the disease. The description of variants is generated by the integration of more than 100,000 ClinVar [46] curated variations (single-nucleotide variants and small insertions and deletions) directly linked to diseases.

As the effects of the variants can differ per considered transcript, MyGeneFriends uses the Ensembl VEP [47] script to create more than a million links between variants and Ensembl transcripts stored in the MyGeneFriends database. To describe the complex relationships between variants, transcripts, and disease-causing genes, we have developed three synoptic and interactive views with variants grouped per affected gene. With this synthetic barcode representation (Figure 3), the human user has a rapid overview of the characteristics of the known variants associated with the disease and can easily identify variants exhibiting specific features, for instance, synonymous variants affecting a splicing region. The third view focuses on variants differentially affecting protein coding transcripts (Multimedia Appendix 3). Such variants can generate a mix of affected and unaffected proteins depending on the tissue or developmental stage and often result in puzzling phenotypes.

Metadiseases have special profile pages on MyGeneFriends, summarizing the main properties of nested diseases, displaying nested diseases as a network, and highlighting the most representative gene friends and phenotypes of the concerned diseases. To date, MyGeneFriends has information on 725 metadiseases, representing 3418 diseases.

Human Profile

The third agent in MyGeneFriends is the human user, who must register on the website (registration is free, and a demo account is available for testing purposes). The user’s profile page contains information provided by the owner: his affiliation, geographic localization, a list of authored publications, and a short description. Even if no description is provided, MyGeneFriends introduces the human to other users by automatically extracting best scored keywords associated with public gene and disease friends of the human and displaying them on his profile.

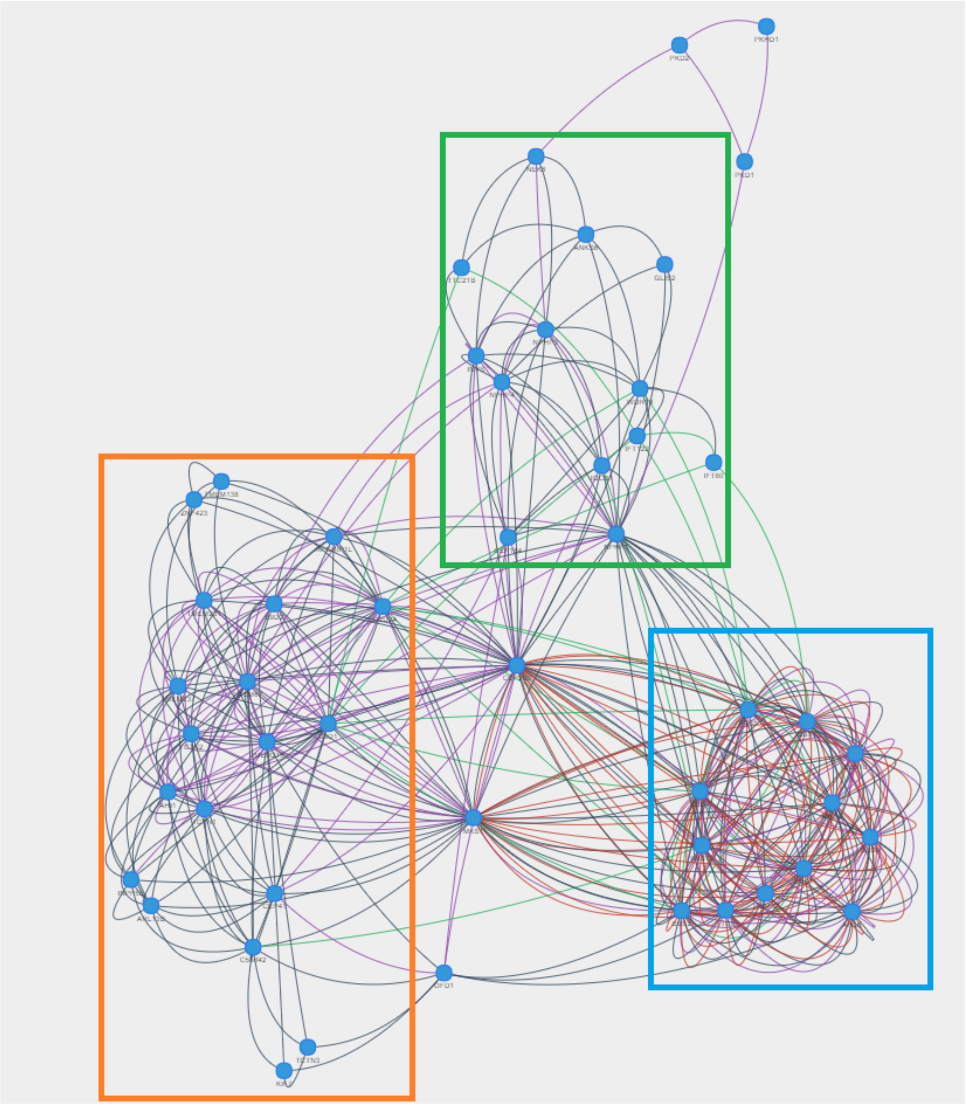
The private view of the profile page allows humans to create and manage groups of collaborators related to research projects, called Topics. All Topics owned by the user are shown in the “My Topics” section. The Topic selected as “active” is used for personalization and suggestion processes. A second section called “My collaborations” allows the user to monitor Topics from the other users with whom he collaborates.

Friendships and Networking

Friendships are an essential concept in MyGeneFriends, since on the one hand, they allow networking through friends and evaluation of the relatedness of 2 agents, and on the other hand, they are used to suggest interesting agents as new friends. Some friendships are automatically generated based on data mining, whereas others result from human activity. Friendships offer different and complementary points of view on the close environment of an agent in terms of protein interactions, function and localization, implication in research projects or diseases, and many others (Figure 1).

Exploitation of the friendship network in MyGeneFriends leverages mined and user-created connections to discover highly connected clusters. Interactive graph views with repulsion physics (using the vis.js library) allow intuitive visualization of friendships within a group of agents (genes from a publication, diseases from a metadisease, or agents associated with a Topic), leading to selection and observation of different types of friendships (common friends, common features, cooccurrence, and so on). Highly connected agents will naturally form subgroups corresponding to biologically relevant categories as exemplified by the Congenital Hepatic Fibrosis [58] gene network (Figure 4).

Figure 4. Dynamic network visualization of relationships between actors. Network of 52 genes related to Congenital Hepatic Fibrosis (CHF), a developmental disorder most frequently associated with ciliopathies. Red links represent shared public human friends, grey links represent shared diseases, violet links represent STRING relationships, and green links represent similar evolutionary profiles. Each link type can be removed or added to the network in real time. Moreover, in the dynamic network view provided by MyGeneFriends, highly connected genes are clustered automatically to form subgroups. In this example, 3 main subnetworks (highlighted by rectangles) emerge corresponding to genes associated with a distinct ciliopathy: Bardet-Biedl Syndrome (blue rectangle), Joubert and Meckel syndromes (orange rectangle), and Senior-Loken syndrome and nephronophthisis (green rectangle).



Topic: Interactive Collaborative Unit

On their profile pages, users can create groups of agents (called Topics). Each group centralizes information around a research project and links to agents collaborating with it (Multimedia Appendix 4), thus presenting a subjective view of biological information from a given research perspective.

This allows MyGeneFriends to display a personalized news feed, providing a technological watch of bibliographical and public database updates related to gene and disease friends that collaborate in the user's Topics. News items include various subjects such as new or lost friendships between diseases and genes, updated symbols, synonyms, descriptions, new or lost GO or HPO annotations, protein sequence updates, or presence in a new publication.

Several tools are provided for the analysis of Topic related agents. The network visualization facilitates the evaluation of the heterogeneity of the Topic's content (Multimedia Appendix 5), and the identification of highly linked subgroups of agents and relationships between these groups. The timeline visualization (Figure 3) places the Topic in a global research perspective, presenting the annual evolution of the number of publications associated with the genes in a Topic.

Finally, in addition to serving as a basis for friendships and publications suggestions (see Methods), information mined from Topics allows the enhancement of the reading experience of an agent's descriptions and publication abstracts by automatically highlighting the keywords most representative of the user's interests (Multimedia Appendix 6).

Discussion

Principal Findings

By leveraging conventions and practices used in popular social networks, MyGeneFriends aims to challenge the way we interact with data by providing a first step toward a system where biological entities such as genes and genetic diseases are no longer passive concepts, but are instead proactive agents of the research process, helping and collaborating with human counterparts.

In mainstream social networks, humans can create a representation of themselves in the form of a profile, then interact with the network by writing posts, adding commentaries or likes, making new friends, and sharing and spreading information. To transpose this concept to MyGeneFriends, we had to create a network that could reflect current research efforts in genetics and medicine. To populate the network, we focused on human genetic diseases because of their broad interest, and their more direct links to genes and genomic variations compared with infectious diseases or cancer. With humans and human genetic diseases selected, the choice of the third agent was obvious as many publications and bioinformatics resources structure their information in a gene-centric manner. To interconnect the network, we adopted two of the main characteristics of real-world friendships: commonality (common friends, qualities, and interactions), and group membership (family, coworkers, and hobbies).

Compared with existing Web services, MyGeneFriends can (1) leverage user behavior to provide personalized profiles and news feeds, given each user’s specific research interests; and (2) consider user behavior as valid biological information integrated in the biological data network to be mined and influencing the discovery of connections between genes and diseases.

Conclusions

The development of MyGeneFriends lies at the frontier between bioinformatics and the emerging science of human-data interaction, and in the future, we plan to extend the functionalities in both areas. First, genes from other model species (mouse, zebrafish, etc) will be added and connected by friendship links based on orthology. Additional friendships will be incorporated to provide a regulatory context such as friendships based on transcription factors or miRNA. Second, we believe that while humans remain special agents in this first version of MyGeneFriends, in the future the three agents will interact on the same level, with more independent and proactive genes and diseases. Research will be facilitated by better communication between different agents, with each agent able to produce and transmit new, relevant data and knowledge. A gene could, for example, find itself linked to a new disease or ask to be sequenced by his friend, the sequencer. With this increased autonomy of nonhuman agents and an independent flow of information, the role of the human in the network must clearly evolve. This evolution can be viewed either as a danger or as a source of new collaborations and opportunities.

Acknowledgments

This work was funded by ANR Investissement d’Avenir Bioinformatique BIP: BIP (ANR10-BINF03-05).

The authors are thankful to Alexia Rohmer for her help with the interactive schematic view of the human body, and are also grateful to Dr. Ioannis Xenarios (Swiss Institute of Bioinformatics), Dr. Frédéric Chalmel (Université de Rennes 1), Dr. Laurent Vallar (Luxembourg Institute of Health), and Prof. Pierre Gançarski (University of Strasbourg) for helpful discussions.

Authors' Contributions

AA developed MyGeneFriends with contributions from KC and YN, and wrote the manuscript. AK and RR managed hardware infrastructure. JL wrote the manuscript. OP and OL supervised the project and wrote the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MyGeneFriends relies on multiple servers, scripts, and databases. Jenkins is used to periodically execute python integration scripts, which maintain the MyGeneFriends database up to date. A backup copy of the database is created daily, and the publications table is synchronized daily with an Elasticsearch server. The MyGeneFriends website is built using the Play framework and calls an API server built with Flask to execute command line programs and download publications into the local database. Bug reports are sent to the YouTrack server.

[PPTX File, 103KB - jmir_v19i6e212_app1.pptx]

Multimedia Appendix 2

Diseases from different, contradictory data sources are merged using two rules. First, disease A is merged with disease B if and only if B has a single link to disease A, and no other diseases have a single link to A. Second, if a disease is linked to several other diseases, a disease group called Metadisease is created.

[PPTX File, 274KB - jmir_v19i6e212_app2.pptx]

Multimedia Appendix 3

The “differential effect on proteins” view for variants on the “Bardet-biedl syndrome 6” disease profile (a) shows that while variants (here rs74315398, rs28937875, rs587777827, rs74315399, rs74315397, rs74315396) linked to this disease affect the coding DNA sequence (CDS) of 2 transcripts of the gene MKKS, one protein coding transcript (green rectangle on Ensembl Genome Browser [b], and 1 of 3 labels in the view [a]) is not affected in the CDS.

[PPTX File, 71KB - jmir_v19i6e212_app3.pptx]

Multimedia Appendix 4

A Topic groups genes, diseases, and humans collaborating on a shared research interest. When a human becomes friends with a new gene or disease, it is added to the active Topic. Human collaborators see all protected friends and annotations related to the Topic.

[PPTX File, 113KB - jmir_v19i6e212_app4.pptx]

Multimedia Appendix 5

Network visualization of agents related to a Topic (here a set of Bardet-Biedl syndrome related genes on the left and a set of muscular fiber related genes on the right) help to understand how many potentially different research interests a Topic contains. The network displays highly connected agents automatically grouped together. Purple links represent STRING based relationships, green links are based on evolutionary profile similarity, red links indicate shared public human friends, and grey links shared disease friends.

[PPTX File, 61KB - jmir_v19i6e212_app5.pptx]

Multimedia Appendix 6

Keywords most related to agents from active Topics are highlighted in (1) publication abstracts, and (2) descriptions on agent profiles. This helps to quickly identify paragraphs that may interest the reader.

[PPTX File, 146KB - jmir_v19i6e212_app6.pptx]

References

1. Allot A, Lecompte O. LBGI. MyGeneFriends website URL: http://lbgi.fr/mygenefriends [accessed 2017-06-01] [WebCite Cache ID 6qtiI2y1c]
2. Bakshy E, Rosenn I, Marlow C, Adamic L. The role of social networks in information diffusion. 2012 Apr 16 Presented at: Proceedings of the 21st International Conference on World Wide Web; 2012; Lyon.
3. King SP, Burgener C, Paretti CT, Davis ME. Google Patents. 2010. System and method for contextual advertising based on status messages URL: https://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US20100228582.pdf [WebCite Cache ID 6kdH8PZdj]
4. Ma H, Zhou D, Liu C, Lyu M, King I. Recommender systems with social regularization. 2011 Presented at: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining; 2011; Hong Kong p. 287-296. [doi: 10.1145/1935826.1935877]
5. Zhao S, Zhong L, Wickramasuriya J. Human as real-time sensors of social and physical events: a case study of twitter and sports games. arXiv 2011:1106-4300.
6. Hendlisz A. Of art and science: is personalized medicine getting personal enough? Curr Opin Oncol 2015 Jul;27(4):349-350. [doi: 10.1097/CCO.0000000000000205] [Medline: 26049276]
7. Hey T. The fourth paradigm – data-intensive scientific discovery. In: Kurbanoğlu S, Al U, Erdoğan PL, Tonta Y, Uçak N, editors. E-Science and Information Management. IMCW 2012. Communications in Computer and Information Science. Berlin, Heidelberg: Springer; 2012.
8. Kitchin R. Big Data, new epistemologies and paradigm shifts. Big Data Soc 2014 Jul 10;1(1):1-12. [doi: 10.1177/2053951714528481]

9. Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online mendelian inheritance in man (OMIM). Hum Mutat 2000;15(1):57-61. [doi: 10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G] [Medline: 10612823]
10. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online mendelian inheritance in man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res 2015 Jan;43(Database issue):D789-D798 [FREE Full text] [doi: 10.1093/nar/gku1205] [Medline: 25428349]
11. Giglia E. PubMed in progress: latest changes in MeSH and MyNCBI. Eur J Phys Rehabil Med 2011 Sep;47(3):525-528 [FREE Full text] [Medline: 21946409]
12. UniProt C. UniProt: a hub for protein information. Nucleic Acids Res 2015 Jan;43(Database issue):D204-D212 [FREE Full text] [doi: 10.1093/nar/gku989] [Medline: 25348405]
13. Benz D, Eisterlehner F, Hotho A, Jaschke R, Krause B, Stumme G. Managing publications and bookmarks with BibSonomy. 2009 Presented at: 20th Acm Conference on Hypertext and Hypermedia (Hypertext 2009); 2009; Torino p. 323-324. [doi: 10.1145/1557914.1557969]
14. Papanikolaou N, Pavlopoulos GA, Pafilis E, Theodosiou T, Schneider R, Satagopam VP, et al. BioTextQuest(+): a knowledge integration platform for literature mining and concept discovery. Bioinformatics 2014 Nov 15;30(22):3249-3256 [FREE Full text] [doi: 10.1093/bioinformatics/btu524] [Medline: 25100685]
15. Giglia E, Spinelli O. PubMed reloaded: new interface, enhanced discovery. Eur J Phys Rehabil Med 2009 Dec;45(4):631-636 [FREE Full text] [Medline: 20032922]
16. Doms A, Schroeder M. GoPubMed: exploring PubMed with the gene ontology. Nucleic Acids Res 2005 Jul 1;33(Web Server issue):W783-W786 [FREE Full text] [doi: 10.1093/nar/gki470] [Medline: 15980585]
17. Gobeill J, Gaudinat A, Pasche E, Vishnyakova D, Gaudet P, Bairoch A, et al. Deep question answering for protein annotation. Database (Oxford) 2015;2015:bav081 [FREE Full text] [doi: 10.1093/database/bav081] [Medline: 26384372]
18. Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD, et al. GeneMANIA prediction server 2013 update. Nucleic Acids Res 2013 Jul;41(Web Server issue):W115-W122 [FREE Full text] [doi: 10.1093/nar/gkt533] [Medline: 23794635]
19. Stelzer G, Inger A, Olender T, Iny-Stein T, Dalah I, Harel A, et al. GeneDecks: paralog hunting and gene-set distillation with GeneCards annotation. OMICS 2009 Dec;13(6):477-487. [doi: 10.1089/omi.2009.0069] [Medline: 20001862]
20. Tsuruoka Y, Tsujii J, Ananiadou S. FACTA: a text search engine for finding associated biomedical concepts. Bioinformatics 2008 Nov 1;24(21):2559-2560 [FREE Full text] [doi: 10.1093/bioinformatics/btn469] [Medline: 18772154]
21. Wei C, Kao H, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. Nucleic Acids Res 2013 Jul;41(Web Server issue):W518-W522 [FREE Full text] [doi: 10.1093/nar/gkt441] [Medline: 23703206]
22. Mandloi S, Chakrabarti S. PALM-IST: pathway assembly from literature mining--an information search tool. Sci Rep 2015;5:10021 [FREE Full text] [doi: 10.1038/srep10021] [Medline: 25989388]
23. Europe PMC Consortium. Europe PMC: a full-text literature database for the life sciences and platform for innovation. Nucleic Acids Res 2015 Jan;43(Database issue):D1042-D1048 [FREE Full text] [doi: 10.1093/nar/gku1061] [Medline: 25378340]
24. Hoffmann R, Valencia A. A gene network for navigating the literature. Nat Genet 2004 Jul;36(7):664. [doi: 10.1038/ng0704-664] [Medline: 15226743]
25. Yachdav G, Goldberg T, Wilzbach S, Dao D, Shih I, Choudhary S, et al. Anatomy of BioJS, an open source community for the life sciences. Elife 2015;4:e07009 [FREE Full text] [doi: 10.7554/eLife.07009] [Medline: 26153621]
26. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016 Aug 18;536(7616):285-291. [doi: 10.1038/nature19057] [Medline: 27535533]
27. Lekschas F, Stachelscheid H, Seltmann S, Kurtz A. Semantic Body Browser: graphical exploration of an organism and spatially resolved expression data visualization. Bioinformatics 2015 Mar 1;31(5):794-796 [FREE Full text] [doi: 10.1093/bioinformatics/btu707] [Medline: 25344497]
28. Shi Z, Wang J, Zhang B. NetGestalt: integrating multidimensional omics data over biological networks. Nat Methods 2013 Jul;10(7):597-598 [FREE Full text] [doi: 10.1038/nmeth.2517] [Medline: 23807191]
29. Chen H, Sharp BM. Content-rich biological network constructed by mining PubMed abstracts. BMC Bioinformatics 2004 Oct 8;5:147 [FREE Full text] [doi: 10.1186/1471-2105-5-147] [Medline: 15473905]
30. Van Landeghem LS, Björne J, Wei C, Hakala K, Pyysalo S, Ananiadou S, et al. Large-scale event extraction from literature with multi-level gene normalization. PLoS One 2013;8(4):e55814 [FREE Full text] [doi: 10.1371/journal.pone.0055814] [Medline: 23613707]
31. Hodis E, Prilusky J, Sussman JL. Proteopedia: A collaborative, virtual 3D web-resource for protein and biomolecule structure and function. Biochem Mol Biol Educ 2010 Sep;38(5):341-342 [FREE Full text] [doi: 10.1002/bmb.20431] [Medline: 21567857]
32. Hoffmann R. A wiki for the life sciences where authorship matters. Nat Genet 2008 Sep;40(9):1047-1051. [doi: 10.1038/ng.f.217] [Medline: 18728691]
33. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, et al. Web Apollo: a web-based genomic annotation editing platform. Genome Biol 2013;14(8):R93 [FREE Full text] [doi: 10.1186/gb-2013-14-8-r93] [Medline: 24000942]

34. Kamphans T, Krawitz PM. GeneTalk: an expert exchange platform for assessing rare sequence variants in personal genomes. Bioinformatics 2012 Oct 1;28(19):2515-2516 [FREE Full text] [doi: 10.1093/bioinformatics/bts462] [Medline: 22826540]
35. Leiserson MD, Gramazio CC, Hu J, Wu H, Laidlaw DH, Raphael BJ. MAGI: visualization and collaborative annotation of genomic aberrations. Nat Methods 2015 Jun;12(6):483-484. [doi: 10.1038/nmeth.3412] [Medline: 26020500]
36. Gormley C, Tong Z. Elasticsearch: The Definitive Guide. California: O'Reilly Media; 2015.
37. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Searle Stephen M J, et al. Ensembl 2015. Nucleic Acids Res 2015 Jan;43(Database issue):D662-D669 [FREE Full text] [doi: 10.1093/nar/gku1010] [Medline: 25352552]
38. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res 2014 Jan;42(Database issue):D756-D763 [FREE Full text] [doi: 10.1093/nar/gkt1114] [Medline: 24259432]
39. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 2004 Apr 20;101(16):6062-6067 [FREE Full text] [doi: 10.1073/pnas.0400782101] [Medline: 15075390]
40. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013 Jan;41(Database issue):D991-D995 [FREE Full text] [doi: 10.1093/nar/gks1193] [Medline: 23193258]
41. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res 2015 Jan;43(Database issue):D1049-D1056 [FREE Full text] [doi: 10.1093/nar/gku1179] [Medline: 25428369]
42. Linard B, Allot A, Schneider R, Morel C, Ripp R, Bigler M, et al. OrthoInspector 2.0: software and database updates. Bioinformatics 2015 Feb 1;31(3):447-448 [FREE Full text] [doi: 10.1093/bioinformatics/btu642] [Medline: 25273105]
43. Bird S. NLTK. 2006 Presented at: Proceedings of the COLING/ACL on Interactive presentation sessions; 2006; Sydney. [doi: 10.3115/1225403.1225421]
44. Rehurek R, Sojka. Software framework for topic modelling with large corpora. 2010 Presented at: LREC 2010 workshop New Challenges for NLP Frameworks; 2010; Valetta p. 46-50.
45. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The human phenotype ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 2014 Jan;42(Database issue):D966-D974 [FREE Full text] [doi: 10.1093/nar/gkt1026] [Medline: 24217912]
46. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014 Jan;42(Database issue):D980-D985 [FREE Full text] [doi: 10.1093/nar/gkt1113] [Medline: 24234437]
47. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the ensembl API and SNP effect predictor. Bioinformatics 2010 Aug 15;26(16):2069-2070 [FREE Full text] [doi: 10.1093/bioinformatics/btq330] [Medline: 20562413]
48. Cunningham F, Moore B, Ruiz-Schultz N, Ritchie GR, Eilbeck K. Improving the sequence ontology terminology for genomic variant annotation. J Biomed Semantics 2015;6:32 [FREE Full text] [doi: 10.1186/s13326-015-0030-4] [Medline: 26229585]
49. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994 Nov 11;22(22):4673-4680. [Medline: 7984417]
50. Fraser N. Google. 2012. google-diff-match-patch URL: https://code.google.com/p/google-diff-match-patch/ [accessed 2016-09-19] [WebCite Cache ID 6kdGfLcD8]
51. Reyes-Palomares A, Rodríguez-López R, Ranea JA, Sánchez-Jiménez F, Sánchez JF, Medina MA. Global analysis of the human pathophenotypic similarity gene network merges disease module components. PLoS One 2013;8(2):e56653 [FREE Full text] [doi: 10.1371/journal.pone.0056653] [Medline: 23437198]
52. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015 Jan;43(Database issue):D447-D452 [FREE Full text] [doi: 10.1093/nar/gku1003] [Medline: 25352553]
53. Khorkova O, Hsiao J, Wahlestedt C. Basic biology and therapeutic implications of lncRNA. Adv Drug Deliv Rev 2015 Jun 29;87:15-24 [FREE Full text] [doi: 10.1016/j.addr.2015.05.012] [Medline: 26024979]
54. Luo H, Sun Y, Wei G, Luo J, Yang X, Liu W, et al. Functional characterization of long noncoding RNA Lnc_bc060912 in human lung carcinoma cells. Biochemistry 2015 May 12;54(18):2895-2902. [doi: 10.1021/acs.biochem.5b00259] [Medline: 25848691]
55. Sun J, Ding W, Zhi J, Chen W. MiR-200 suppresses metastases of colorectal cancer through ZEB1. Tumour Biol 2015:-Epub ahead of print. [doi: 10.1007/s13277-015-3822-3] [Medline: 26242262]
56. Maiella S, Rath A, Angin C, Mousson F, Kremp O. [Orphanet and its consortium: where to find expert-validated information on rare diseases]. Rev Neurol (Paris) 2013 Feb;169(Suppl 1):S3-S8. [doi: 10.1016/S0035-3787(13)70052-3] [Medline: 23452769]
57. Jee K, Kim G. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc Inform Res 2013 Jun;19(2):79-85 [FREE Full text] [doi: 10.4258/hir.2013.19.2.79] [Medline: 23882412]

58. Gunay-Aygun M, Gahl W, Heller T. Congenital hepatic fibrosis overview. In: GeneReviews. Seattle: University of Washington, Seattle; 2014.

Abbreviations

CDS: coding DNA sequence
EFO: experimental factor ontology
FSS: functional semantic similarity
FTP: file transfer protocol
GEO: gene expression omnibus
GO: gene ontology
HPO: human phenotype ontology
IC: information content
IDF: inverse document frequency
MeSH: MEdical Subject Headings
NCBI: National Center for Biotechnology Information
NLTK: natural language toolkit
OMIM: Online Mendelian Inheritance in Man
ORM: object-relational mapping
PDF: portable document format
REST: representational state transfer
STRING: search tool for recurring instances of neighboring genes
TF: term frequency
UCSC: University of California, Santa Cruz
UTR: untranslated region
VCF: variant call format
VEP: variant effect predictor

Edited by G Eysenbach; submitted 05.10.16; peer-reviewed by M Mazzucato, J Wang; comments to author 17.11.16; revised version received 21.12.16; accepted 04.03.17; published 16.06.17.

Please cite as:
Allot A, Chennen K, Nevers Y, Poidevin L, Kress A, Ripp R, Thompson JD, Poch O, Lecompte O
MyGeneFriends: A Social Network Linking Genes, Genetic Diseases, and Researchers
J Med Internet Res 2017;19(6):e212
URL: http://www.jmir.org/2017/6/e212/
doi:10.2196/jmir.6676
PMID:28623182

Original Paper

What Motivates Young Adults to Talk About Physical Activity on Social Network Sites?

Ni Zhang[1], MA, MPH, PhD; Shelly Campo[2], PhD; Jingzhen Yang[3], MPH, PhD; Petya Eckler[4], PhD; Linda Snetselaar[2], PhD; Kathleen Janz[2], EdD; Emily Leary[5], PhD

[1]San Jose State University, San Jose, CA, United States
[2]The University of Iowa, Iowa City, IA, United States
[3]Ohio State University, Columbus, OH, United States
[4]University of Strathclyde, Glasgow, United Kingdom
[5]University of Missouri, Columbia, MO, United States

Corresponding Author:
Ni Zhang, MA, MPH, PhD
San Jose State University
San Jose State University
One Washington Square, SPX204
San Jose, CA, 95192
United States
Phone: 1 408 924 8109
Fax: 1 408 924 2979
Email: ni.zhang01@sjsu.edu

Abstract

Background: Electronic word-of-mouth on social network sites has been used successfully in marketing. In social marketing, electronic word-of-mouth about products as health behaviors has the potential to be more effective and reach more young adults than health education through traditional mass media. However, little is known about what motivates people to actively initiate electronic word-of-mouth about health behaviors on their personal pages or profiles on social network sites, thus potentially reaching all their contacts on those sites.

Objective: This study filled the gap by applying a marketing theoretical model to explore the factors associated with electronic word-of-mouth on social network sites about leisure-time physical activity.

Methods: A Web survey link was sent to undergraduate students at one of the Midwestern universities and 439 of them completed the survey.

Results: The average age of the 439 participants was 19 years (SD=1 year, range: 18-24). Results suggested that emotional engagement with leisure-time physical activity (ie, affective involvement in leisure-time physical activity) predicted providing relevant opinions or information on social network sites. Social network site users who perceived stronger ties with all their contacts were more likely to provide and seek leisure-time physical activity opinions and information. People who provided leisure-time physical activity opinions and information were more likely to seek opinions and information, and people who forwarded information about leisure-time physical activity were more likely to chat about it.

Conclusions: This study shed light on the application of the electronic word-of-mouth theoretical framework in promoting health behaviors. The findings can also guide the development of future social marketing interventions using social network sites to promote leisure-time physical activity.

(J Med Internet Res 2017;19(6):e226) doi:10.2196/jmir.7017

KEYWORDS

physical activity; social marketing; social media

Introduction

Background

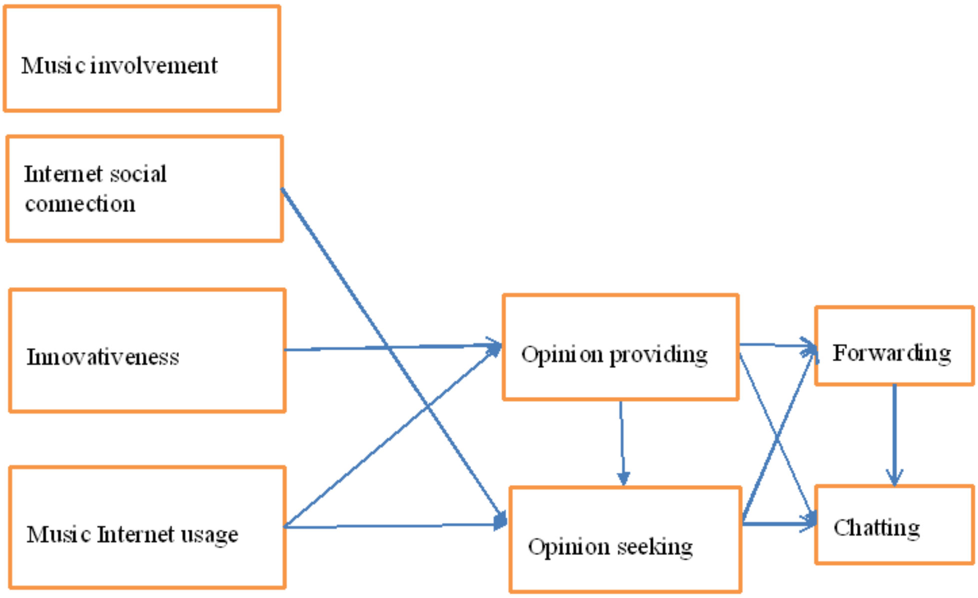
As of 2015, 90% of Internet users in the United States between the ages of 18 and 29 years use social network sites (SNSs) [1]. SNSs allow individuals to share their social connections and provide a platform for peer communication in their daily lives [2], offering great potential to influence a large number of young adults' health choices. For example, young adults liked receiving personalized health information about the HPV vaccine from peers on SNSs [3]. Further, social support from peers on SNSs can indirectly increase college students' intentions to participate in leisure-time physical activity (LTPA) [4]. Here, LTPA refers to a specific domain of physical activity in which people engage voluntarily during their free time because of physical activity has the potential to provide feelings of enjoyment, control, or mastery [5,6].

Engaging users to generate and share content is crucial to the outreach and success of health promotion activities that use SNSs [7]. Peer communication on SNSs can increase the reach of health information exponentially if forwarded consecutively by many people [8]. However, little is known about what motivates people to actively initiate peer communication about health topics on their own pages/profiles, where health information can be seen by all their contacts, rather than only by members of certain SNS health groups. This exploratory study applies a marketing theoretical model to fill this gap in understanding.

In marketing, online peer communication with "positive or negative statements made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" is referred to as electronic word-of-mouth (eWOM) [9]. eWOM on SNSs has been perceived by customers as a more reliable source of product information than the corporate-sponsored messages in traditional marketing, such as advertising, sales promotions, and public relations [10,11]. There are two components of eWOM: opinion providing and opinion seeking [12]. *Opinion providing* is the *process* by which some individuals influence the attitudes or behaviors of others [12]. Similarly, *opinion seeking* is the *process* by which people seek advice about a product [12,13]. In social marketing, the product can be the health behavior of the individual [14].

In this study, we focus on one specific outcome or product of a health behavior: LTPA. To systematically study the factors that affect the different components and formats of eWOM about LTPA, we applied a marketing theoretical framework, *the Path Model of the Antecedents and Consequences of Online Word-of-Mouth (hereafter the eWOM Path Model)* [12]. The eWOM Path Model is demonstrated in Figure 1 [12].

Figure 1. EWOM (electronic word-of-mouth) path model.



Antecedents of eWOM

Based upon the eWOM Path Model [12] and previous marketing literature [15], the first predictor of eWOM we investigated is involvement, which refers to a person's perception of a product's relevance, here LTPA, based on inherent needs, values, and interests. Product involvement, or personal interest in the product, or the excitement resulting from product ownership and use, motivates people to initiate traditional word-of-mouth (WOM) about commercial products [16]. Previous research suggests that involvement is associated with traditional WOM and eWOM [12,17]. Here, the role of LTPA involvement in eWOM, specifically on SNSs, is investigated.

Furthermore, this study distinguishes between two kinds of involvement: affective and cognitive. In marketing research, affect refers to "a class of mental phenomena characterized by a consciously experienced, subjective feeling state, commonly accompanying emotions and moods" [18]. Aligned with this description, this study defines affective involvement in LTPA as relevance to LTPA resulting from a consciously experienced, subjective feeling state, commonly accompanying emotions and moods. Correspondingly, cognitive involvement in LTPA can be defined as relevance to LTPA with respect to the cognitive attribution of LTPA. This research investigates if affective involvement, cognitive involvement, or both predict eWOM about LTPA.

Applying the eWOM Path Model [12] in the context of SNSs, general Internet social connection is replaced with two indicators of social connections on SNSs. The first indicator is perceived strength of ties to all contacts. With eWOM, participants engage in public communication with a network of people [19] and this makes one-to-one social ties less relevant in eWOM than in traditional WOM [20]. Thus, we regard all contacts on SNSs as a holistic network and investigate if perceived strength of ties to all contacts on SNSs is a predictor of eWOM about LTPA on SNSs.

To further scrutinize the network, this study examines another indicator of social connections, which is the ratio of strong ties on SNSs. For traditional WOM, strong ties often refer to close contacts like family and close friends, whereas weak ties include relationships with acquaintances [21,22]. Similarly, for eWOM, strong ties often refer to family members, relatives, and close friends, whereas weak ties include acquaintances, classmates, and neighbors [23]. For traditional WOM, referrals are more likely to happen between strong ties, which are also considered as more influential for the product [24]. However, for eWOM on SNSs, information can reach audiences with both strong and weak ties simultaneously. Thus, we hypothesize that people with larger numbers of strong ties are more likely to talk about LTPA on SNSs.

Applying the eWOM Path Model to study eWOM about online music, Sun and colleagues [12] found that people who actively use the Internet for music consumption were more likely to provide and seek opinions about online music. In the context of LTPA, the consumption of the product is in the form of participating in LTPA, which cannot be done online. Moreover, according to the literature [15], self-enhancement, which refers to the tendency to seek experiences that improve or bolster the self-concept by drawing attention to one's skills and talents [25], was one main antecedent of eWOM about a commercial product. Thus, we will investigate this relationship in the context of LTPA and hypothesize that people who participate in more LTPA are more likely to talk about it on SNSs as a means of self-enhancement.

Opinion Providing and Opinion Seeking

These are two distinguished components of eWOM: opinion providing and opinion seeking [12]. Adapted from a marketing context, here *opinion providing* is the *process* by which some individuals influence others' LTPA participation. Similarly, *opinion seeking* is the *process* by which people seek advice when participating in LTPA [12,13]. This study explores various antecedents for these two components of eWOM respectively, in order to guide future interventions that encourage young adults to provide or seek opinions of LTPA.

Formats of eWOM

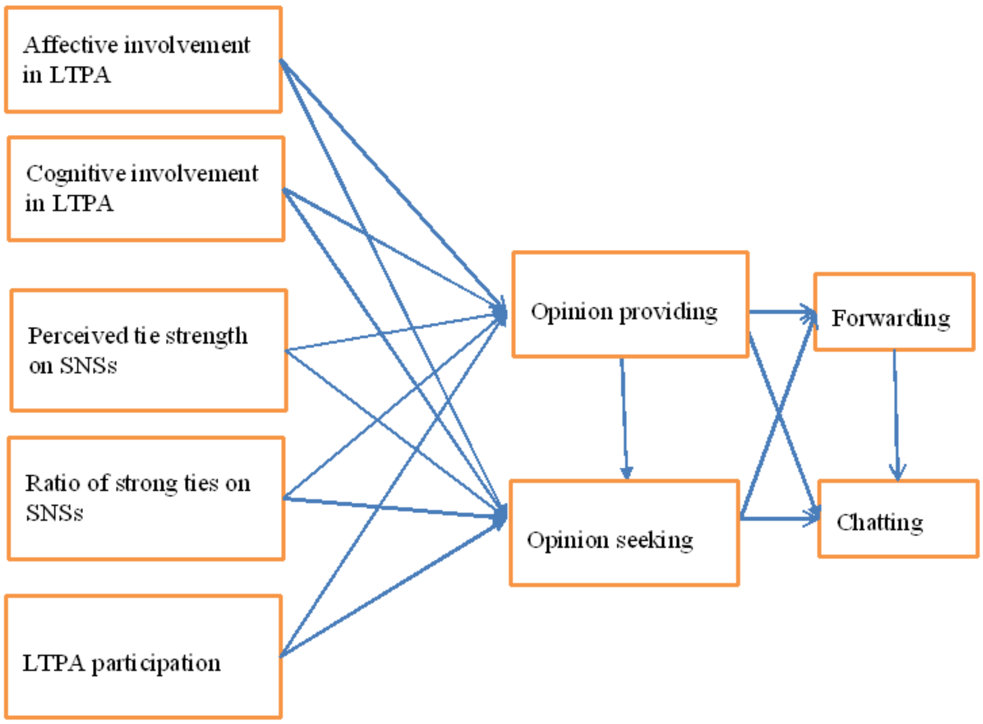
There are two formats in eWOM: chatting and forwarding [12]. Chatting is referred to as creating posts using one's own words [3,12]. eWOM on SNSs can be forwarded directly by reposting exact quotes from other contacts and posting existing URLs. In the eWOM Path Model [12], forwarding was the predictor of chatting.

Sun and colleagues [12] found that both opinion providing and seeking were predictors of chatting and forwarding about online music respectively. In a qualitative study on eWOM about HPV vaccine on Facebook, female college students revealed their preference for forwarding rather than chatting in general, because of their lack of HPV vaccine knowledge [3]. Applying the eWOM Path Model in the context of eWOM on SNSs about LTPA, this study investigates whether people who provide opinions are more likely to chat or forward information about LTPA on SNSs and whether people who seek opinions about LTPA on SNSs are more likely to chat or forward information about LTPA on SNSs.

Furthermore, Sun and colleagues [12] found that providing online opinions predicted online opinion-seeking. They speculated that people who are more likely to provide opinions are more likely to seek opinions in order to update their information and knowledge. In the context of eWOM on SNSs, one study found that providing opinions about dining is a predictor of seeking dining-related information on SNSs [26]. This study investigates if the same association exists in the context of LTPA.

In summary, this study applies the eWOM Path Model to LTPA to examine the factors associated with two components of eWOM, opinion providing and opinion seeking, by using two formats of eWOM: forwarding and chatting. The conceptual framework for this study is demonstrated in Figure 2.

Figure 2. Conceptual framework.



Methods

Data Collection

A mass recruitment email was sent to all undergraduate students enrolled at one of the Midwestern universities in the fall of 2011. In order to prevent self-selection due to participation in LTPA, the topic of LTPA was not mentioned in the recruitment email. A link to the Web survey was provided in the email. Participants were compensated with entry into a lottery in which US $10 gift cards were given to 50 participants. Participation was anonymous and this study was approved by the university's Institutional Review Board. In total, 439 participants out of all the undergraduate students between the age of 18 and 24 completed the survey within one month.

Measurements

Antecedents of eWOM

Affective Involvement

Participants were asked to rate four items adapted from Zaichkowsy's [27] Personal Involvement Inventory: "To me, participating in leisure-time physical activities is (1) 'boring-interesting,' (2) 'exciting-unexciting,' (3) 'appealing-unappealing,' or (4) 'fascinating-mundane'" on a 5-point (from 1 to 5) semantic differential scale. Items (2), (3), and (4) were reversely coded before composite scores were summed, where higher scores indicated higher affective involvement. The Cronbach alpha value was .90.

Cognitive Involvement

Participants were asked to rate six items adapted from Zaichkowsy's [27] Personal Involvement Inventory "To me, participating in leisure-time physical activities is (1) 'unimportant-important,' (2) 'irrelevant-relevant,' (3) 'means nothing-means a lot to me,' (4) 'worthless-valuable,' (5) 'involving-uninvolving,' or (6) 'not needed-needed'" on a 5-point (from 1 to 5) semantic differential scale. Item (5) was reversely coded before composite scores were aggregated, where higher scores indicated higher cognitive involvement. The Cronbach alpha value was .88.

Perceived Strength of Ties

Perceived strength of ties was assessed using three questions [28]. Participants were asked to rate how frequently they individually communicated with all their contacts on their most-used SNS on a scale of 1 to 7 (never-very frequently). Participants were then asked to rate the importance of all of their contacts on this SNS, ranging from 1 to 7 (not at all important-very important). Afterwards, participants rated how close they individually felt to all their contacts on this SNS, ranging from 1 to 7 (not at all close-very close). The Cronbach alpha value for perceived strength of ties was .70; composite scores were summed and higher scores indicated higher perceived strength of ties.

Perceived Ratio of Strong Ties

Perceived ratio of strong ties was assessed using two questions [23]. The first question asked "Approximately how many total contacts do you have on this SNS?" For Facebook/MySpace and other SNSs besides Twitter, "contacts" refers to "friends." For Twitter, "contacts" refers to "followers." A follow-up question asked, "How many contacts in each category (parents/guardians, siblings, other family members, and close friends) do you have on this SNS?" The ratio of strong ties was calculated by dividing the sum of parents/guardians, siblings, other family members, and close friends by the number of all contacts on this SNS. Higher scores indicated a higher perceived ratio of strong ties.

LTPA Participation

A physical activity checklist consisting of 27 types of LTPA was adapted from the Physical Activity Questionnaire for Adults to measure LTPA participation in the past 7 days [29]. For each type of LTPA, participants were asked to indicate (1) whether they participated in it and (2) how many times they had participated in it over the previous seven days. Participation counts for each type of LTPA over the previous seven days were categorized into five groups: (1) 0=none; (2) 1=1-2 times; (3) 2=3-4 times; (4) 3=5-6 times; (5) 4=7 times or more. LTPA participation was measured using the mean score calculated by dividing total participation counts for all types of LTPA by the number of types of LTPA. Higher LTPA participation scores indicated increased participation.

Components of eWOM

Opinion Providing About LTPA

The 5-point Likert scale ranging from "strongly disagree" to "strongly agree" used by Flynn et al [13] for opinion leaders was adapted in this study. The following is an example of one of the three items: "My contacts on this SNS pick a leisure-time physical activity based on what I have told them on the SNS." The Cronbach alpha value was .89. The composite scores were summed.

Opinion Seeking About LTPA

The 5-point Likert scale ranging from "strongly disagree" to "strongly agree" for opinion seeking developed by Flynn et al [13] was adapted in this study. A representative example, one of the three items, is "When considering participating in a leisure-time physical activity, I ask my contacts for advice on this SNS." Composite scores were summed. The Cronbach alpha value was .84.

Formats of eWOM

Forwarding About LTPA on This SNS

The measurement contained six items with a 5-point Likert scale, ranging from "strongly disagree" to "strongly agree" [12]. One item was, "I tend to use the 'Share this site' function to share websites on this SNS when I find information about leisure-time physical activity." Composite scores were summed for each item. Higher scores indicated that participants forwarded more information about LTPA on their preferred SNS. The Cronbach alpha value for forwarding was .92.

Chatting About LTPA on This SNS

The chatting measure contained five items with a 5-point Likert scale, ranging from "strongly disagree" to "strongly agree" [12]. One item was, "On this social network site, I like to share information with my contacts about my favorite leisure-time physical activity." Composite scores were summed. Higher scores indicated that participants shared more about LTPA with their SNS contacts. The Cronbach alpha value was .91.

Covariates

Demographics

Participants provided their age (between 18 and 24), gender (male or female), and race/ethnicity (non-Hispanic white or other).

Data Analysis

Descriptive analyses were conducted using SPSS version 22 (IBM Corp). Spearman rank correlations were tested among the antecedents, components, and consequences of eWOM. Path models with 439 cases were conducted using AMOS version 22 (IBM Corp), a structural equation modeling program. AMOS has been widely used in other health communication studies [30]. Diagnostic analyses were performed to test the assumptions of the path model analysis. Outliers (n=22) were deleted and the square root transformation was used for the ratio of strong ties.

Comparative fit index (CFI) and the root mean square error of approximation (RMSEA) were used to estimate the model fit. According to Kline [31], a CFI value over .90 and RMSEA value between .05 and .08 indicate an acceptable model fit [31]. Multi-group analyses were conducted to test the difference between groups based on sex and ethnicity, respectively. No significant differences were observed. Age was not significantly associated with any of the eWOM components and consequences in the path model. Thus, the final path model did not include these covariates. The path coefficients were standardized regression coefficients with $P<.05$, indicating significant associations.

Results

Demographics

The average age of the 439 participants was 19 years (SD=1 year, range: 18-24). Just over 79.0% (343/434) of the participants were female and the majority of the participants were non-Hispanic whites (81.0%%, 354/437). More than 77.9% (341/438) of the participants indicated that they visited their SNS several times a day.

Bivariate Analysis Results

Spearman correlation matrix is shown in Table 1 and includes only 394 cases as 45 cases, with missing values for at least one of these variables, were deleted list-wise. These correlations provide pair-wise associations between the variables and give an indication of potential relationships in the path model.

Table 1. Spearman correlation, mean, and standard deviation for variables of interest.

N=394	1	2	3	4	5	6	7	8	9	Mean	SD
1. Affective involvement	1.00									15.26	3.89
2. Cognitive involvement	.73									25.31	4.41
3. Perceived strength of ties	.06	−.02								11.73	3.54
4. Ratio of strong ties	.06	.05	.18[a]							18.02	25.46
5. LTPA participation	.31[a]	.32[a]	−.06	.02						1.86	1.44
6. Opinion providing	.23[a]	.18[a]	.27[a]	−.00	.09					6.56	3.26
7. Opinion seeking	.12*	.04	.29[a]	−.00	−.01	.53[a]				5.78	3.01
8. Forwarding	.13[a]	.08	.27[a]	−.01	.09	.54[a]	.62[a]			10.54	5.45
9. Chatting	.25[a]	.19[a]	.34[a]	.06	.10[b]	.68[a]	.71[a]	.70[a]	1.00	10.69	5.15

[a] Correlation is significant at the 0.01 level (2-tailed).

[b] Correlation is significant at the 0.05 level (2-tailed).

Path Analysis Results

Statistically significant relationships were observed between perceived strength of ties and both opinion providing and opinion seeking, whereas cognitive involvement was associated with opinion providing only (see Table 2). Forwarding and chatting behaviors were also associated with opinion providing and seeking behaviors (see Table 2). The path model produced a good fit (RMSEA=0.054; CFI=0.990; χ^2_{10}=22.1, *P*=.01). Path coefficients are displayed in Table 2, while Figure 3 illustrates only significant paths.

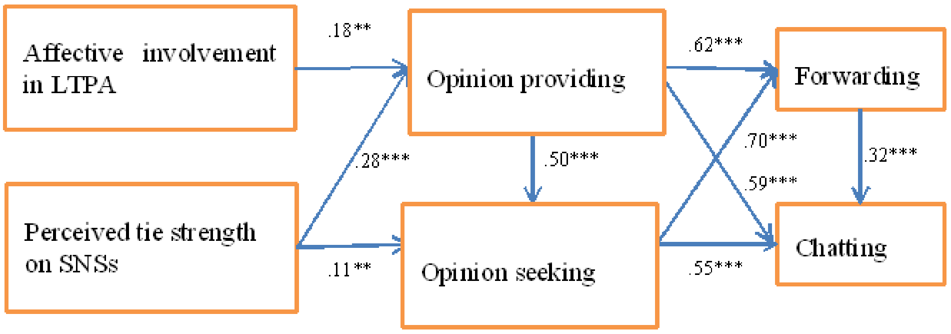
Table 2. Parameter estimates for causal paths.

Hypotheses	Causal Paths[a]	Path coefficient estimates[b]	Standard error[b]	*P* value[b]
H1a	Affective involvement →Opinion providing	.175	.058	.003
H1b	Affective involvement →Opinion seeking	.058	.049	.24
H2a	Cognitive involvement →Opinion providing	.018	.052	.73
H2b	Cognitive involvement →Opinion seeking	−.080	.044	.07
H3a	Perceived strength of ties →Opinion providing	.281	.043	<.001
H3b	Perceived strength of ties →Opinion seeking	.110	.038	.004
H4a	Ratio of strong ties →Opinion providing	−.054	.068	.43
H4b	Ratio of strong ties →Opinion seeking	−.040	.057	.49
H5a	LTPA participation →Opinion providing	−.051	.106	.63
H5b	LTPA participation →Opinion seeking	−.099	.089	.26
H6a	Opinion providing →Forwarding	.642	.083	<.001
H6b	Opinion providing →Chatting	.591	.059	<.001
H7a	Opinion seeking →Forwarding	.702	.087	<.001
H7b	Opinion seeking →Chatting	.501	.063	<.001
H8	Opinion providing →Opinion seeking	.497	.043	<.001
H9	Forwarding → Chatting	.317	.034	<.001

[a] Goodness-of-fit statistics: CFI=0.990; RMSEA=0.054.

[b] Results controlled for age, gender, race/ethnicity, length of SNS membership, and frequency of using SNSs.

Figure 3. Path model with significant path coefficients only.



Only affective involvement was significantly associated with online opinion providing (*P*=.003). Perceived tie strength was significantly associated with both online opinion providing (*P*<.001) and opinion seeking (*P*=.004), as expected. As outlined in the path model, both online opinion providing and online opinion seeking were significantly associated with forwarding (*P*<.001 for both) and chatting (*P*<.001 for both). Online opinion providing is significantly associated with online opinion seeking (*P*<.001). Additionally, forwarding was significantly associated with chatting (*P*<.001).

Discussion

Principal Findings

This study bridges a gap in the literature by exploring the antecedents and consequences of eWOM about LTPA on SNSs among traditionally-aged (18-24 years) college students. The study observed that students with high affective involvement and stronger perceived strength of ties on SNSs were more likely to provide information and/or opinions about LTPA. However, cognitive involvement and the ratio of strong ties were not significantly related to higher opinion providing. The study supported the ideas that forwarding and chatting were two consequences for both opinion providing and opinion seeking.

Antecedents of eWOM

This was one of the first studies that differentiated between affective and cognitive involvement in the role of predicting opinion providing and opinion seeking. In studies on traditional WOM, different kinds of involvements, such as product-involvement, self-involvement, other-involvement (eg, to “share” the pleasure), and message-involvement have been documented [32]. Previous research on eWOM has, so far, not distinguished between the various kinds of involvement [12].

This study revealed the roles that these two types of involvement play in predicting eWOM. Only affective involvement, and not cognitive involvement, in LTPA was a predictor of opinion providing on SNSs. This suggests that, if students found a certain LTPA interesting, exciting, appealing, and/or fascinating, they were more likely to share the experience on SNSs. The importance of affective involvement is supported by one study, which found that students’ top motives for sending pass-along emails were “because it’s fun, and because they enjoyed it, because it’s entertaining” [33]. Similarly, in e-commerce, the effect of affective involvement on purchase intention was nearly double that of cognitive involvement [34]. It is worth noting that the described affect is a positive one here, in that participants described positive emotions that arose from engaging in LTPA, which in turn led to providing their opinion. This is a well-known path to generating eWOM, as past research shows that the main motivation to share information online is because people are positively involved with a brand [35].

Neither affective nor cognitive involvement predicted opinion seeking. This could be explained by the difference between LTPA and other traditional commercial products. As is not the case with these products, individual LTPA participation could be considered more personal, fueling needs for specific opinions and information, which may be communicated offline or through private online messages. Alternatively, due to its experiential nature, LTPA could lead to different experiences for different people because of the various physical capabilities, accessibility to the LTPA, and so on, whereas purchasing a product generates a more uniform response. Past research has shown that experiential eWOM comments are seen as less helpful than utilitarian ones, thus possibly driving online users not to seek them in the first place [36].

Adapting to the SNSs context, this study investigated two indicators of strength of ties, perceived strength of ties, and ratio of strong ties and found different effects for each on eWOM. SNS users who perceived strong ties with all their contacts on this SNS were more likely to provide or seek opinions about LTPA. In a previous study, only opinion seeking was influenced by strong ties [12]. Context may play a role in these superficially different results as physical activity could be considered a more personal experience than music. Thus, users might feel more comfortable providing opinions about LTPA when they feel closer to all their contacts on SNSs because of the personal nature of LTPA experiences.

Consistent with previous work [37], perceived ratio of strong ties in users’ social networks on the SNSs was not a predictor for any of the eWOM components in this study. This could be explained by the idea that users use more privacy-related functions (eg, sharing photos) to communicate with close friends on Facebook [38]. Our findings imply that users talk about LTPA on SNSs regardless of the number of close friends.

Surprisingly, LTPA participation was not related to either component of eWOM—opinion providing or opinion seeking. These results are inconsistent with a major motivation to use SNSs: self-presentation [39]. SNSs allow users to construct their identities through pictures, profile information, and wall content [40]. When examining eWOM about physical activity on Twitter, one study found that 38% of the tweets about LTPA were about users' previous and current LTPA participation [41]. One explanation could be that there are different channels for opinion providing and opinion seeking for someone who participates in LTPA. The Google/Keller Fay group [42] found that the predominant method of word-of-mouth is still offline. Therefore, it is possible that those who participate in LTPA would communicate with their friends through other channels such as face-to-face, phone calls, or text messaging, rather than through eWOM on SNSs. Further, this result shows that positive emotional involvement with LTPA is more important for sharing opinions than actual participation in LTPA.

Consequences of eWOM

This study found that both chatting and forwarding were consequences of increased opinion providing and opinion seeking about LTPA on SNSs. SNS users who were more likely to provide opinions about LTPA on SNSs were also more likely to seek opinions and information. Additionally, users who were more likely to forward information about LTPA were more likely to chat about it too. Students with either increased opinion providing or opinion seeking were more likely to either forward information about LTPA or chat about it. . These results are consistent with the findings of Sun and colleagues on music-related eWOM [12].

This study provides the first empirical evidence for adapting the eWOM Path Model [12] to a heath context. These findings can guide future interventions using eWOM on SNSs to promote physical activity and other health behaviors. SNS users who have positive emotional connections to LTPA and who feel closely connected to all their contacts can be recruited to be peer leaders to disseminate information about LTPA. Further, public health practitioners should emphasize the "fun" part of LTPA classes or programs to harness the potential for eWOM on SNS and increase participation. Examples might include, "Even if you have never danced before, you will have fun and feel successful in a Zumba class!" [43].

In terms of future research, this study provides the foundation for examining the antecedents and formats of health-related eWOM, which has the highest potential for being shared or discussed. One of the values of eWOM is its organic spread among networks of users or, in other words, its viral aspect. Certain messages (eg, positive emotional ones) may have a higher viral potential than others and future research can examine this further, which will benefit health marketers by harnessing users' natural tendencies to share information and enabling them to motivate one another into healthier behaviors.

The study population was limited by self-selection; however, efforts were made to decrease the selection bias through careful wording in the recruitment email. Specifically, in an effort to reach those who do not participate or are not interested in LTPA, the LTPA focus of the study was not mentioned in the recruitment email survey introduction. However, the study sample may not be representative of the population of undergraduate students at one of the midwestern universities. Furthermore, we only included participants at traditional college, aged between 18 and 24. Study respondents tended to be younger, freshmen, living on campus, and female, as compared with the population of Midwestern University students. Thus, caution is required when applying the results to a larger population.

Another limitation of this study is associated with self-reporting. Although the online questionnaires were filled anonymously, the possibility of social desirability to over-report the values of certain variables, such as the LTPA and opinion providing about LTPA on SNSs, cannot be ruled out.

Conclusions

This study contributes to the existing literature by examining the antecedents and formats of eWOM about LTPA on SNSs, applying a theoretical framework typically used in marketing. Results indicate that young adults who feel that LTPA is interesting and fun and are closer to all of their contacts on SNSs, are more likely to provide opinions and information about LTPA on SNSs. Those who provide opinions and information are also more likely to seek opinions and information from others. Those who forward information about LTPA are more likely to talk about LTPA using their own words. Future interventions using SNSs to spread eWOM about LTPA should increase young adults' affective involvement in order to encourage their organic comments to other users. Young adults who are more affectively involved in LTPA and feel closer to all their contacts on SNSs could be recruited to be peer leaders to provide opinions and information. Future interventions can also encourage young adults who forward information about LTPA on SNSs to chat about it too, as that is a more engaged behavior, which may motivate them or others into participation in physical activity.

Conflicts of Interest

None declared.

References

1. Pew Research Center. 2015. Social media usage: 2005-2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [accessed 2017-05-22] [WebCite Cache ID 6qf3Nh9Pz]
2. Boyd DM, Ellison NB. Social network sites: definition, history, and scholarship. J Comput Mediat Commun 2007;13(1):210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]

3. Zhang N, Tsark J, Campo S, Teti M. Facebook for health promotion: female college students' perspectives on sharing HPV vaccine information through facebook. Hawaii J Med Public Health 2015 Apr;74(4):136-140 [FREE Full text] [Medline: 25954600]
4. Zhang N, Campo S, Yang J, Janz KF, Snetselaar LG, Eckler P. Effects of social support about physical activity on social networking sites: applying the theory of planned behavior. Health Commun 2015;30(12):1277-1285. [doi: 10.1080/10410236.2014.940669] [Medline: 26086237]
5. Kleiber DA, Walker GJ, Mannell RC. A Social Psychology of Leisure. State College, PA: Venture Publishing Inc; 2011:3-28.
6. Howley ET. Type of activity: resistance, aerobic and leisure versus occupational physical activity. Med Sci Sports Exerc 2001;33(Supplement):S364-S369. [doi: 10.1097/00005768-200106001-00005]
7. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
8. Ho JY, Dempsey M. Viral marketing: motivations to forward online content. J Bus Res 2010 Sep;63(9-10):1000-1006. [doi: 10.1016/j.jbusres.2008.08.010]
9. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? J Interact Market 2004 Jan;18(1):38-52. [doi: 10.1002/dir.10073]
10. Foux G. Consumer-generated media: get your customers involved. Brand Strategy 2006;8:38-39.
11. Mangold W, Faulds D. Social media: the new hybrid element of the promotion mix. Business horizons 2009;52(4):357-365.
12. Sun T, Youn S, Wu G, Kuntaraporn M. Online word-of-mouth (or mouse): an exploration of its antecedents and consequences. J Comp Mediated Comm 2006 Jul;11(4):1104-1127. [doi: 10.1111/j.1083-6101.2006.00310.x]
13. Flynn LR, Goldsmith RE, Eastman JK. Opinion leaders and opinion seekers: two new measurement scales. J Acad Market Sci 1996 Mar 01;24(2):137-147. [doi: 10.1177/0092070396242004]
14. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011;13(1):e17.
15. King R, Racherla P, Bush V. What we know and don't know about online word-of-mouth: a review and synthesis of the literature. J Interact Market 2014;28(3):167-183.
16. Sundaram D, Mitra K, Webster C. Word-of-mouth communications: a motivational analysis. Adv Consum Res 1998;25(1):527-531.
17. Zaichkowsky JL. Measuring the involvement construct. J Consum Res 1985 Dec;12(3):341. [doi: 10.1086/208520]
18. Westbrook RA. Product/Consumption-based affective responses and postpurchase processes. J Mark Res 1987 Aug;24(3):258-270. [doi: 10.2307/3151636]
19. Kozinets R, De VK, Wojnicki A, Wilner S. Networked narratives: understanding word-of-mouth marketing in online communities. J Marketing 2010;74(2):71-89.
20. Brown J, Broderick A, Lee N. Word of mouth communication within online communities: conceptualizing the online social network. J Interact Market 2007;21(3):2-20.
21. Granovetter MS. The Strength of weak ties. Am J Sociol 1973 May;78(6):1360-1380. [doi: 10.1086/225469]
22. Granovetter M. The strength of weak ties: a network theory revisited. Sociological Theory 1983;1:201. [doi: 10.2307/202051]
23. Chu S, Choi S. Electronic word-of-mouth in social networking sites: a cross-cultural study of the United States and China. J Global Market 2011;24(3):263-281.
24. Brown J, Reingen P. Social ties and word-of-mouth referral behavior. J Cons Res 1987;14(3):350-362.
25. Ross MW. Typing, doing, and being: sexuality and the internet. J Sex Res 2005 Nov;42(4):342-352. [doi: 10.1080/00224490509552290] [Medline: 19827239]
26. Bilgihan A, Peng C, Kandampully J. Generation Y's dining information seeking and sharing behavior on social networking sites: an exploratory study. Int J Contemp Hosp M 2014;26(3):349-366.
27. Zaichkowsky J. The personal involvement inventory: reduction, revision, and application to advertising. J Advertising 1994;23(4):59-70.
28. Chu SC, Kim Y. Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites. IJA 2011;30(1):47-75.
29. Copeland J, Kowalski K, Donen R, Tremblay M. Convergent validity of the physical activity questionnaire for adults: the new member of the PAQ Family. J Phys Act Health 2005;2(2):216-229.
30. Osborn CY, Cavanaugh K, Wallston KA, Rothman RL. Self-efficacy links health literacy and numeracy to glycemic control. J Health Commun 2010;15(Suppl 2):146-158 [FREE Full text] [doi: 10.1080/10810730.2010.499980] [Medline: 20845200]
31. Kline R. Principles and practice of structural equation modeling. New York, NY: Guilford press; 2011:103-111.
32. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD. Electronic word-of-mouth via consumer opinion platforms: what motivates consumers to articulate themselves on the internet? J Interact Market 2004;18(1):38-52.
33. Phelps JE, Lewis R, Mobilio L, Perry D, Raman N. Viral marketing or electronic word-of-mouth advertising: examining consumer responses and motivations to pass along email. JAR 2004 Dec 01;44(4):333-348. [doi: 10.1017/S0021849904040371]
34. Jiang Z, Chan J, Tan BC, Chua WS. Effects of interactivity on website involvement and purchase intention. J Assoc Inf Syst 2010;11(1):34-59.

35. Verhaeghe A, Van Belleghem S, Vergult C, Schillewaert N, Claus D. A new approach for measuring “buzz” word of mouth and word of mouse. 2007 Presented at: Proceedings of the ESOMAR Worldwide Multi Media Measurement Conference; June 2007; Dublin.
36. Pan Y, Zhang JQ. Born unequal: a study of the helpfulness of user-generated product reviews. Journal of Retailing 2011;87(4):598-612.
37. Vilpponen A, Winter S, Sundqvist S. Electronic word-of-mouth in online environments: exploring referral networks structure and adoption behavior. J Interact Advert 2013;6(2):8-77.
38. Hsu CJ, Wang C, Tai Y. The closer the relationship, the more the interaction on facebook? investigating the case of Taiwan users. Cyberpsychol Behav Soc Netw 2011;14(7-8):473-476. [doi: 10.1089/cyber.2010.0267] [Medline: 21288075]
39. Nadkarni A, Hofmann SG. Why do people use facebook? Pers Individ Dif 2012 Feb 01;52(3):243-249 [FREE Full text] [doi: 10.1016/j.paid.2011.11.007] [Medline: 22544987]
40. Zhao S, Grasmuck S, Martin J. Identity construction on Facebook: Digital empowerment in anchored relationships. Comput Human Behav 2008;24(5):1816-1836.
41. Zhang N, Campo S, Janz K, Eckler P, Yang J, Snetselaar L, et al. Electronic word of mouth on twitter about physical activity in the United States: exploratory infodemiology study. J Med Internet Res 2013;15(11):e261.
42. Google/Keller Fay group. Gstatic. Word of mouth and the internet URL: http://www.gstatic.com/ads/research/en/2011_Word_of_Mouth_Study.pdf [accessed 2017-05-26] [WebCite Cache ID 6qkizyXQU]
43. University of Iowa Recreation Services. Uiowa. 2012. Fitness programs URL: http://recserv.uiowa.edu/Apps/Programs/FitnessPrograms.aspx [accessed 2017-05-24] [WebCite Cache ID 6qkj4I9nq]

Abbreviations

eWOM: electronic word-of-mouth
LTPA: leisure-time physical activity
SNS: social network sites
WOM: word-of-mouth

Edited by G Eysenbach; submitted 20.11.16; peer-reviewed by J Tayaben, D Li; comments to author 05.02.17; revised version received 11.03.17; accepted 14.04.17; published 22.06.17.

Please cite as:
Zhang N, Campo S, Yang J, Eckler P, Snetselaar L, Janz K, Leary E
What Motivates Young Adults to Talk About Physical Activity on Social Network Sites?
J Med Internet Res 2017;19(6):e226
URL: http://www.jmir.org/2017/6/e226/
doi:10.2196/jmir.7017
PMID:28642215

Original Paper

Scaling Up Psychological Treatments: A Countrywide Test of the Online Training of Therapists

Christopher G Fairburn[1*], FMedSci, FRCPsych; Elizabeth Allen[2], MSc, PhD; Suzanne Bailey-Straebler[1], MSN, DPhil; Marianne E O'Connor[1], BA; Zafra Cooper[1,3*], DPhil, Dip Clin Psych

[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom
[2]London School of Hygiene and Tropical Medicine, London, United Kingdom
[3]Yale School of Medicine, Department of Psychiatry, New Haven, CT, United States
[*]these authors contributed equally

Corresponding Author:
Christopher G Fairburn, FMedSci, FRCPsych
Department of Psychiatry
University of Oxford
Warneford Hospital
Oxford,
United Kingdom
Phone: 44 186 561 3199
Email: credo@medsci.ox.ac.uk

Abstract

Background: A major barrier to the widespread dissemination of psychological treatments is the way that therapists are trained. The current method is not scalable.

Objective: Our objective was to conduct a proof-of-concept study of Web-centered training, a scalable online method for training therapists.

Methods: The Irish Health Service Executive identified mental health professionals across the country whom it wanted to be trained in a specific psychological treatment for eating disorders. These therapists were given access to a Web-centered training program in transdiagnostic cognitive behavior therapy for eating disorders. The training was accompanied by a scalable form of support consisting of brief encouraging telephone calls from a nonspecialist. The trainee therapists completed a validated measure of therapist competence before and after the training.

Results: Of 102 therapists who embarked upon the training program, 86 (84.3%) completed it. There was a substantial increase in their competence scores following the training (mean difference 5.84, 95% Cl –6.62 to –5.05; $P<.001$) with 42.5% (34/80) scoring above a predetermined cut-point indicative of a good level of competence.

Conclusions: Web-centered training proved feasible and acceptable and resulted in a marked increase in therapist competence scores. If these findings are replicated, Web-centered training would provide a means of simultaneously training large numbers of geographically dispersed trainees at low cost, thereby overcoming a major obstacle to the widespread dissemination of psychological treatments.

(J Med Internet Res 2017;19(6):e214) doi:10.2196/jmir.7864

KEYWORDS

psychotherapy; training; dissemination; Internet; eating disorders; cognitive behavior therapy

Introduction

Psychological treatments are difficult to disseminate and implement [1,2]. They are inherently complex in form; they are labor-intensive and therefore costly; and, in contrast with drug treatments, there is no commercial drive to disseminate them. An additional obstacle is the way in which therapists are trained. This typically involves a therapist with generic training in psychological treatment delivery attending a specialist training workshop given by an expert and then practicing the treatment under the supervision of someone experienced at delivering the treatment [3]. As there is a scarcity of people qualified to provide both the workshops and the clinical supervision, this method is incapable of meeting the worldwide demand for training [4].

To address these problems, Web-centered training was devised [4]. Two main features distinguish it from conventional training. First, the training is program-led; that is, the training program itself leads the therapist through the training obviating the need for input from an external expert. Second, by delivering the training over the Internet, large numbers of geographically dispersed therapists can be trained at one time.

This investigation was a proof-of-concept study with 2 aims. The first was to determine whether Web-centered training is feasible and acceptable to therapists. The second aim was to establish its likely effectiveness. In common with most large-scale interventions, Web-centered training was not expected to be as potent as more intensive methods [5], a shortcoming that would be offset by its far greater scalability.

Methods

Design and Context

This was a proof-of-concept study in which a cohort of therapists across Ireland was offered Web-centered training in a specific psychological treatment. Their competence was assessed before and after the training.

The opportunity to undertake the study arose as a result of the Irish Health Service Executive (HSE) deciding to establish a countrywide eating disorder service. This necessitated identifying mental health professionals from across the Republic of Ireland whom the HSE wished to receive training in enhanced cognitive behavior therapy for eating disorders (CBT-E) [6,7]. CBT-E was chosen by the HSE as its treatment of choice because it has a strong evidence base across the full range of eating disorders, a decision consonant with the 2017 National Institute for Health and Care Excellence guidelines on eating disorders [8]. The Centre for Research on Eating Disorders at Oxford (CREDO) was asked to provide the training because it developed CBT-E [6]. Fortuitously, this request came at a time when CREDO was completing the development of a Web-centered training program in CBT-E. It therefore provided an opportunity to conduct a proof-of-concept study of the new method of training.

The research protocol was submitted to Oxford University Central Research Ethics Committee. It was judged not to require formal ethical approval.

Recruitment of the Trainees

The HSE identified 139 therapists as potential candidates for training. There were no formal inclusion or exclusion criteria. Each of these therapists was sent an email from CREDO in which they were invited to take part in an evaluation of the new form of training. They were provided with details about CBT-E and the study (Multimedia Appendix 1) together with an outline of the content of the training website (Multimedia Appendix 2). Those therapists who indicated that they were willing to take part and who provided informed consent were asked for details of their professional background, age, gender, and clinical experience. They were then given access to the training program for 20 weeks. Figure 1 shows the figures for the recruitment and retention of the trainees.

Figure 1. Recruitment and retention of the trainees.

The Web-Centered Training Program

The Web-centered training program in CBT-E has 2 main parts, The Course and The Library. The Course is linear in nature and takes a minimum of 9 hours to complete (Multimedia Appendix 2). It provides a detailed practical description of how to implement the main focused form of CBT-E given by an expert on the treatment (CGF). This description takes the form of multiple brief video presentations accompanied by handouts and separated by formative learning exercises, video recordings of acted illustrations of the treatment, and tests of knowledge accompanied by feedback. While working through The Course, trainees are encouraged to read relevant sections from the CBT-E treatment guide [6] and treat 1 or 2 patients. As implementing CBT-E with patients who are not severely underweight takes up to 20 weeks, access to the program was for this length of time. It was stressed to the trainees that clinical responsibility for their patients remained with their local clinical team and not with CREDO.

The second part of the training website, The Library, provides supplementary information on how to use CBT-E with subgroups of patients including those who are severely underweight and those with clinical perfectionism, core low self-esteem, or marked interpersonal difficulties. There is also a detailed account of how to use CBT-E to treat young patients. At the time of conducting this study, The Library was not available but trainees had full access to all the material of relevance to the implementation of the focused version of CBT-E.

Web-centered training is designed to be undertaken ether autonomously (independent training) or with support from a nonspecialist support worker (supported training). Although the supported form requires external input, it is still scalable because the role of the support worker is restricted to encouraging the trainee to follow the training program. In this study the supported form of training was used. Thus, in addition to having access to the website, the trainees were offered up to 12 telephone calls over the 20 weeks of training, each lasting no longer than 30 minutes. These calls were weekly at first and then every 2 weeks. The support workers followed a protocol that defined their role, and they were supervised by 2 senior clinicians (ZC and SBS); 2 of the support workers were research assistants with no clinical experience and 2 had clinical experience but restricted themselves to the supportive role.

Assessment of Competence

The competence measure was developed in advance of the study. It was designed to be scalable in nature so that it could be used in large-scale training studies such as this one. Its development and validation are described in detail in a separate publication [9]. This included detailed blueprinting, state-of-the-art item writing, and independent item review with initial field testing, followed by formal Rasch analysis. The measure was developed independently of the Web-centered training program in order to avoid teaching to the test.

The final version of the measure comprises 22 multiple choice questions, each of which addresses the trainee's knowledge and understanding of the focused version of CBT-E and its implementation (ie, applied knowledge). Total scores on the measure have been related to performance at implementing CBT-E, and there is a cut-point indicative of a good level of competence at delivering CBT-E. Three equivalent versions of the measure are available so that different versions can be used before and after training.

Data Analysis

Regression analysis was used to estimate the mean difference in therapist competence scores and to explore whether trainee characteristics (age, gender, professional background, years of clinical experience), support characteristics (support worker background, number of supportive phone calls), and the trainee's use of a practice case were associated with change. Logistic regression (adjusted for baseline score) was used to explore factors associated with the likelihood of trainees scoring over the competence cut-point following the training.

Results

Participants

Of the 139 potential participants identified by the HSE, 102 embarked upon Web-centered training (Figure 1), 93 of whom were female. Their mean age was 40.1 (SD 8.1) years, and their mean length of clinical experience was 12.0 (SD 7.9) years. The professional background of the trainees is shown in Table 1. They comprised 2 main professional groups, clinical psychologists (37.3%) and psychiatric nurses or nurse therapists (34.4%). Three-quarters (74.4%) currently treated patients with eating disorders, although a third (33.3%) did so rarely.

Table 1. Professional background of the trainees (N=102).

Profession	Number n (%)
Clinical psychologist	38 (37.3)
Dietitian	3 (2.9)
Family therapist	1 (1.0)
Nurse therapist	2 (2.0)
Occupational therapist	9 (8.8)
Psychiatric nurse	33 (32.4)
Psychiatrist	3 (2.9)
Social worker	13 (12.7)

Completion of Training

Of 102 trainees, 86 (84.3%) completed the full training program. Of the remainder, 11 failed to start (6 cited practical reasons for withdrawing), and 5 were lost to contact (Figure 1).

Outcome of Training

Of the 86 trainees who finished the entire training program, 80 completed the competence measure before and after their training. Their scores increased substantially from 5.39 before training (95% CI 4.78-6.00) to 11.23 after training (95% CI 10.46-11.99), the mean difference being 5.84 (95% Cl –6.62 to –5.05, *P*<.001). One of these trainees scored above the competence cut-point prior to the training (1.3%, 1/80) whereas 34 did so afterwards (42.5%, 34/80).

The exploratory logistic regression analyses identified 1 significant predictor of outcome. Psychiatric nurses and nurse therapists had a smaller increase in their competence scores than the other professional groups (mean difference 2.52, *P*=.003) and were less likely to score above the cut-point (OR 0.17, 95% CI 0.05-0.58; *P*=.005).

Discussion

The Study and its Findings

This study was an opportunistic initial evaluation of a new scalable way of training therapists to deliver specialist psychological treatments. It capitalized on a training initiative of the Irish HSE. It found that Web-centered training, when accompanied by nonspecialist support, was practicable, acceptable, and reasonably effective. Over 80% of the trainees completed the training program, and their scores on the validated measure of competence increased substantially. Almost half obtained competence scores indicative of a good level of competence.

Certain practical difficulties were encountered. One was technical. A beta version of the website was used and this proved unstable at times. The other was logistic. Some trainees reported difficulties implementing the training during their working hours. This was for 2 main reasons. First, given their clinical commitments, it was difficult to protect time for the training and second, in some rural areas there was limited access to the Internet. Both difficulties proved surmountable.

Comparison with Other Training Studies

It is difficult to provide a context within which to interpret these findings because therapist training has been relatively neglected as a research topic [10,11]. Furthermore, many of the training studies to date have been small in scale and few have used competence measures with a validated cut-point [12]. The competence figures that have been reported following training vary widely with higher figures coming from training programs that have included extensive supervision, a method that is not scalable and therefore of little relevance in this context.

Strengths of the Study

The study had several strengths. First, it was conducted in a real-world setting and should therefore be viewed as an effectiveness study rather than an efficacy one. Second, the sample size was large for a training study although it was too small to draw firm conclusions regarding subgroups within the sample. This qualification particularly applies to the exploratory predictor analyses. Third, it used a validated measure of therapist competence.

Limitations of the Study

The study had certain limitations. First, the trainees were selected by the HSE rather than being volunteers as would usually be the case. Second, there was no posttraining follow-up to determine whether the effects of the training persisted and whether the trainees put into practice what they had learned. Third, the study was not controlled. We cannot therefore discount the possibility that the increase in therapist competence scores would have happened spontaneously although this seems improbable. Nor do we know how this form of training would compare in effectiveness with other training methods. A comparison with the conventional method of training might seem the obvious design but we believe that the findings would be of limited interest given the poor scalability of the latter method. Instead, a comparison of independent and supported training would be of greater value as both are scalable yet one has higher inherent costs and is harder to implement.

Conclusions

Two findings emerge from this opportunistic study. First, Web-centered training is feasible and well accepted. Over 80 therapists, dispersed across a wide geographical area, completed the training program despite receiving only a limited amount of nonspecialist support. Training on this scale, and at a distance, would be impossible with the current method of training. Second, the training was reasonably effective as almost half the trainees obtained a competence score indicative of a good level of competence. While this figure is lower than that reported with more intensive but unscalable forms of training, it is impressive given the scalability of the new method.

If these findings were replicated, Web-centered training would provide a means of simultaneously training large numbers of geographically dispersed therapists thereby overcoming a major obstacle to the dissemination of psychological treatments.

Acknowledgments

We are grateful to the Health Service Executive for its assistance with the project and to Ian Daly and Rhona Jennings in particular. We are also grateful to Erin Charlton and Beth Edwards for their help.

The research was funded by a strategic award from the Wellcome Trust (094585). CGF holds a Principal Research Fellowship from the Wellcome Trust (046386). The Wellcome Trust had no involvement with the design or conduct of the study or the preparation of this manuscript.

Conflicts of Interest

CGF receives royalties from the treatment guide.

Multimedia Appendix 1

Overview of enhanced cognitive behavior therapy for eating disorders.

[PDF File (Adobe PDF File), 93KB - jmir_v19i6e214_app1.pdf]

Multimedia Appendix 2

The training program.

[PDF File (Adobe PDF File), 36KB - jmir_v19i6e214_app2.pdf]

Multimedia Appendix 3

Web-centred training.

[PNG File, 84KB - jmir_v19i6e214_app3.PNG]

References

1. Fairburn CG, Patel V. The global dissemination of psychological treatments: a road map for research and practice. Am J Psychiatry 2014 May;171(5):495-498. [doi: 10.1176/appi.ajp.2013.13111546] [Medline: 24788281]
2. Kazdin A. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
3. Herschell A, Kolko D, Baumann B, Davis A. The role of therapist training in the implementation of psychosocial treatments: a review and critique with recommendations. Clin Psychol Rev 2010 Jun;30(4):448-466 [FREE Full text] [doi: 10.1016/j.cpr.2010.02.005] [Medline: 20304542]
4. Fairburn CG, Cooper Z. Therapist competence, therapy quality, and therapist training. Behav Res Ther 2011 Jun;49(6-7):373-378 [FREE Full text] [doi: 10.1016/j.brat.2011.03.005] [Medline: 21492829]
5. Rose G. Rose's Strategy of Preventive Medicine: The Complete Original Text. Oxford: Oxford University Press; 2008.
6. Fairburn CG, Cooper Z, Shafran R. Cognitive behaviour therapy for eating disorders: a transdiagnostic theory and treatment. Behav Res Ther 2003 May;41(5):509-528. [Medline: 12711261]
7. Fairburn CG. Cognitive Behavior Therapy and Eating Disorders. New York: The Guilford Press; 2008.
8. National Institute for Health and Care Excellence. Eating Disorders: Recognition and Treatment. London: National Institute for Health and Care Excellence; 2017.
9. Cooper Z, Doll H, Bailey-Straebler S, Kluczniok D, Murphy R, O'Connor ME, et al. The development of an online measure of therapist competence. Behav Res Ther 2015 Jan;64:43-48 [FREE Full text] [doi: 10.1016/j.brat.2014.11.007] [Medline: 25528502]
10. Rosen RC, Ruzek JI, Karlin BE. Evidence-based training in the era of evidence-based practice: challenges and opportunities for training of PTSD providers. Behav Res Ther 2017 Jan;88:37-48. [doi: 10.1016/j.brat.2016.07.009] [Medline: 28110675]
11. McHugh RK, Barlow DH. In: Barlow D, editor. Dissemination and Implementation of Evidence-Based Psychological Interventions. New York: Oxford University Press; 2012.
12. McHugh R, Barlow D. Training in evidence-based psychological interventions. In: McHugh RK, Barlow DH, editors. Dissemination and Implementation of Evidence-Based Psychological Interventions. New York: Oxford University Press; 2012:43-58.

Abbreviations

CBT-E: Enhanced cognitive behaviour therapy for eating disorders
CREDO: Centre for Research on Eating Disorders at Oxford
HSE: Health Service Executive
NICE: National Institute for Health and Care Excellence

Edited by G Eysenbach; submitted 13.04.17; peer-reviewed by J Apolinário-Hagen, G Waller; comments to author 04.05.17; revised version received 22.05.17; accepted 26.05.17; published 16.06.17.

Please cite as:
Fairburn CG, Allen E, Bailey-Straebler S, O'Connor ME, Cooper Z
Scaling Up Psychological Treatments: A Countrywide Test of the Online Training of Therapists
J Med Internet Res 2017;19(6):e214
URL: http://www.jmir.org/2017/6/e214/
doi:10.2196/jmir.7864
PMID:28623184

Original Paper

Using Social Media to Characterize Public Sentiment Toward Medical Interventions Commonly Used for Cancer Screening: An Observational Study

Omar Metwally[1,2], BS, MD; Seth Blumberg[3,4], MD, PhD; Uri Ladabaum[5], MD, MS; Sidhartha R Sinha[5], MD

[1]Department of Clinical Informatics, University of California, San Francisco, San Francisco, CA, United States
[2]Department of Internal Medicine, Highland General Hospital, Oakland, CA, United States
[3]Francis I. Proctor Foundation, University of California, San Francisco, San Francisco, CA, United States
[4]St Mary's Medical Center, San Francisco, CA, United States
[5]School of Medicine, Division of Gastroenterology, Stanford University, Stanford, CA, United States

Corresponding Author:
Sidhartha R Sinha, MD
School of Medicine, Division of Gastroenterology
Stanford University
Alway Building, Room M211
300 Pasteur Drive
Stanford, CA, 94305-5187
United States
Phone: 1 650 723 6661
Fax: 1 650 723 5488
Email: sidsinha@stanford.edu

Abstract

Background: Although cancer screening reduces morbidity and mortality, millions of people worldwide remain unscreened. Social media provide a unique platform to understand public sentiment toward tools that are commonly used for cancer screening.

Objective: The objective of our study was to examine public sentiment toward colonoscopy, mammography, and Pap smear and how this sentiment spreads by analyzing discourse on Twitter.

Methods: In this observational study, we classified 32,847 tweets (online postings on Twitter) related to colonoscopy, mammography, or Pap smears using a naive Bayes algorithm as containing positive, negative, or neutral sentiment. Additionally, we characterized the spread of sentiment on Twitter using an established model to study contagion.

Results: Colonoscopy-related tweets were more likely to express negative than positive sentiment (negative to positive ratio 1.65, 95% CI 1.51-1.80, *P*<.001), in contrast to the more positive sentiment expressed regarding mammography (negative to positive ratio 0.43, 95% CI 0.39-0.47, *P*<.001). The proportions of negative versus positive tweets about Pap smear were not significantly different (negative to positive ratio 0.95, 95% CI 0.87-1.04, *P*=.18). Positive and negative tweets tended to share lexical features across screening modalities. Positive tweets expressed resonance with the benefits of early detection. Fear and pain were the principal lexical features seen in negative tweets. Negative sentiment for colonoscopy and mammography spread more than positive sentiment; no correlation with sentiment and spread was seen for Pap smear.

Conclusions: Analysis of social media data provides a unique, quantitative framework to better understand the public's perception of medical interventions that are commonly used for cancer screening. Given the growing use of social media, public health interventions to improve cancer screening should use the health perceptions of the population as expressed in social network postings about tests that are frequently used for cancer screening, as well as other people they may influence with such postings.

(J Med Internet Res 2017;19(6):e200) doi:10.2196/jmir.7485

KEYWORDS

Twitter; sentiment analysis; cancer screening; colonoscopy; mammography; Pap smear; Papanicolaou test; social media; early detection of cancer

Introduction

The US Preventive Services Task Force and numerous professional societies endorse colonoscopy, mammography, and Pap smear as effective screening modalities for colon, breast, and cervical cancer, respectively. Over 350,000 cases of these cancers are diagnosed yearly in the United States [1-5]. Despite the effectiveness of these and other screening modalities in reducing cancer-related mortality, millions of Americans do not undergo screening [6-13]. The reasons for this lack of adherence, particularly for these 3 tests, are multifold. Colonoscopy, mammography, and Pap smear are generally considered more invasive or involved than exclusively laboratory-based screening tests, such as those for high cholesterol or diabetes. Briefly, colonoscopy generally involves visualization via a flexible endoscope inserted into the rectum, and often biopsy, of colonic mucosa. To increase the chances of complete visualization of colonic mucosa, patients are typically advised to adhere to a restricted diet with avoidance of solid food the day prior to the procedure and are frequently required to drink large volumes of bowel-cleansing solutions that result in frequent bowel movements [14]. Mammography involves radiographic imaging after compression of the breast tissue, a process that can be uncomfortable for many women [15]. Pap smear entails insertion of a speculum into the vagina and use of a brush to scrape a sample of cervical cells, which allows the operator to examine cells microscopically for malignant and premalignant changes, but it can also be associated with pain and anxiety [16].

Patient questionnaires have identified discomfort, embarrassment, and various socioeconomic factors as barriers to participation in cancer screening [4,17]. However, studying patient perceptions of modalities commonly used for cancer screening using formal surveys is limited by several factors. First, these surveys can be costly to administer and do not provide real-time actionable information [18]. Second, monitoring the spread and changes in sentiment over time is limited by cost and diminishing response rates. Third, surveys ask specific questions and typically provide limited possible responses, which qualitatively and quantitatively limits the range of data generated by these interventions [19]. Infodemiology, which includes exploration of the distribution and determinants of information on the Internet to improve public health, provides an alternative method to study societal perceptions of health care, such as their sentiment regarding commonly used cancer screening interventions [20]. Infodemiologic studies have investigated numerous aspects of health, including attitudes toward and spread of illness as expressed on social media, search engines, and blogs; sentiment in chronic diseases; and the effectiveness of smoking cessation campaigns [20-24]. Analysis of social media postings offers a unique opportunity to overcome the limitations of conventional surveys and to understand core health care issues, such as why screening recommendations are often not followed. Among such networks, Twitter is relatively unique in that vast amounts of data are publicly available. Revealing differences in sentiment on social media toward various tools commonly used for screening and analyzing how interventions to improve screening affect perception may lead to understanding how screening adherence can be increased.

In this study, we applied established methods in sentiment analysis and machine learning to Twitter data to characterize sentiment toward common interventions used for cancer screening. Similar methods have been applied to characterize patient attitudes toward various medical topics, including vaccines, illness, pain, and drug use [19,25-29]. We also quantified the way in which sentiment regarding interventions commonly used for cancer screening spreads on social networks, offering a unique opportunity to both understand health-related discourse propagation and gain insight on how to engineer outreach efforts more effectively.

Methods

Recruitment

We used the Twitter (Twitter, Inc) search application programming interface (API) to collect over 30,000 English-language tweets relating to colonoscopy (10,262), mammography (12,002), and Pap smear (10,583) [12]. All tweets were collected on consecutive days over a 6-week period from December 2015 to January 2016. Colonoscopy-related tweets were identified by querying for the term “colonoscopy.” Mammography-related tweets were identified by querying for the term “mammogram” or “mammography.” Pap smear-related tweets were identified by querying for “pap smear,” “pap test,” “Papanicolaou test,” “Papanicolaou AND screening,” “pap AND cervical cancer,” “pap AND pelvic exam,” or “pap AND HPV.” We obtained the data set by writing code to manually collect these publicly available tweets through the Twitter API, which is a sampling of up to 1% of the total number of tweets at any time (ie, the Twitter Firehose) [30-32]. The question of how representative the Twitter API is of the Firehose has been rigorously studied, and the limitations are discussed in the “Limitations” section below. Briefly, the quantity and quality of tweets delivered via the API depends on the keywords used to query the tweets, the user IDs specified, and geographic tags (if present). The Twitter API begins sampling using an undisclosed method once the 1% tweet threshold is reached. Following the precedent set by other Twitter-based studies, and according to our institutional review board’s recommendations to exempt this study from review, we did not obtain consent from Twitter users, since we used the data in aggregate, and these data are publicly available. For each tweet, we recorded the content of the message and the number of retweets (how many times the tweet was propagated by other users).

Sentiment Classification

We classified sentiment in 2 separate steps as described in the literature [33,34]. First, 1500 tweets (500 for each screening modality) were manually labeled by an investigator as containing positive, negative, or neutral sentiment and were used to train the classifiers. For each screening modality, we trained a naive Bayes classifier, a classification algorithm in which training is based on prior probabilities with different variables assumed to be independent of one another a priori, to categorize all tweets as carrying either positive or nonpositive sentiment [35]. Then, we trained a second naive Bayes classifier to categorize all tweets as carrying either a negative sentiment or a nonnegative sentiment. Tweets that were positive and nonnegative were ultimately labeled as positive. Tweets that were negative and nonpositive were ultimately labeled as negative. Tweets that were classified as both nonpositive and nonnegative were grouped into a third, neutral category. No tweets were classified as both positive and negative.

Validity

To assess consistency in labeling, a random subset of the 1500 tweets were relabeled by the same investigator, with 96% concordance. To establish validity of our labeling system, a second investigator independently labeled tweets, with an interobserver concordance of 95%. We characterized the accuracy of the classification algorithm by using 1200 of the labeled tweets as a training set and the remaining 300 as a testing set [36]. Our decision to use 75% of labeled tweets for training and 25% of labeled tweets for testing is consistent with validity assessment common in the machine learning literature; 20% to 33% of a labeled set is commonly used for testing purposes, with the remaining dataset used for training [36]. We inferred the true proportion of positive and negative tweets via a 2-step bootstrap method [37]. The first step of the bootstrap sampled individual classifications from the observed data with replacement. The second step labeled the bootstrapped classified data as positive, negative, or neutral based on contingency tables (Multimedia Appendix 1). To compute 2-sided *P* values for the ratio of negative to positive sentiment, we applied a 1-step sample with replacement bootstrap to a null dataset that had the same number of observed neutral tweets, but an equal proportion of positive and negative tweets. The total size of the null dataset matched the observed data, and we assumed classification of the null datasets to be 100% accurate.

Dissemination of Sentiment

We analyzed word frequency in all original tweets for the most common words in positive and negative tweets. Demographic information about Twitter users was obtained from Demographics Pro (Demographics Pro Inc), a third-party tool providing inferred predictive analytics on demographic information about social media users with 95% or greater confidence based on multiple data sources [26,38,39].

To assess the likelihood of a tweet to be propagated (ie, retweeted), we employed established concepts from the spread of infectious disease [40]. The effective reproduction number equals the expected secondary cases resulting from exposure to an infected individual [41,42]. Analogously, we defined the rate of propagation as the mean number of times a message is retweeted by a Twitter user. To account for heterogeneity of retweeting, we inferred rate of propagation by assigning a negative binomial distribution for the number of retweets each tweet generated. We determined the statistical difference of rate of propagation by Akaike information criterion score [41] and calculated corresponding *P* values by chi-square modeling of the log likelihood ratio. *P* values for the incidence of new tweets were determined based on assuming an underlying Poisson distribution for the introduction of new tweets.

Results

Classifier Performance

Our classifier labeled tweet sentiment with an accuracy of about 80%. Importantly, no negatively classified tweets were manually labeled as positive, and only 4% of the positively classified tweets were manually labeled as negative (Multimedia Appendix 1). The misclassifications were predominantly for tweets with nonneutral sentiment classified as being neutral or for tweets with neutral sentiments being classified as nonneutral. As such, the overwhelming majority of misclassified tweets did not entail complete reversal of sentiment. One example of a tweet with neutral sentiment being classified as nonneutral (in this particular case, as negative) is “Worried about preparing for a colonoscopy? Don’t. The preparation can be inconvenient, but it is not difficult or painful.” Since we were using a naive Bayes classification algorithm, the most likely explanation for misclassification of this tweet is the presence of words with negative connotations, such as “painful” and “inconvenient.” Similarly, this nonneutral (in this case, negative) tweet was incorrectly classified as neutral: “cant afford doctor just go to the airport. You get a free xray and breast exam. And if you mention Al Qaeda and you get a free colonoscopy.”

Differences in Sentiment Among Screening Modalities

When adjusted for imperfections in classification, colonoscopy-related tweets were estimated to be 1.65 (95% CI 1.51-1.80, *P*<.001) times more likely to express negative sentiment than positive sentiment (Figure 1 and Multimedia Appendix 2). In contrast, mammography-related tweets were 0.43 (95% CI 0.39-0.47, *P*<.001) times more likely to be negative than positive. The proportions of positive versus negative sentiment in Pap smear-related tweets were not significantly different (negative to positive ratio 0.95, 95% CI 0.87-1.04, *P*=.18). The majority of tweets in all screening modalities were neutral.

Demographic Analysis

Table 1 provides aggregate sex and age information about Twitter users discussing each screening modality. A large proportion of Twitter users discussing an intervention commonly used for cancer screening were less than 45 years of age, generally younger than those who commonly pursue routine colon cancer screening (typically starting at age 50 years). In contrast to the demographics of the entire Twitter network, which is characterized by roughly equal proportions of male and female users, more male users commented on colonoscopy and, not surprisingly, more female users commented on mammography and Pap smear [43]. Interestingly, Twitter users commenting on colonoscopy, mammography, and Pap smear were younger than the average Twitter user.

Figure 1. Sentiment expressed on Twitter regarding colonoscopy, mammography, and Pap smear (based on classification of over 30,000 tweets). A naive Bayes classifier was trained on labeled data and used to classify tweets relating to either colonoscopy, mammography, or Pap smear. Tests of statistical significance were undertaken using a bootstrap method with negative to positive sentiment ratio for colonoscopy (1.65, *P*<.001), mammography (0.43, *P*<.001), and Pap smear (0.95, *P*=.18). The full results for the bootstrap analysis are shown in Multimedia Appendix 2.

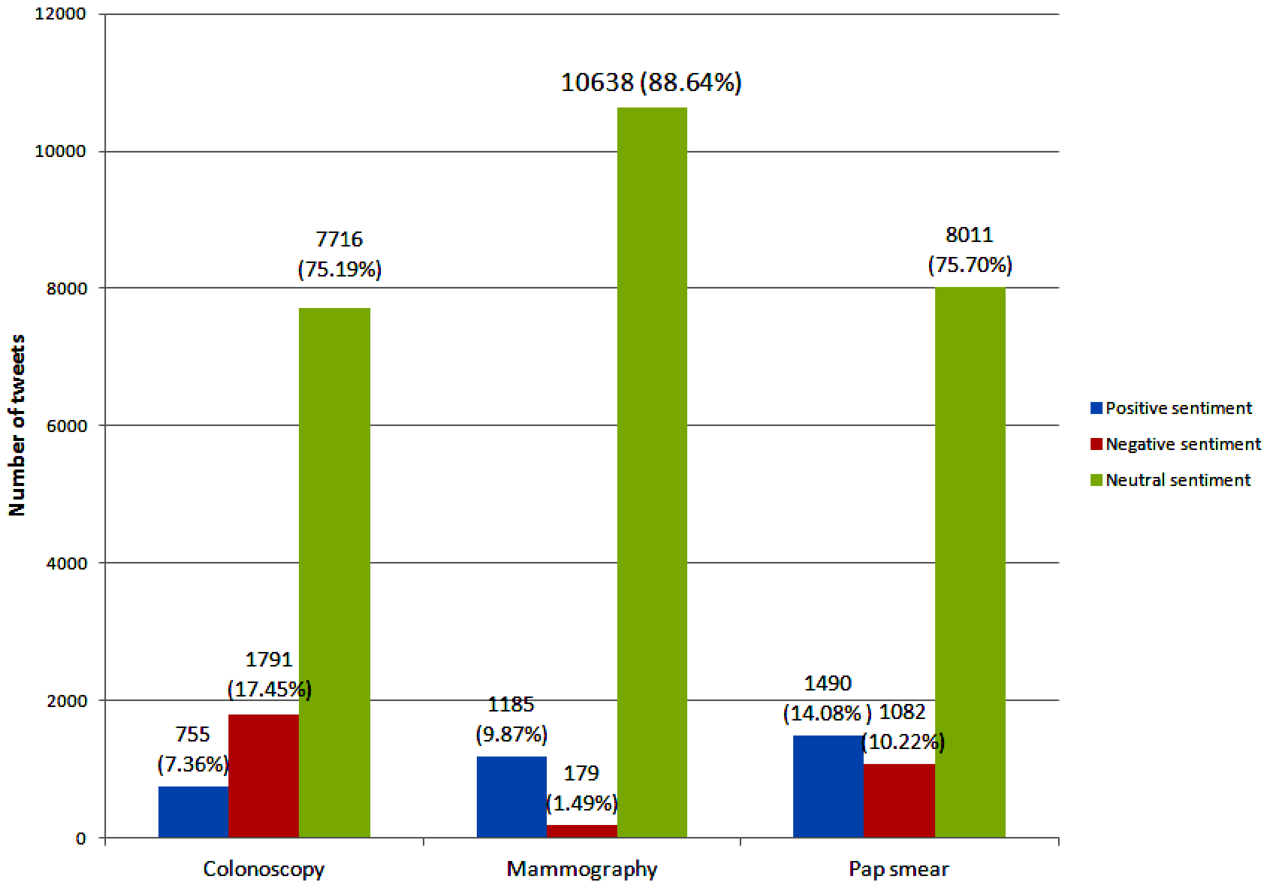


Table 1. Demographics[a] of users tweeting about cancer screening by screening modality (N=32,847).

Characteristics	Colonoscopy	Mammography	Pap smear
Sex			
Male	56.7%	36.4%	33.3%
Female	43.3%	63.6%	66.7%
Age group (years)			
≤20	18.4%	10.9%	24.3%
21-29	30.8%	20.9%	34.3%
30-34	15.9%	14.7%	11.8%
35-44	18.8%	30.7%	18.6%
45-54	10.1%	15.5%	7.8%
55-64	4.3%	5.5%	2.2%
≥65	1.7%	1.5%	1.1%

[a] Percentage data obtained from Demographics Pro.

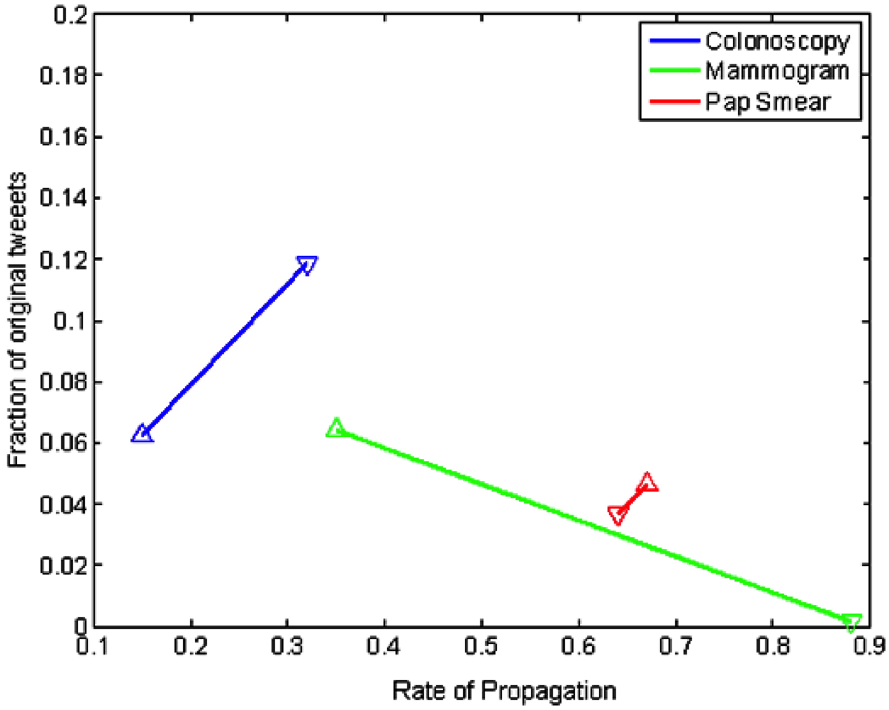
Word Frequency Analysis

Word frequency analysis for all 3 screening modalities demonstrated similarly perceived benefits of tools frequently used for cancer screening (Multimedia Appendix 3). Word frequency analysis showed that positively charged tweets most frequently contained words such as “health,” “awareness,” “screening,” “detection,” and “recommend.” Negatively charged tweets most frequently contained words related to pain (“painful,” “hurts”), anxiety associated with the procedure (“worried”), and issues with procedure preparation (“dehydration,” “preparing”). Examination of individual tweets showed that positive sentiment was more likely to be expressed when providing information about a procedure or explaining the benefits of cancer screening (Table 2). Positive tweets might also provide a "call to action" to encourage friends or family to get screened. Negative tweets often expressed physical discomfort related to a screening modality and compared cancer screening to anxiety-provoking tasks or suggested a level of social inappropriateness with the topic. Tweets with negative sentiment often included sarcastic humor.

Table 2. Examples of positive and negative tweets.

Type of tweet and modality	Positive sentiment	Negative sentiment	Neutral sentiment
Information on the procedure (positive) or fear of pain or harm (negative)			
Colonoscopy	Worried about preparing for a colonoscopy? Don't. The preparation can be inconvenient, but it is not difficult or painful.	Getting a colonoscopy can cause the patient to explode (methane+oxygen+electrical spark).	Colonoscopies are not just simple, harmless tests. Here are the pros and cons to consider.
Mammography	For women with dense breasts, ultrasound could help diagnose breast cancer.	I don't think my breast have ever been so smashed and squeezed. And here I thought a mammogram would be like taking an X-ray--NOPE! Corsets = struggling to breathe while getting a constant mammogram.	Is This Why Mammogram Recommendations Have Changed?
Pap smear	Getting a #Paptest is one of the best things you can do to prevent #cervicalcancer.	So my mom never had a pap smear until after she had me, when she was 37. I grew up hearing about how horrific it was. It really hurt her.	For me, a Pap Test just ended up being an unexpected trigger. For various reasons, some of which I will never know.
Benefits of cancer screening (positive) or "I'd rather" tweets (negative)			
Colonoscopy	LOVE your Parents enough to take them in for a Colonoscopy! It could just save their lives! Studies show that the colon cancer death rate was cut by more than half in those who had a colonoscopy. Has your dad turned 50 yet? If so, bug him into getting a colonoscopy. You could be saving his life.	Things I would rather do than my exams: re-organise a forever 21 store, eat my own vomit, peel 4,000 potatoes with a spork, colonoscopy prep I can think of a few better places: Gates of Hell, during a colonoscopy, Mordor, a Joel Osteen Conference...	
Mammography	Annual mammography in women 40 to 48 y of age reduced breast cancer mortality.	i'd rather have a mammogram done while being awake as they remove my kidney	
Pap smear	The only way to find changes that may lead cervical cancer is by having a Pap. Screening saves lives!	I would rather give myself a pap smear in the middle of Macy's than read your Christmas newsletter, Aunt Karen	
Call to Action (positive) or Other (negative)			
Colonoscopy	Have a friend turning 50? Encourage them to get their colonoscopy; it could save their life.	i would hate to get a colonoscopy… My first colonoscopy will be done by a coroner at my autopsy Today a colleague told me that he's having a colonoscopy this week. I need a new job.	
Mammography	Ladies get that mammogram because it saved my little sister from a very aggressive breast cancer. Make it a XMas present to yourself.	How a mammogram actually causes breast cancer.	
Pap smear	I went for my first ever Pap Test today *feeling proud & brave*. Thanks to all the lesbian women who urged/reminded me to go! Hello ladies schedule your mammogram today. Include health in your new year's resolution.	Your mothers so dumb she went to Dr. Dre for a pap smear	

Figure 2. The fraction of original tweets and the rate of propagation for each modality. Upward (downward) pointing triangles represent positive (negative) tweets. Lines are for visualization purposes only. The difference in rate of propagation between positive and negative tweets was significant for colonoscopy (*P*=.001) and mammography (*P*=.02) but not for Pap smear (*P*=.83).



Propagation of Sentiment

The proportion of tweets with positive versus negative sentiment is dependent on both the frequency of new tweets and the rate at which these new tweets are retweeted (Figure 2 and Multimedia Appendix 4) [44]. Comparison of positively versus negatively classified tweets showed that, for colonoscopy, negative sentiment both was more frequent in original tweets and spread more than positive ones. New tweets regarding mammography were typically positive rather than negative. However, the few negative-sentiment tweets toward mammography spread much more than those with positive sentiment. For Pap smear, negative and positive tweets had similar characteristics with regard to their spread and the frequency of new tweets.

Discussion

Principal Results

By using automated sentiment classifiers, we were able to analyze much larger bodies of data than in typical patient survey-based studies. Understanding basic differences in sentiment between interventions commonly used for screening, such as the greater prevalence of negative sentiment regarding colonoscopy compared with mammography, could lead to more targeted, effective interventions, as well as the real-time means to assess the effects of such interventions. Such comparisons could, for example, foster organizations promoting screening to learn from each other to more effectively maintain social media interventions to promote positive sentiment for these lifesaving medical interventions. Beyond sentiment analysis, word frequency analyses can provide quantitative as well as qualitative insight into potential reasons for differences in sentiment and can identify areas on which to focus education efforts. For instance, pain and fear were common themes in negative tweets about all 3 modalities, findings that have been echoed, at times inconsistently, by formal patient surveys [17].

Comparison With Prior Work

We employed novel machine learning algorithms to understand sentiment on social media regarding tools commonly used for screening. Understanding opinions regarding these changes by analyzing social media could be valuable in assessing health policy changes and implementing new policies. With respect to cancer screening, public reactions to changes in screening recommendations from professional societies could also be monitored, and false perceptions could be addressed immediately.

Limitations

Despite the large number of public tweets available for analysis, this is an imperfect representation of the population at large. The age of Twitter users sampled is generally younger than the target screening population, who commonly use the medical tests investigated in this study, limiting the generalizability of our results to older populations. However, this limitation also provides useful insight. For example, one hypothesis to explain the negative sentiments regarding colonoscopy in younger users is that some of these persons may have conditions such as inflammatory bowel disease and require invasive, potentially embarrassing interventions that their peers do not require. It is unclear how younger persons’ sentiments toward colonoscopy might affect the sentiments of older persons, including the relatives of the tweet authors, who may be eligible for age-appropriate colon cancer screening. While references to colonoscopy on Twitter may not entirely reflect screening and surveillance, these are the most common indications for colonoscopy [13]. The same is true for mammography and Pap smear. Yet even understanding perceptions of a tool *commonly* used for screening is telling. This information still represents public opinion, albeit a younger population, and understanding the perceptions of this younger population may influence future screening decisions as well. Influencing younger users may in fact be a strategy to further improve cancer screening adherence. The relatively small subgroup sample sizes limited the demographic analysis possible through Demographics Pro. Additionally, potentially vulnerable groups, including minorities at risk for poor preventive health use, may not be represented. Nonetheless, we believe that public expressions of sentiment provide insight that may not necessarily be reflected by formal surveys into how the screening modalities that we studied are perceived by the public. We chose to capture whether tweets mentioning tools commonly used for cancer screening were generally positive or negative versus attempting to discern whether the sentiment was expressed specifically regarding the actual screening procedure itself. More nuanced sentiment analysis methods capable of discerning meaning by analyzing sentences as aggregates of phrases and their modifiers may improve our understanding of public discourse specifically related to cancer screening [45]. We believe that even capturing such nonscreening-related mentions of these interventions provides valuable insight into public opinion for these tools used by millions to improve health.

How well the Twitter API samples the total corpus of tweets (the Twitter Firehose) has been studied by Morstatter et al and has yielded heterogeneous results [32]. The API’s sampling is imperfect and depends to a large extent on the type of analysis undertaken. For example, those authors found that the 1% sampling becomes substantially more representative when tweets are collected over consecutive days, as was the case in our study. The quality of the API’s sampling decreases when the number of hashtags or query keywords decreases, which would theoretically affect the quality of sampled colonoscopy tweets (where “colonoscopy” is the only queried term) more than Pap smear-related tweets (which we sampled using 10 unique keywords). Correcting for this sampling bias is difficult given that Twitter does not disclose how sampling is performed, but it should be acknowledged in infodemiologic studies that use the Twitter search API.

Conclusion

We have analyzed tweets about interventions commonly used for cancer screening to assess public sentiment about these interventions. There were substantially more negative than positive tweets about colonoscopy, but not mammography or Pap smear. Tweet propagation in the social network was greater for negative than for positive tweets about colonoscopy and mammography, suggesting a possible disproportionate impact of negative sentiment for these screening tests. Examination of large data sets available from the Twitter social network using automated algorithms provides an opportunity to examine public attitudes toward cancer screening and other health care interventions that might lead to policy changes, novel programs, and more refined counseling guidelines that improve public attitudes and health-related behaviors.

Acknowledgments

Seth Blumberg was supported by the RAPIDD Program of the Science and Technology Directorate, US Department of Homeland Security (Research and Policy for Infectious Disease Dynamics). Sidhartha Sinha was supported by a KL2 Mentored Career Development Award of the Stanford Clinical and Translational Science Award to Spectrum (National Institutes of Health KL2 TR 001083) and the Division of Gastroenterology and Hepatology at Stanford University School of Medicine, Stanford, CA, USA. We are grateful to Demographics Pro for providing us with demographic information. The Python code we used to implement the naive Bayes classification algorithm was built on code from the public domain, and the authors are acknowledged in the source code we provide [46].

Authors' Contributions

OM contributed to study design, conceptualization, methodology, software, validation, formal analysis, data curation, writing of the original draft, review, and editing, and visualization. SB contributed to study design, methodology, software, validation, formal analysis, data curation, writing review and editing, and visualization. UL contributed to study design, writing review and editing, and formal analysis. SRS contributed to study design, conceptualization, validation, formal analysis, data curation, writing of the original draft, review, and editing, and visualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Contingency tables for labeling bootstrapped classified data.

[PDF File (Adobe PDF File), 26KB - jmir_v19i6e200_app1.pdf]

Multimedia Appendix 2

Tweet sentiment classification with cancer screen modality.

[PDF File (Adobe PDF File), 27KB - jmir_v19i6e200_app2.pdf]

Multimedia Appendix 3

Word frequency analysis of common themes.

[PDF File (Adobe PDF File), 24KB - jmir_v19i6e200_app3.pdf]

Multimedia Appendix 4

Rate of propagation for positive and negative tweets.

[PDF File (Adobe PDF File), 59KB - jmir_v19i6e200_app4.pdf]

References

1. Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med 2013 Sep 19;369(12):1095-1105 [FREE Full text] [doi: 10.1056/NEJMoa1301969] [Medline: 24047059]
2. Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, American Cancer Society Colorectal Cancer Advisory Group, US Multi-Society Task Force, American College of Radiology Colon Cancer Committee. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology 2008 May;134(5):1570-1595. [doi: 10.1053/j.gastro.2008.02.002] [Medline: 18384785]
3. Inadomi JM, Vijan S, Janz NK, Fagerlin A, Thomas JP, Lin YV, et al. Adherence to colorectal cancer screening: a randomized clinical trial of competing strategies. Arch Intern Med 2012 Apr 9;172(7):575-582 [FREE Full text] [doi: 10.1001/archinternmed.2012.332] [Medline: 22493463]
4. Wee CC, McCarthy EP, Phillips RS. Factors associated with colon cancer screening: the role of patient factors and physician counseling. Prev Med 2005 Jul;41(1):23-29. [doi: 10.1016/j.ypmed.2004.11.004] [Medline: 15916989]
5. Centers for Disease Control and Prevention. United States cancer statistics: 1999-2012 incidence and mortality data. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2015. URL: http://www.cdc.gov/uscs [accessed 2015-12-14] [WebCite Cache ID 6peu3g5B1]
6. American Cancer Society. What are the key statistics about cervical cancer?. 2016. URL: https://www.cancer.org/cancer/cervical-cancer/about/key-statistics.html [accessed 2017-04-12] [WebCite Cache ID 6peuBzo3G]
7. U.S. Preventive Services Task Force. Final summary: cervical cancer: screening. Rockville, MD: U.S. Preventive Services Task Force; 2012. URL: http://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/cervical-cancer-screening [accessed 2016-02-08] [WebCite Cache ID 6peuKAWoA]
8. Surveillance, Epidemiology, and End Results (SEER) Program. Cancer stat facts: female breast cancer. Rockville, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute; 2017. URL: https://seer.cancer.gov/statfacts/html/breast.html [accessed 2017-05-29] [WebCite Cache ID 6qpBoDveX]
9. Paci E. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. J Med Screen 2012;19 Suppl 1:5-13. [doi: 10.1258/jms.2012.012077] [Medline: 22972806]
10. Ferrante JM, Chen P, Crabtree BF, Wartenberg D. Cancer screening in women: body mass index and adherence to physician recommendations. Am J Prev Med 2007 Jun;32(6):525-531 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.004] [Medline: 17533069]
11. Koong S, Yen AM, Chen TH. Efficacy and cost-effectiveness of nationwide cervical cancer screening in Taiwan. J Med Screen 2006;13 Suppl 1:S44-S47. [Medline: 17227642]
12. Arroyave AM, Penaranda EK, Lewis CL. Organizational change: a way to increase colon, breast and cervical cancer screening in primary care practices. J Community Health 2011 Apr;36(2):281-288. [doi: 10.1007/s10900-010-9309-7] [Medline: 20835777]
13. Ladabaum U, Levin Z, Mannalithara A, Brill JV, Bundorf MK. Colorectal testing utilization and payments in a large cohort of commercially insured US adults. Am J Gastroenterol 2014 Oct;109(10):1513-1525. [doi: 10.1038/ajg.2014.64] [Medline: 24980877]
14. Inadomi JM. Screening for colorectal neoplasia. N Engl J Med 2017 Dec 20;376(16):1599-1600. [doi: 10.1056/NEJMc1702535] [Medline: 28423294]
15. Domar AD, Eyvazzadeh A, Allen S, Roman K, Wolf R, Orav J, et al. Relaxation techniques for reducing pain and anxiety during screening mammography. AJR Am J Roentgenol 2005 Feb;184(2):445-447. [doi: 10.2214/ajr.184.2.01840445] [Medline: 15671361]

16. Basen-Engquist K, Shinn EH, Warneke C, de Moor C, Le T, Richards-Kortum R, et al. Patient distress and satisfaction with optical spectroscopy in cervical dysplasia detection. Am J Obstet Gynecol 2003 Oct;189(4):1136-1142. [Medline: 14586367]
17. Ghanouni A, Plumb A, Hewitson P, Nickerson C, Rees CJ, von Wagner C. Patients' experience of colonoscopy in the English Bowel Cancer Screening Programme. Endoscopy 2016 Mar;48(3):232-240. [doi: 10.1055/s-0042-100613] [Medline: 26841268]
18. Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. Qual Manag Health Care 2012;21(1):9-19. [doi: 10.1097/QMH.0b013e3182417fc4] [Medline: 22207014]
19. 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]
20. 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]
21. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [Medline: 12517369]
22. Eysenbach G. Websites on screening for breast cancer: “infodemiology” studies have surely had their day. BMJ 2004 Mar 27;328(7442):769; author reply 769-769; author reply 770 [FREE Full text] [doi: 10.1136/bmj.328.7442.769-b] [Medline: 15044306]
23. Delir HP, Kang Y, Buchbinder R, Burstein F, Whittle S. Investigating subjective experience and the influence of weather among individuals with fibromyalgia: a content analysis of Twitter. JMIR Public Health Surveill 2017 Jan 19;3(1):e4 [FREE Full text] [doi: 10.2196/publichealth.6344] [Medline: 28104577]
24. 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]
25. 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 [FREE Full text] [doi: 10.2196/jmir.3769] [Medline: 25843553]
26. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145. [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
27. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. “When ‘bad’ is ‘good’”: identifying personal communication and sentiment in drug-related tweets. JMIR Public Health Surveill 2016 Oct 24;2(2):e162 [FREE Full text] [doi: 10.2196/publichealth.6327] [Medline: 27777215]
28. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
29. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of menthol cigarettes among Twitter users: content and sentiment analysis. J Med Internet Res 2017 Feb 27;19(2):e56 [FREE Full text] [doi: 10.2196/jmir.5694] [Medline: 28242592]
30. Twitter developers documentation: the search API. San Francisco, CA: Twitter Inc; 2017. URL: https://dev.twitter.com/rest/public/search [accessed 2015-02-17] [WebCite Cache ID 6qJGWK7Zu]
31. Gnip. Boulder, CO: Gnip, Inc; 2016. URL: https://gnip.com/ [accessed 2017-04-12] [WebCite Cache ID 6peuTPZB7]
32. Morstatter F, Pfeffer J, Liu H, Carley K. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. Palo Alto, CA: AAAI Press; 2013 Presented at: 7th International AAAI Conference on Weblogs and Social Media; July 8–11, 2013; Cambridge, MA, USA p. 400-408 URL: https://arxiv.org/pdf/1306.5204.pdf
33. Natural language toolkit.: NLTK Project; 2017 Jan 02. URL: http://www.nltk.org/ [accessed 2017-04-12] [WebCite Cache ID 6peuVeLOE]
34. Ranco G, Aleksovski D, Caldarelli G, Grčar M, Mozetič I. The effects of Twitter sentiment on stock price returns. PLoS One 2015;10(9):e0138441 [FREE Full text] [doi: 10.1371/journal.pone.0138441] [Medline: 26390434]
35. Manning C, Raghavan P, Schuze H. Introduction to Information Retrieval. Cambridge, UK: Cambridge University Press; 2008.
36. Pagolu V, Challa K, Panda G, Majhi B. Sentiment analysis of Twitter data for predicting stock market movements. 2016 Oct 03 Presented at: International Conference on Signal Processing, Communication, Power and Embedded Systems. SCOPES 2016; Oct 3-5, 2016; Paralakhemundi, Odisha, India p. 3-5 URL: https://arxiv.org/pdf/1610.09225.pdf
37. Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979 Jan;7(1):1-26. [doi: 10.1214/aos/1176344552]
38. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. J Med Internet Res 2014;16(6):e157 [FREE Full text] [doi: 10.2196/jmir.3247] [Medline: 24974893]
39. Understand and grow your social audience. Carrboro, NC: Demographics Pro Inc URL: http://www.demographicspro.com/ [accessed 2017-04-12] [WebCite Cache ID 6peuZYwPn]

40. Corley CD, Mihalcea R, Mikler AR, Sanfilippo AP. Predicting individual affect of health interventions to reduce HPV prevalence. Adv Exp Med Biol 2011;696:181-190. [doi: 10.1007/978-1-4419-7046-6_18] [Medline: 21431558]
41. Blumberg S, Funk S, Pulliam JR. Detecting differential transmissibilities that affect the size of self-limited outbreaks. PLoS Pathog 2014 Oct;10(10):e1004452 [FREE Full text] [doi: 10.1371/journal.ppat.1004452] [Medline: 25356657]
42. Blumberg S, Lloyd-Smith JO. Inference of R(0) and transmission heterogeneity from the size distribution of stuttering chains. PLoS Comput Biol 2013;9(5):e1002993 [FREE Full text] [doi: 10.1371/journal.pcbi.1002993] [Medline: 23658504]
43. Greenwood S, Perrin A, Duggan M. Social media update 2016. Washington, DC: Pew Research Center; 2016 Nov 11. URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [accessed 2017-04-12] [WebCite Cache ID 6peufALgU]
44. Blumberg S, Enanoria WT, Lloyd-Smith JO, Lietman TM, Porco TC. Identifying postelimination trends for the introduction and transmissibility of measles in the United States. Am J Epidemiol 2014 Jun 01;179(11):1375-1382 [FREE Full text] [doi: 10.1093/aje/kwu068] [Medline: 24786800]
45. Socher R, Perelygin A, Wu J, Chuang J, Manning C, Ng A, et al. Recursive deep models for semantic compositionality over a sentiment treebank. 2013 Presented at: Conference on Empirical Methods in Natural Language Processing; Oct 18-21, 2013; Seattle, WA, USA.
46. Metwally O. Bitbucket. Overview. Sydney, Australia: Atlassian; 2017. URL: https://bitbucket.org/osmode/cancer_screening/ [accessed 2017-06-02] [WebCite Cache ID 6qvESTouQ]

Abbreviations

API: application programming interface

Edited by G Eysenbach; submitted 13.02.17; peer-reviewed by J Zia, S Cutrona; comments to author 06.03.17; revised version received 11.04.17; accepted 23.05.17; published 07.06.17.

Please cite as:
Metwally O, Blumberg S, Ladabaum U, Sinha SR
Using Social Media to Characterize Public Sentiment Toward Medical Interventions Commonly Used for Cancer Screening: An Observational Study
J Med Internet Res 2017;19(6):e200
URL: http://www.jmir.org/2017/6/e200/
doi:10.2196/jmir.7485
PMID:28592395

Original Paper

Frequencies of Private Mentions and Sharing of Mammography and Breast Cancer Terms on Facebook: A Pilot Study

Marco Huesch[1], MBBS, PhD; Alison Chetlen[2], DO; Joel Segel[3,4], PhD; Susann Schetter[2], DO

[1]Milton S Hershey Medical Center, Department of Radiology, Public Health Sciences, Penn State College of Medicine, Hershey, PA, United States
[2]Milton S Hershey Medical Center, Department of Radiology, Breast Imaging Division, Penn State College of Medicine, Hershey, PA, United States
[3]Milton S Hershey Medical Center, Department of Health Policy and Administration, Pennsylvania State University, University Park, PA, United States
[4]Penn State Cancer Institute, Penn State University, Hershey, PA, United States

Corresponding Author:
Marco Huesch, MBBS, PhD
Milton S Hershey Medical Center
Department of Radiology, Public Health Sciences
Penn State College of Medicine
500 University Drive
Mailcode H-066
Hershey, PA, 17033
United States
Phone: 1 717 531 0003 ext 284649
Fax: 1 717 531 0002
Email: mhuesch@pennstatehealth.psu.edu

Abstract

Background: The most popular social networking site in the United States is Facebook, an online forum where circles of friends create, share, and interact with each other’s content in a nonpublic way.

Objective: Our objectives were to understand (1) the most commonly used terms and phrases relating to breast cancer screening, (2) the most commonly shared website links that other women interacted with, and (3) the most commonly shared website links, by age groups.

Methods: We used a novel proprietary tool from Facebook to analyze all of the more than 1.7 million unique interactions (comments on stories, reshares, and emoji reactions) and stories associated with breast cancer screening keywords that were generated by more than 1.1 million unique female Facebook users over the 1 month between November 15 and December 15, 2016. We report frequency distributions of the most popular shared Web content by age group and keywords.

Results: On average, each of 59,000 unique stories during the month was reshared 1.5 times, commented on nearly 8 times, and reacted to more than 20 times by other users. Posted stories were most often authored by women aged 45-54 years. Users shared, reshared, commented on, and reacted to website links predominantly to e-commerce sites (12,200/1.7 million, 36% of all the most popular links), celebrity news (n=8800, 26%), and major advocacy organizations (n=4900, 15%; almost all accounted for by the American Cancer Society breast cancer site).

Conclusions: On Facebook, women shared and reacted to links to commercial and informative websites regarding breast cancer and screening. This information could inform patient outreach regarding breast cancer screening, indirectly through better understanding of key issues, and directly through understanding avenues for paid messaging to women authoring and reacting to content in this space.

(J Med Internet Res 2017;19(6):e201) doi:10.2196/jmir.7508

KEYWORDS

Facebook; online social network; social media; breast cancer screening; mammography; user comments; websites; links

Introduction

Nearly 3 million women have a history of breast cancer today in the United States [1], and a further 15 million receive screening mammography annually [2]. Understanding how women interact with their support and social networks may be clinically important in breast cancer [3]. Breast cancer screening with imaging is widely recognized as lifesaving [4,5], yet still far too few women take advantage of this program. Widespread estimates that as many as 1 in 3 women remain unscreened or underscreened [6] suggest that more must be done to drive awareness, improve access, and increase screening.

Online social media and social networks potentially provide an opportunity for women to become aware, or more aware, of breast cancer risk and screening options and methods. Such novel channels can allow women to share intimate information regarding their symptoms, signs, screening, diagnosis, and treatment with close friends and relatives. In this study, we explored content relating to breast cancer screening on the leading US online social networking platform. Our approach has several key differentiators from past and current work.

First, we listened rather than reaching out and teaching or communicating. We sought to illustrate that researchers can use an online platform to listen to users in a way that respects their privacy and doesn't identify them or any of their actual text. This social experience can be viewed through the lens of social normative theory, recognizing that these online channels allow users to build relationships and potentially influence the attitudes and behaviors of connected others [7,8]. The information spread through social media, whether true or false, can affect the social norms of others for good or bad, making listening to such content important for understanding perceptions, awareness, and attitudes [9].

Yet most research using Facebook, including our own, has hinged on outreach instead of listening. While an online social network is designed to be a social experience for its users, commercial outreach by advertisers and researchers is simple and cost effective. Such outreach methods exploit the personal and intimate setting afforded by the network and its highly tailored ability to finely target users based on expressed and inferred interests. For example, we reached more than 50,000 white, Latino, and Hispanic American women with an interest in maternity care in Los Angeles in part through targeted Facebook advertisements [10]. Other researchers have seen success with similar Facebook-based outreach in the settings of mental health [11,12], tobacco use [13], and drug and alcohol use [14].

Second, Facebook is an intrinsically different platform from other online platforms. Recently, Rosenkrantz and colleagues provided an innovative and important look at how women perceive the mammography experience through examination of several hundred carefully selected tweets both before and after the screening [15]. Others have similarly examined Twitter [16], YouTube video comments [17], smartphone apps [18], and Google Trends [19]. However, these other platforms differ in use, beliefs, attitudes, experiences, typical audience, and context of use. Facebook allows its users to experience gratification from satisfying the need to belong and the need for self-presentation [20,21]. Facebook also differs from the more public platforms by allowing users to share content with a circle of connected users. This offers potentially an opportunity to listen to more nuanced and private, sensitive conversations. This aspect of nonpublic, sensitive information is similar to that revealed through private searches on Google, but differs from the usually public comments on a video on YouTube, tweets on Twitter, or a weblog (blog). Other differences that distinguish Facebook are that it may, unfortunately, also allow inaccurate information, myths, or undesirable social norms to spread, unlike more public communications such as tweets, in which such issues can be more quickly and easily identified [22].

Third, the scale of our data source exceeds those of other studies leveraging Facebook data. Some studies have examined the rate of engagement with sampled posted Facebook content on breast cancer screening [23,24], or relatively small samples of conversations about complementary medicine and breast cancer on Facebook [25], or within samples of Facebook groups specifically relating to breast cancer [26].

Yet there is a wide and deep penetration of Facebook in the United States. More than half of all American adults are users [27]. It is also the most demographically representative of all online social networks; of adult women who are online, 77% are users of Facebook [28]. The Pew Research Center also found the median number of Facebook friends to be around 200 [27]. This suggests that there are large numbers of connected users who can see, comment on, and react to content that their friends create. Since such content could have positive or negative public health effects, we contend that understanding what is being shared is critical.

We believe that online investigations are crucial to understanding women's experiences better, and to inform strategies that seek to deal with obstacles to improved utilization of screening. This pilot study is a cursory first step: an exploration of the terms and phrases used by female users on Facebook relating to breast cancer screening over a 1-month period. Our hypothesis was that adult women would be actively generating content and interacting with other users' content on Facebook on the topic of breast cancer screening. Our objectives were to understand (1) the most commonly used terms and phrases relating to breast cancer screening, (2) the most commonly shared website links that other women interacted with, and (3) the most commonly shared website links, by age groups.

Methods

We contracted with Sysomos Scout (Sysomos, Toronto, ON), a commercial infomediary that resells Twitter, Facebook, blog, and other social media data [10]. We provided a list of 69 keywords and key phrases (Textbox 1) to Sysomos, looking only at Facebook data generated by self-identified female users of Facebook, covering professions and organizations, formal and informal terms for services rendered relating to breast cancer and breast cancer screening, related symptoms and signs, risk strata, and investigation findings. A practicing breast imaging radiologist (AC) and a physician scientist with substantial prior experience using online infomediaries (MH) prespecified these keywords based on guidelines content, institutional patient education materials, and the bibliography of breast cancer screening literature of this study.

We controlled searches using the proprietary tool's user interface (Figure 1). Facebook and Sysomos make only 30-day rolling period data available, and we randomly initiated coverage on November 15, 2016, which ran through December 15, 2016.

Definition of Content

Sysomos matched these keywords to any Facebook *story* (a posted item of content by a user) or any type of Facebook *interaction* that can be a *reshare* (a reposting of an original story to another user connected to the sharing user), a *comment* (a text comment made by a connected user on the original story or on a prior comment), or a *reaction* (one of several emoji representing emotions, such as positive, negative, empathetic, surprise, and love).

Definition of Counts

All counts for numbers of stories and interactions are unique, by Facebook's construction of nonoverlapping categories of story, reshare, comment, and reaction. Counts of authors are more complex. Within a category, the number of authors is the unique number of authors. For example, if 45,000 women commented on an article, these are nonduplicated authors. Across categories this may not hold, as the same author may post several stories, comment on other stories, and react to many others.

Accordingly, we cannot add the numbers of authors across the different categories of interactions. For example, 1.1 million unique authors making reactions and the 0.4 million unique authors making comments cannot be added to obtain 1.5 million authors, because this resulting sum double counts women doing both. However, the actual total is no smaller than 1.1 million and no larger than 1.5 million. We conservatively report only the lower number and use phrases such as "…at least…" in reporting these totals.

Most Commonly Mentioned Terms

Sysomos reported to us summary aggregate statistics such as totals, time-based trends such as subtotal by day, content-based subtotals, keyword prevalence, other word prevalence in context of keyword, and most popular website links that were posted or shared. Importantly, Facebook explicitly limits some aggregate data to just the top 10 items within a category and limits all aggregate data to items with at least 100 instances. This is due to confidentiality concerns and the ability otherwise to potentially reidentify individuals. We provide selected excerpts of these data, including tabular and graphical summaries.

Most Commonly Shared Website Links by Interaction Type and by Age Group

In this pilot study, we were most interested in the type of content that was being shared. Links to website content originate in a story. Such stories can be authored by women who embed a link in a posted story, or authored by a marketer or news media organization that uses a shortened (eg, bitly) Web address to allow ease of use and visibility. Sysomos allowed us to identify the actual 10 most popular links and the frequency of each, by interaction type and content of link.

We clicked through all of these links and examined their content in detail. One study team member, a physician scientist (MH), manually categorized their content retroactively. This led to us identifying 5 mutually exclusive and collectively exhaustive categories to which all shared links belonged. These categories were *e-commerce related to breast cancer*, *celebrity breast cancer information*, *breast cancer advocacy and charity*, *noncelebrity breast cancer information*, and *unrelated to breast cancer*. This last category arose because, although a user may have been commenting on a breast cancer news story, they may also have been sharing an unrelated news item in the same post, and hence both were captured. These categories have not been externally validated and should be considered hypothesis generating only.

This study was conducted using completely deidentified, aggregated summary data provided by a third party, and accordingly did not involve human participant research and did not require an institutional review board determination or approval in our institution.

Textbox 1. Keywords and key phrases used to capture Facebook data.

Profession and organizational terms

Radiology

Doctor xray

Hospital Xray

Hospital Radiology

Radiologist

breast center

breast imaging center

breast cancer screening

breast screen guidelines

breast screening guidelines

Services rendered: formal terms

Mammogram

Mammography

Breast Exam

Digital mammography

digitized mammographic image

Breast tomosynthesis

three-dimensional mammography

three-dimensional mammogram

3-D mammogram

3d mammogram

3d mammogram

breast imaging

breast image

full-field digital mammogram

Screening Mammography

Screening Mammogram

Diagnostic Mammography

Diagnostic Mammogram

3-D mammography

mastectomy

3d mammography

Lumpectomy

full-field digital mammography

digital breast tomosynthesis

3d mammography

breast tumor

Digital mammogram

breast needle biopsy

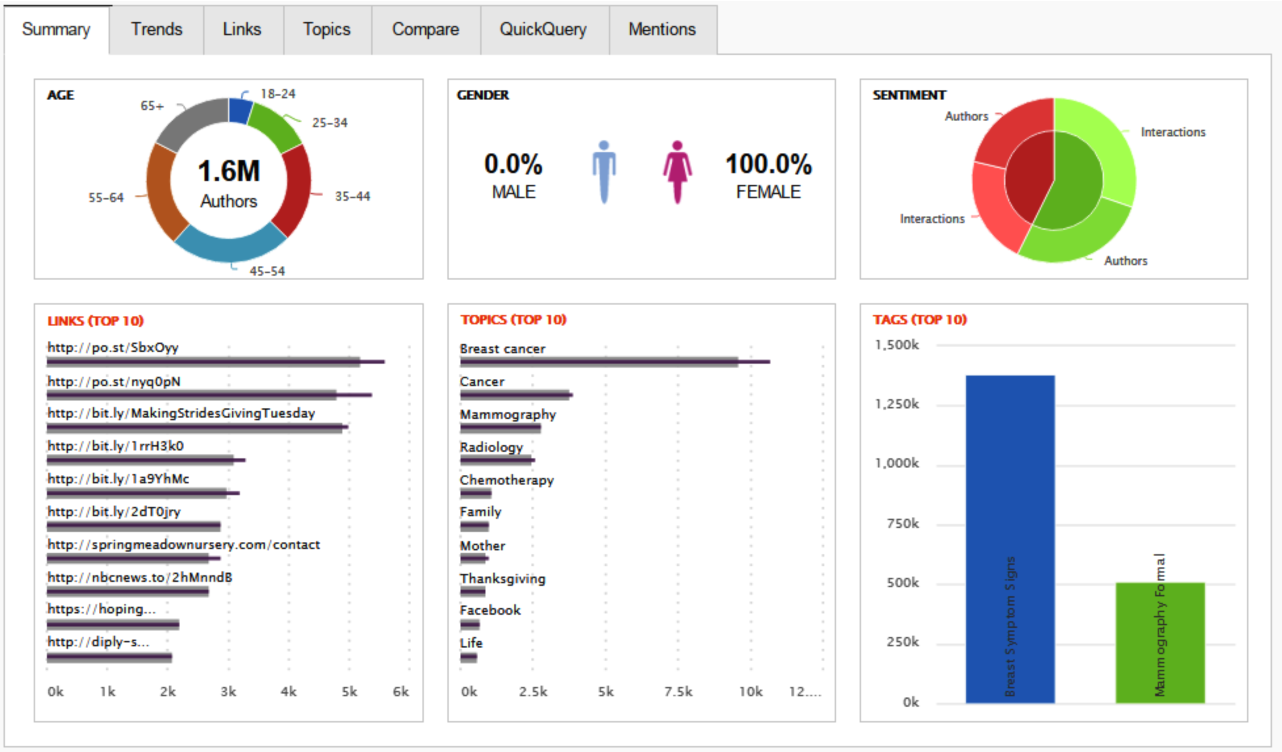
Services rendered: informal terms

breast xray

breast x-ray

Xray of my breasts
X-ray of my breasts
X-ray of my breast
Doctor x-rayed my breasts
Hospital X-rayed my breasts
x-rayed my breasts
needle biopsy done of my breast
needle biopsy of my breast
Symptoms and signs terms
breast lump
breast lumps
lump in my breast
Preexisting risk terms
BRCA tested positive
BRCA positive
family risk breast
high-risk breast
high-risk breasts
Findings terms
abnormal breast screen
abnormal breast x-ray
abnormal breast xray
dense breast
dense breasts
breast density
DCIS
ductal carcinoma
fatty breasts
fatty breast
breast cancer

Figure 1. Sysomos Subscriber Dashboard screenshot showing total authors, sex and age distributions, sentiment, top links shared, and top inferred topics (source: Sysomos).



Results

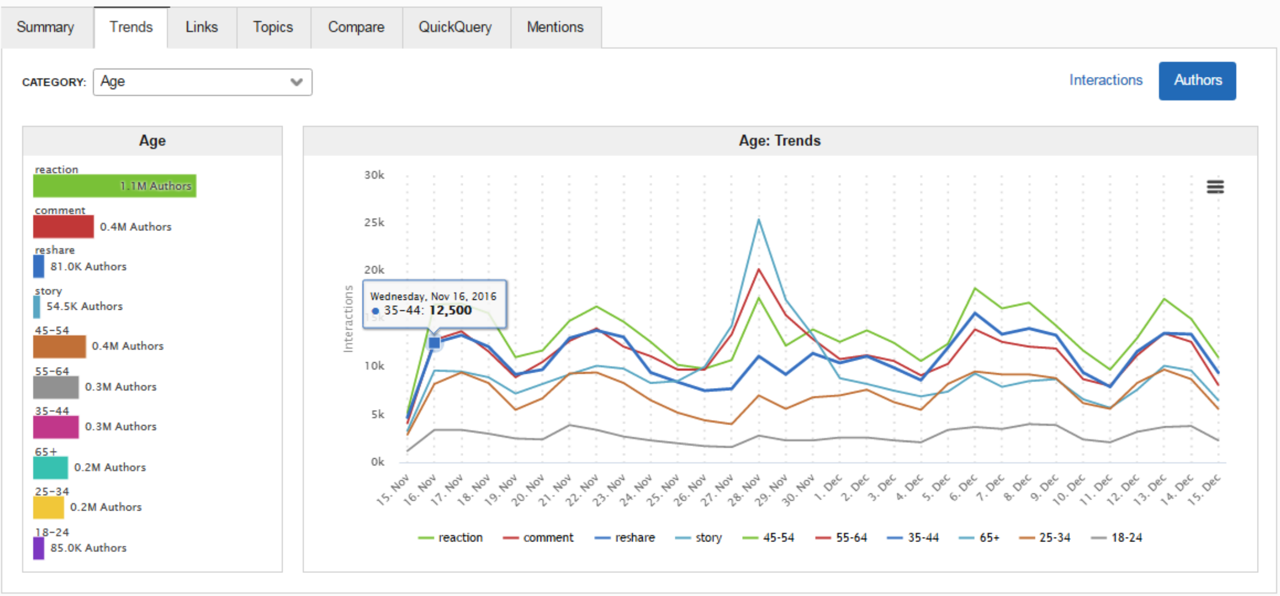
More than 1.7 million unique interactions (comments on stories, reshares, and emoji reactions) and stories associated with the 69 breast cancer screening keywords were generated by at least 1.1 million Facebook users over the 30-day period from November 16, 2016 to December 15, 2016.

On average, each of the 59,000 unique stories during the month was reshared 1.5 times, commented on nearly 8 times, and reacted to more than 20 times by other users seeing the original content.

Most Commonly Mentioned Terms

Stories and interactions were most often authored by women aged 45-54 years (Figure 2). We observed a substantial spike in volume on November 28, 2016, the Monday on which many news sites shared a picture of a bald Ms Shannen Doherty (an American actress) and her mother, immediately prior to Ms Doherty’s radiotherapy.

Figure 2. Sysomos Subscriber Dashboard screenshot showing trends in interaction types and age groups daily over the 30-day rolling time period (November 15-December 15, 2016) (source: Sysomos).



A search of mentions of “Doherty” in any link shared, reshared, or otherwise interacted with showed 6700 mentions by 6600 unique authors over the month, respectively 0.4% and 0.6% of the overall totals for the month.

Common terms relating to breast cancer and screening mammography mentioned in any context included “mammogram” (266,000 interactions, or 16% of the month’s total interactions), “lump” (26,600, 1.6%), “abnormal mammogram” (4400, 0.3%), “scars” (4000, 0.2%), “BRCA” (3800, 0.2%), “dense” (3200, 0.2%), “DCIS” (3000, 0.2%), “high risk” (2900, 0.2%), and “compression” (1000, 0.06%).

Most Commonly Shared Website Links by Interacted Type

Across all interactions, the 10 most popular links accounted for a total of 33,600 interactions, or almost 2% of monthly total interactions (Table 1). Multimedia Appendix 1 provides the actual links.

Links to *e-commerce related to breast cancer* represented the plurality of interacted links at 36% of all top interactions and were the most popular to be reshared with connected users (59% of all top reshares). These tended to represent a for-profit or not-for-profit organization that was selling items in some way connected to breast cancer themes over the Web. The most common of these were thebreastcancersite.com, hopinghand.com, and makaiclothing.com. Other links in this category were also aimed at recruiting email addresses for future direct marketing by offering giveaways (thebreastcancersite.com).

The next largest category was links to *celebrity breast cancer information*, which represented 26% of all top interactions and originated in stories such as from the television program *The Today Show*, E! Online, and similar sites. Here the most prominent foci were Ms Shannen Doherty, Ms Danielle Spencer, and Ms Robin Roberts, in that order. Almost all of these interactions were emoji-based reactions.

The third largest category represented links to *breast cancer advocacy and charity* sites, with 15% of all top interactions concentrated among links to 2 sites. The American Cancer Society donation page (“Making Strides Giving Tuesday”) was the second most popular link interacted with overall, with 4900 interactions or 0.3% of all interactions, was the most reacted to, and was the fifth most often reshared link among all top links. The other site was the advocacy and information site of Susan G. Komen, which did not make the top 10 links overall, and represented only 100 commented-on links (the 10th most commented-on link).

Noncelebrity breast cancer information links constituted only 6% of all top interactions and included mostly personal blog stories, inspirational messages, and some traditional news media. The most prominent themes in this category were positive stories around the use of tattoos to mask the surgical scars associated with breast cancer surgery and a new breast cancer vaccine trial at City of Hope hospitals in Duarte, California.

Importantly, in this category were at least 700 shared links relating to mercola.com, a natural health advocacy site that presented a view against breast cancer screening, including multiple references to scientific studies and a recent article by Welch and colleagues [29]. This single link alone represented nearly 6% of all the top links in the 35- to 54-year age group.

Finally, more than 1 in 6 links were not in relation to breast cancer or screening terms. These presumably were stories, reshares, and comments in which a user conveyed multiple messages, some about breast cancer (hence they were selected by Sysomos) and some not about this.

Most Commonly Shared Website Links by Age Group

We repeated our analyses to understand how interest and interactions changed across age groups (Table 2). Examining the totals of the top 10 links by age group, we found a clear increase with age. The most salient findings were the increase by age groups in e-commerce, peaking in the 45- to 54-year age group, and the complete or nearly complete lack of interest in those between 18 and 44 years of age in breast cancer advocacy and charity sites.

Additionally, we noted that noncelebrity-related news and information about breast cancer represented a larger share among the youngest users (50%) than among older users. We also noted the apparent complete lack of interest among the age group 65 years and older in celebrity-related breast cancer information.

Table 1. Distribution of most popular links by category and interaction type.

Top 10 links	Most popular reshared with others	Most commented on	Most reacted to	Overall most popular across all interactions
E-Commerce related to breast cancer	3100 (59%)	400 (29%)	9400 (32%)	12,200 (36%)
Celebrity breast cancer information	300 (6%)	500 (36%)	8000 (28%)	8800 (26%)
Noncelebrity breast cancer information	1000 (19%)	300 (21%)	1700 (6%)	2100 (6%)
Breast cancer advocacy and charity	300 (6%)	100 (7%)	4500 (16%)	4900 (15%)
Unrelated content	600 (11%)	100 (7%)	5400 (19%)	5600 (17%)
Total of top 10 link volume	5300	1400	29,000	33,600

Table 2. Distribution of most popular links by category and age group.

Top 10 links	Age group (years)						Overall most popular across all interactions
	18-24	25-34	35-44	45-54	55-64	≥65	
E-Commerce related to breast cancer	100 (7%)	900 (28%)	2200 (37%)	3500 (59%)	3400 (40%)	3100 (35%)	12,200 (36%)
Celebrity breast cancer information	200 (14%)	900 (28%)	2200 (37%)	2500 (42%)	1100 (13%)	0 (0%)	8800 (26%)
Noncelebrity breast cancer information	700 (50%)	200 (6%)	600 (10%)	400 (7%)	900 (11%)	1600 (18%)	2100 (6%)
Breast cancer advocacy and charity	0 (0%)	0 (0%)	300 (5%)	600 (10%)	1500 (17%)	3100 (35%)	4900 (15%)
Unrelated content	0 (0%)	400 (13%)	700 (12%)	900 (15%)	1700 (20%)	1100 (12%)	5600 (17%)
Total of top 10 link volume	1400	3200	6000	5900	8600	8900	33,600

Discussion

In this novel pilot study, we examined aggregated mentions of terms and phrases, and shared website links among women in the United States on Facebook in relation to breast cancer screening over a 1-month window. We found substantial content posted by, shared among, and interacted with by large numbers of women. The most popular stories provided information on women undergoing treatment for breast cancer and information on online destinations to purchase small items and make small donations to further research.

We observed that the timing of upswings in interest often appeared to coincide with celebrity news, such as a picture shared by Shannen Doherty of herself about to receive radiotherapy for her breast cancer. In general, our work supports the importance of sharing of and commenting on stories about well-known celebrities with breast cancer [30,31].

It is well-known that the Internet allows a so-called long tail to form, in which many niche sites, topics, or products are, respectively, visited, mentioned, or bought by a small number of users, in contrast with more popular sites, topics, or products [32]. We found a limited count of women creating and interacting with very popular content, such as content relating to “mammogram” (more than 15% of all interactions in our study). For example, we found that even many of the most popular terms, such as “DCIS” and “dense,” represented very small (0.2%) proportions of the overall number of story interactions. These may nevertheless be a meaningful subgroup. We saw the same phenomenon with interactions to links. The 10 most popular links accounted for just 2% of all interactions. This suggests that many items of less popular content were still, in aggregate, accounting for a large number of shared links.

Yet, despite these restrictions, we found that there was a plurality of links to commercial e-commerce websites marketing items related to breast cancer themes, such as thebreastcancersite.com. We saw little sharing of original medical news content from formal online media or formal health information publishers, despite the positive impact this can have [33]. There was some sharing of a story from a natural health website that appeared to be strongly against breast cancer screening. The extent of sharing of this site (nearly 6% of all the top links in the 35- to 54-year age group) and the strength of the views against breast cancer screening in the content on that site appeared to us to echo well-known online campaigns against vaccines by vaccine skeptics [34].

We also found less content than we had expected from some of the most prominent advocacy organizations, such as Susan G. Komen, although the American Cancer Society’s breast cancer site was the link with the second most frequent interactions. Finally, we saw fewer mentions than we had expected of terms anecdotally thought to be points of concern for women (eg, breast compression during imaging) and that had been found among their tweets in a recent innovative study by Rosenkrantz and colleagues [15]. One potential limitation is that, given the terms of use of the tool, we were able to examine only a 1-month study window.

As we continue to examine this new data source, we expect to obtain more detailed insights about what women are interacting about and how they are interacting regarding breast cancer terms. We expect that such data can inform the outreach of advocacy organizations, and can inform campaigns to improve rates of screening and to educate high-risk women concerning their options, among many other examples.

Methodologically, this study adds to our understanding of patients’ and consumers’ articulated thoughts and feelings about important public health initiatives such as breast cancer screening. We showed that summarized information is available from the world’s leading online social network, and note that this commercially available information is distinct from more easily analyzed public online social media. Given the greater demographic representativeness of Facebook, compared with other online social media and social networks [8], the data on this platform are a potentially useful research tool.

Limitations

While our study had several important strengths, including novelty, exhaustiveness, and national scale in the United States, there are several important limitations. Our data source, Sysomos, is a commercial reseller of data obtained indirectly from Facebook through another intermediary, Datasift. Data provenance, custody, and governance must be assumed but cannot be verified or guaranteed. For example, software errors could occur at each one of these handoffs, as well as within each segment of the data custody chain.

In particular, Facebook is the data owner, whose terms of service do not permit actual visualization of the original post or comment. To protect users' privacy, all data were aggregated, deidentified, and mapped coarsely into topics. We therefore had no independent ability to confirm whether the reported statistics we obtained were accurate, representative, or exhaustive. Moreover, under our contract, data availability was limited to rolling 1-month lookback periods. Other restrictions motivated by privacy and imposed by the data owner include sampling only high-frequency items, limiting results to the top 10 items in a category, and masking results in which fewer than 100 Facebook users mentioned a term or shared a link. As a result, none of our results were able to provide a full view of the frequency distribution.

Neither we nor other researchers can subsequently return to historical periods beyond examining reports that were downloaded contemporaneously. Similarly, only a 30-day rolling period of aggregated data is made available by Facebook, Datasift, and Sysomos. This clearly further limits replication and error checking. For research purposes, while substantial information abides, much is lost during this process. This weakness does not seem to be one that will be alleviated, given legitimate concerns regarding online privacy [35].

Finally, while our research is internally valid, the extent to which it is externally applicable is not known. The particular month of data we looked at was almost immediately after a polarizing general election in the United States, in which health-related conversations (eg, Affordable Care Act, Planned Parenthood, women's right to choose) were widely occurring. In other months, there might have been fewer mentions of breast cancer screening terms. Our research also explicitly required women to have access to the Internet, be a member of Facebook, and use English in their interactions. There are clearly large parts of US society in which one or more of these requirements are not met.

Future researchers may exploit other less coarse methods for obtaining online social media and social network data. Companies operating online survey panels such as Knowledge Networks, Inc [36], Qualtrics, and ClearVoice Research [37] can allow more representative surveys and more specific questions as to what women share on Facebook. Free resources can also be accessed through Google's own Trends data—for example, to analyze searches related to Angelina Jolie's disclosed prophylactic mastectomies [8], and to understand interest in public hospital quality reports [38]—or by exploiting the freely available Twitter data [39].

Conclusions

Examining novel data from the universe of mentions on the leading online social network regarding breast cancer screening-related terms provided an important but superficial and initial look at topics of great interest among all female Facebook users over 1 month. More work is needed using this novel data source and applying its insights to solving pressing public health problems, including the inadequate screening for breast cancer.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Categorization of top 10 links by interaction type and by age category.

[PDF File (Adobe PDF File), 69KB - jmir_v19i6e201_app1.pdf]

References

1. U.S. breast cancer statistics. Ardmore, PA: Breastcancer.org; 2016. URL: http://www.breastcancer.org/symptoms/understand_bc/statistics [accessed 2017-02-10] [WebCite Cache ID 6oBGgZtVq]
2. National Center for Health Statistics. Mammography. Atlanta, GA: U.S. Department of Health & Human Services; 2017 Jan 19. URL: https://www.cdc.gov/nchs/fastats/mammography.htm [accessed 2017-02-10] [WebCite Cache ID 6oBGiCAry]
3. Kroenke CH, Kubzansky LD, Schernhammer ES, Holmes MD, Kawachi I. Social networks, social support, and survival after breast cancer diagnosis. J Clin Oncol 2006 Mar 01;24(7):1105-1111. [doi: 10.1200/JCO.2005.04.2846] [Medline: 16505430]
4. Chetlen A, Mack J, Chan T. Breast cancer screening controversies: who, when, why, and how? Clin Imaging 2016;40(2):279-282. [doi: 10.1016/j.clinimag.2015.05.017] [Medline: 26093511]
5. Webb ML, Cady B, Michaelson JS, Bush DM, Calvillo KZ, Kopans DB, et al. A failure analysis of invasive breast cancer: most deaths from disease occur in women not regularly screened. Cancer 2014 Sep 15;120(18):2839-2846 [FREE Full text] [doi: 10.1002/cncr.28199] [Medline: 24018987]
6. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Health, United States, 2015: with special feature on racial and ethnic health disparities. Hyattsville, MD: National Center for Health Statistics; 2016. Table 70. Use of mammography among women aged 40 and over, by selected

characteristics: United States, selected years 1987-2013 URL: https://www.cdc.gov/nchs/hus/index.htm [accessed 2017-02-10] [WebCite Cache ID 6qvPapoOP]

7. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
8. Huesch M. Celebrity influence? An observational study of consumer responses on social media and on an internet search engine to news of Angelina Jolie?s prophylactic mastectomies. CESR-Schaeffer Working Paper Series, Paper No-2015-012. Los Angeles, CA: University of Southern California, Center for Economic and Social Research; 2015. URL: https://cesr.usc.edu/documents/WP_2015_012.pdf [accessed 2017-05-31] [WebCite Cache ID 6qsBowcVp]
9. Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. J Med Internet Res 2013;15(4):e62 [FREE Full text] [doi: 10.2196/jmir.2503] [Medline: 23594933]
10. Huesch MD, Galstyan A, Ong MK, Doctor JN. Using social media, online social networks, and internet search as platforms for public health interventions: a pilot study. Health Serv Res 2016 Jun;51 Suppl 2:1273-1290. [doi: 10.1111/1475-6773.12496] [Medline: 27161093]
11. Batterham PJ. Recruitment of mental health survey participants using Internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191. [doi: 10.1002/mpr.1421] [Medline: 24615785]
12. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]
13. Carter-Harris L, Bartlett ER, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. J Med Internet Res 2016;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
14. Fazzino TL, Rose GL, Pollack SM, Helzer JE. Recruiting U.S. and Canadian college students via social media for participation in a web-based brief intervention study. J Stud Alcohol Drugs 2015 Jan;76(1):127-132 [FREE Full text] [Medline: 25486401]
15. Rosenkrantz AB, Labib A, Pysarenko K, Prabhu V. What do patients tweet about their mammography experience? Acad Radiol 2016 Nov;23(11):1367-1371. [doi: 10.1016/j.acra.2016.07.012] [Medline: 27658329]
16. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]
17. Basch CH, Hillyer GC, MacDonald ZL, Reeves R, Basch CE. Characteristics of YouTube™ videos related to mammography. J Cancer Educ 2015 Dec;30(4):699-703. [doi: 10.1007/s13187-014-0769-9] [Medline: 25502853]
18. Coughlin SS. Intervention approaches for addressing breast cancer disparities among African American women. Ann Transl Med Epidemiol 2014 Sep 08;1(1):1001 [FREE Full text] [Medline: 25568890]
19. Fazeli DS, Carlos RC, Hall KS, Dalton VK. Novel data sources for women's health research: mapping breast screening online information seeking through Google Trends. Acad Radiol 2014 Sep;21(9):1172-1176 [FREE Full text] [doi: 10.1016/j.acra.2014.05.005] [Medline: 24998689]
20. Joison A. 'Looking at', 'looking up' or 'keeping up with' people? Motives and uses of Facebook. 2008 Presented at: 26th Annual CHI Conference on Human Factors in Computing Systems; Apr 5-10, 2008; Florence, Italy.
21. Nadkarni A, Hofmann SG. Why do people use Facebook? Pers Individ Dif 2012 Feb 01;52(3):243-249 [FREE Full text] [doi: 10.1016/j.paid.2011.11.007] [Medline: 22544987]
22. Gabarron E, Serrano JA, Wynn R, Lau AY. Tweet content related to sexually transmitted diseases: no joking matter. J Med Internet Res 2014;16(10):e228 [FREE Full text] [doi: 10.2196/jmir.3259] [Medline: 25289463]
23. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]
24. Theiss SK, Burke RM, Cory JL, Fairley TL. Getting beyond impressions: an evaluation of engagement with breast cancer-related Facebook content. Mhealth 2016 Nov;2:41 [FREE Full text] [doi: 10.21037/mhealth.2016.10.02] [Medline: 27840816]
25. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, et al. Web conversations about complementary and alternative medicines and cancer: content and sentiment analysis. J Med Internet Res 2016;18(6):e120 [FREE Full text] [doi: 10.2196/jmir.5521] [Medline: 27311444]
26. Bender JL, Jimenez-Marroquin M, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
27. Smith A. 6 new facts about Facebook. Washington, DC: Pew Research Center; 2014 Feb 03. URL: http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/ [accessed 2017-02-10] [WebCite Cache ID 6oBGm5m6g]
28. Duggan M. The demographics of social media users. Washington, DC: Pew Research Center; 2015 Aug 19. URL: http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/ [accessed 2017-02-10] [WebCite Cache ID 6oBGrROV4]
29. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med 2016 Dec 13;375(15):1438-1447. [doi: 10.1056/NEJMoa1600249] [Medline: 27732805]

30. Juthe RH, Zaharchuk A, Wang C. Celebrity disclosures and information seeking: the case of Angelina Jolie. Genet Med 2015 Jul;17(7):545-553 [FREE Full text] [doi: 10.1038/gim.2014.141] [Medline: 25341112]
31. PressRun: The New York Times's most visited content of 2013. New York, NY: The New York Times Company; 2014 Jan 17. URL: http://www.nytco.com/the-new-york-timess-most-visited-content-of-2013/?smid=tw-share [accessed 2017-02-10] [WebCite Cache ID 6oBGywMFR]
32. Karch M. What is the long tail and how does it apply to Google?. Bellevue, WA: Lifewire; 2016 Nov 16. URL: https://www.lifewire.com/long-tail-and-google-1616459 [accessed 2017-02-10] [WebCite Cache ID 6oBH0HQFx]
33. Price J, Simon K. Patient education and the impact of new medical research. J Health Econ 2009 Dec;28(6):1166-1174. [doi: 10.1016/j.jhealeco.2009.08.005] [Medline: 19783313]
34. Abbasi A, Adjeroh D, Dredze M, Paul MJ, Zahedi FM, Zhao H, et al. Social media analytics for smart health. IEEE Intell Syst 2014 Mar;29(2):60-80. [doi: 10.1109/MIS.2014.29]
35. Huesch MD. Privacy protection during internet search for health-related information--reply. JAMA Intern Med 2014 Mar;174(3):477. [doi: 10.1001/jamainternmed.2013.13513] [Medline: 24590090]
36. Doctor JN, Huesch MD, Meeker D. Rethinking the value of survival: clinical trials should measure patient preferences for survival on entry to trials. J Clin Epidemiol 2016 Sep;77:137-138. [doi: 10.1016/j.jclinepi.2016.03.030] [Medline: 27164276]
37. Huesch M, Brady R. Should past recipients of a kidney be allowed to have another? Surveys of preferences of individuals with and without end stage renal disease. USC CESR-Schaeffer Working Paper Series, Paper No-2015-006. Los Angeles, CA: University of Southern California, Center for Economic and Social Research, Leonard D Schaeffer Center for Health Policy & Economics; 2015. URL: https://cesr.usc.edu/documents/WP_2015_006.pdf [accessed 2017-05-31] [WebCite Cache ID 6qsDXrGVO]
38. Huesch MD, Currid-Halkett E, Doctor JN. Public hospital quality report awareness: evidence from national and Californian Internet searches and social media mentions, 2012. BMJ Open 2014 Mar 11;4(3):e004417 [FREE Full text] [doi: 10.1136/bmjopen-2013-004417] [Medline: 24618223]
39. Huesch M, Ver Steeg G, Galstyan A. Vaccination (anti-) campaigns in social media. 2013 Presented at: Workshop on Expanding the Boundaries of Health Informatics using AI, AAAI 2013 Workshop; Jul 14-15, 2013; Bellevue, WA, USA URL: https://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/view/7094/6502

Edited by A Keepanasseril; submitted 14.02.17; peer-reviewed by YTD Cheung, C Basch, A Cyr; comments to author 09.03.17; revised version received 01.05.17; accepted 10.05.17; published 09.06.17.

Please cite as:
Huesch M, Chetlen A, Segel J, Schetter S
Frequencies of Private Mentions and Sharing of Mammography and Breast Cancer Terms on Facebook: A Pilot Study
J Med Internet Res 2017;19(6):e201
URL: http://www.jmir.org/2017/6/e201/
doi:10.2196/jmir.7508
PMID:28600279

Original Paper

Too Far to Care? Measuring Public Attention and Fear for Ebola Using Twitter

Liza GG van Lent[1], MA; Hande Sungur[2], MSc; Florian A Kunneman[1], PhD; Bob van de Velde[2], PhD; Enny Das[1], PhD

[1]Centre for Language Studies, Radboud University, Nijmegen, Netherlands
[2]Communication Sciences, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Liza GG van Lent, MA
Centre for Language Studies
Radboud University
PO Box 9103
Nijmegen, 6500 HD
Netherlands
Phone: 31 024 3612928
Fax: 31 024 3612177
Email: lizavanlent@gmail.com

Abstract

Background: In 2014, the world was startled by a sudden outbreak of Ebola. Although Ebola infections and deaths occurred almost exclusively in Guinea, Sierra Leone, and Liberia, few potential Western cases, in particular, caused a great stir among the public in Western countries.

Objective: This study builds on the construal level theory to examine the relationship between psychological distance to an epidemic and public attention and sentiment expressed on Twitter. Whereas previous research has shown the potential of social media to assess real-time public opinion and sentiment, generalizable insights that further the theory development lack.

Methods: Epidemiological data (number of Ebola infections and fatalities) and media data (tweet volume and key events reported in the media) were collected for the 2014 Ebola outbreak, and Twitter content from the Netherlands was coded for (1) expressions of fear for self or fear for others and (2) psychological distance of the outbreak to the tweet source. Longitudinal relations were compared using vector error correction model (VECM) methodology.

Results: Analyses based on 4500 tweets revealed that increases in public attention to Ebola co-occurred with severe world events related to the epidemic, but not all severe events evoked fear. As hypothesized, Web-based public attention and expressions of fear responded mainly to the psychological distance of the epidemic. A chi-square test showed a significant positive relation between proximity and fear: χ^2_2=103.2 (*P*<.001). Public attention and fear for self in the Netherlands showed peaks when Ebola became spatially closer by crossing the Mediterranean Sea and Atlantic Ocean. Fear for others was mostly predicted by the social distance to the affected parties.

Conclusions: Spatial and social distance are important predictors of public attention to worldwide crisis such as epidemics. These factors need to be taken into account when communicating about human tragedies.

(J Med Internet Res 2017;19(6):e193) doi:10.2196/jmir.7219

KEYWORDS

psychological theory; epidemics; fear; distance perception; social media

Introduction

General Background

In early spring of 2014, the world was startled by an outbreak of Ebola: a fairly unfamiliar and incurable virus with a high risk of death. Starting in March in Guinea, the epidemic quickly spread to West African cities, causing over 10,000 infections and 5000 deaths in a course of 9 months [1]. Although Western media covered the main events relating to the unfolding Ebola crisis in West Africa, especially psychologically close events appeared to cause a stir in the West. For instance, the reporting of a few Ebola cases in the United States led to demonstrations in favor of a travel ban even for countries like Saudi Arabia, where Ebola did not actually prevail [2], and in Western Europe significant public attention was paid to the euthanasia of a dog that belonged to an infected Spanish nurse [3].

When do humans start paying attention to real-world events? What aspects of an epidemic are most likely to increase fear? How can one adequately isolate the factors responsible for sudden surges in public attention and fear following health crises like the Ebola outbreak? The present research uses Twitter to examine the role of real-time changes in spatial and social distance to the epidemic to understand shifts in public attention and fear during a crisis. Previous research used Twitter as a proxy for disease activity [4-6] and to assess real-time public sentiment and opinion [7-10]. The present research complements these findings by applying construal level theory (CLT) [11] to examine Web-based attention and fear in public crises. The present research also contributes to CLT literature by being one of the few studies to examine CLT assertions beyond experimental laboratory conditions with longitudinal data [12,13].

Construal Level Theory of Psychological Distance

CLT is a psychological theory that explains the relationship between psychological distance of stimuli (eg, events, objects, and people) and how they are mentally represented or construed [11,14]. Psychological distance refers to the subjective distance stimuli maintain from a person's direct experience [15], which is centered around "here," "now," the "self," and "reality." Based on CLT, as the Ebola epidemic comes closer on these four dimensions (ie, spatially, temporally, socially, or hypothetically), it also becomes "psychologically closer." According to CLT, psychological distance plays a fundamental role in shaping mental representations [14,16]. Whereas psychologically close stimuli tend to be represented in a detailed, contextualized, and concrete manner (ie, low-level construal), psychologically far stimuli are represented more generally and abstractly (ie, high-level construal) [14,17]. The association between psychological distance and construal level has been successfully applied to explain various psychological processes such as biases and decision making [12,14], and the CLT framework has been increasingly applied by communication scholars [18-21].

Psychological distance has important implications for risk perceptions and experiencing affect. Previous research has shown that fear and arousal for (real or imagined) negative events decrease with increased psychological distances [22,23]. Similarly, perceptions of risk for negative events, such as a health hazard, have shown to increase with psychologically close framing [24], whereas increased psychological distance is linked to lower risk perceptions [13]. Low-level construals (and close psychological distances) have also been linked to increased truth [25,26] and likelihood perceptions [15,27,28] (eg, risk of contracting a disease), and higher behavioral intentions [29] compared with high-level construals (and far psychological distances). The underlying reason for these effects is suggested to be the more concrete and detailed nature of mental representations in closer psychological distances. The reduced intensity of affect caused by high psychological distance is also suggested to be due to the critical role physical distance has for human biology and survival [22]. In line with these findings, a previous study showed that as time passed, people used less anxiety- and sadness-related words on Twitter about the Sandy Hook Elementary School shooting [30]. However, the same study also showed that an increase in spatial distance caused a decrease in the number of sadness-related words but an increase in the number of anxiety-related words. In addition, "focusing on the abstract causes of this tragedy (rather than the concrete details) decreased sadness (...) but increased anxiety" (p.370). Although highly relevant to this research, the Sandy Hook shooting was one big impactful event in the United States. This research examines responses to an impactful health crisis as it unfolds across different continents over the course of several months.

Psychological Distance, Severity, and Fear for Epidemics Expressed on Twitter

Previous social media findings suggest that although media patterns sometimes converge with epidemic curves [5], they appear better suited to track public opinion and sentiment. This is because the Web-based media patterns more often follow the agenda of classical news media about an epidemic in an area of focus [7,8]. Classical media curves often do not converge with epidemic curves but rather are governed by the laws of news values. For example, the first infection in an epidemic has higher news value than the 1000th infection [31]. Although the relationship between classical and Web-based responses has been examined within one specific region, little is known about the relationship between gradual changes in psychological distance to certain real world events (eg, increased infection rates) and public attention and sentiment. Gaining more insight into this relationship is important to obtain generalizable insights about human responses to health crises. Specifically, 2 key theoretical questions have not been answered by previous CLT studies: (1) "when do individuals start experiencing events as psychologically close?" and (2) "does the severity of a (psychologically far or close) event play a role in public sentiment?"

Building on previous research regarding the use of social media as a real-time public opinion and sentiment proxy, we examined Twitter patterns as the Ebola epidemic approached the area of focus, in this case the Netherlands. We connected these patterns to specific locations (eg, West Africa and Spain) mentioned in the Tweets in order to examine the relationship between psychological distance of the outbreak, public attention, and fear. Based on CLT, as psychological distance decreases, mental representations of the events concerning the Ebola outbreak should become more concrete. We therefore expect increased public attention, as expressed by the number of tweets, as the Ebola outbreak becomes psychologically closer (Hypothesis 1). An earlier Twitter study found that surveillance of flu infection rates could be improved by using tweets that contained reference to "self" and "infection" (vs "other" and "concerned awareness") [4]. This gives reason to think that a similar difference could be observed for fear; people may not fear becoming infected with Ebola unless an epidemic becomes psychologically close. This implies that especially expressions of fear for self should respond to reports of increased psychological closeness of an epidemic (Hypothesis 2).

It also remains unclear to what extent the affective value of an event itself plays a role in the relationship between psychological distance and public sentiment. Journalism studies suggest that severity is a factor that crucially determines the prominence of a certain real-world event. For example, 1000 casualties have bigger news value than a small number of casualties and will consequently trigger higher levels of media attention [32]. However, media coverage of severe epidemics also includes key political and economic events such as political debates about the epidemic or travel bans [32]. As social media follows classical media patterns, the severity of health events may only partially direct public attention on Twitter, together with key (nonhealth) events related to the epidemic. Previous studies showed that media volume curves appear to align more with public opinion than the epidemic curves, but these studies mostly investigated this relationship in a rather static manner, as they occurred within the same region. As the vast majority of Ebola cases were observed in West Africa, a comparison of Twitter curves in the Netherlands with the epidemic curves and severe world events relating to the epidemic should provide new insight into the relationship between the severity of the epidemic, and public attention and fear in a more global and transient manner. [7]. We therefore examined the relationship between public attention for Ebola on Twitter on the one hand and infection and mortality rates on the other (RQ1).

Methods

Material

A corpus of 185,253 Dutch tweets containing "(#)Ebola" was built, dating from the first outbreak (and report) of Ebola on March 22, 2014 to October 31, 2014. We collected the tweets from the TwiNL archive, a corpus of Dutch tweet IDs posted from December 2010 onward [33]. The peaks of Ebola prevalence in the corpus were connected to corresponding real-world events in the news and to infection and mortality rates, as counted by the Centers for Disease Control and Prevention (2015). A sample of 4500 tweets was selected by using the randomize function of Microsoft Excel and was then hand-coded by 3 independent coders who each coded 2000 tweets.

Manual Encoding

Tweets were encoded for "fear" as (1) fear for self, (2) fear for others, and (3) no fear. The coders used fear-related keywords and emoticons such as: fear, scare(d), threat, help, scary, brr, OMG (oh my god), WTF (what the fuck), danger(ous), panic, dare (not), creepy, infected, contagious, death(s), care, symptoms, hospital, ill(ness), die(d), mortal(ity (rate)), spread, infection (rate), risk (group), :(, :'(, :O [8,34]. Whether these were indeed designators of fear (and not just of risk) was evaluated in the context of each tweet. Furthermore, simple notifications of reported fear without emotional connotations were coded as "no fear," even when they contained the expressed worries from others. The difference between fear for self and others was based on the use of subjects, verbs, and (personal) pronouns [4]. When the (implied) subject was "I" or "we," that tweet was encoded as "fear for self" and otherwise as "fear for others." The intercoder reliability was high (Cohen kappa=.930).

To assess the "psychological distance" of Ebola toward the Netherlands, tweets were coded into one of the twelve different categories: (1) the Netherlands, (2) neighboring countries, (3) West and North Europe, (4) South Europe, (5) East Europe, (6) North America, (7) North Africa, (8) West Africa (Guinea, Sierra Leone, Liberia), (9) South Africa, (10) South America, (11) Asia and Oceania, and (12) no location. Tweets were assigned a specific "distance" category when the tweet explicitly expressed a location within the concerning area (eg, when a tweet mentioned a potential case in Zurich, that tweet was encoded as "(3) West and North Europe"). When a tweet did not mention a specific location, it was encoded as "(12) no location." When a tweet mentioned multiple areas, the code of the nearest area was granted.

To our knowledge, no previous research identified the psychologically close or far countries or areas for the Dutch population. Therefore, this classification was based on the dimensions of psychological distance proposed by CLT. Following considerations of spatial and social distance, we formed the categories by starting from the areas the Dutch people presumably regard the nearest and continued categorizing toward the furthest areas. According to CLT, people form higher number of specific categories for low and fewer or broader categories for high psychological distances [35]. We therefore categorized a higher number of more specifically formed groups within close spatial distance of the Netherlands (ie, Europe is divided in groups 1 to 5), and broader groups out of areas with high spatial distance (eg, Asia and Oceania form group 11). North America is classified as the sixth nearest group; although it was not the next spatially closest area, we reasoned it will probably be seen as psychologically closer due to the closer social distance it implies for the Netherlands. West Africa was classified separately because it was thematically relevant as the Ebola epidemic originated there. When a tweet mentioned multiple areas, the code of the nearest area was applied. Intercoder reliability was high (Cohen kappa=.911).

Time Series Analysis

Relations between time series were estimated and corrected using vector error correction models (VECMs) to account for cointegration (ie, tendencies to equilibrium) in different time series with lags based on empirical results using Aikake information criterion. The model to support claims regarding fear was specified using the Johansen procedure for determining cointegration with alpha =.05 threshold. These series were logged and corrected for trends after determining the original series contained violations of the stationarity assumption using the Kwiatkowski et al unit root test with constant; violations of normality detected by the Jarque-Berra residual normality, skew, and kurtosis test; and remaining autocorrelation by using the multivariate Portmanteau- and Breusch-Godfrey test for serially correlated errors. The logged model showed no signs of serial correlations (χ^2_{224}= 234.4 , *P*>.05) or of lingering heteroscedasticity , although non-normality persisted. The new-deaths and new-cases series were used as exogenous dummy variables. The research question was modeled using the same tests. In this model, non-normality also persisted but no heteroscedasticity was observed. In both cases a visual inspection of the residual plots did not indicate strong biases relevant to the reported results. In addition, we tested robustness of findings with different imputed breakpoints identified through the Zivot and Andrews unit root test and saw no significant changes. The expressed correlations are based on the error-corrected vectors of the error correction model (ECM) time series.

Results

Social Media Curve Versus Epidemic Curve

Figure 1 shows that at the onset of the epidemic Ebola received fairly little Twitter attention in the Netherlands. From March 22 to July 21, 2014, 8600 tweets (4.64%) were sent, of which the most encoded tweets either did not contain a location (57.7%, n=105) or referred to West Africa (34.1%, n=62). Only 13 of those tweets contained fear (both for self and others). Figures 2 and 3 show data on July 22, 2014, after which an increase in the daily tweet volume can be observed. The Twitter curve does not coincide well with the epidemic curves for the number of Ebola cases and deaths, whereas the total cases and deaths grow very fast at the end of October, the tweet volume goes down. At first glance, the Twitter curve mainly coincides with real-world events in relatively psychologically close areas; for example, when Ebola crossed the Mediterranean Sea and the Atlantic Ocean (suggesting increased proximity), peaks in the tweet volume can be observed. A VECM (r= 2) causality analysis of differences confirmed that new reported infections and new reported deaths do not seem to granger cause (F_{6663}=0.19, P>.05) or instantaneously cause (χ^2_2=0.3, P>.05) the number of tweets. This answers RQ1, showing that the epidemic curve and social media curve do not coincide.

Figure 1. The daily Dutch tweet volume about Ebola from March 22 to October 31, 2014 (N=185,253); and the reported new cases (N=13,540) and deaths (N=4941) caused by Ebola in Guinea, Liberia, and Sierra Leone, according to the Centers for Disease Control and Prevention (2015). The primary bar lines indicate one month, the secondary one week.

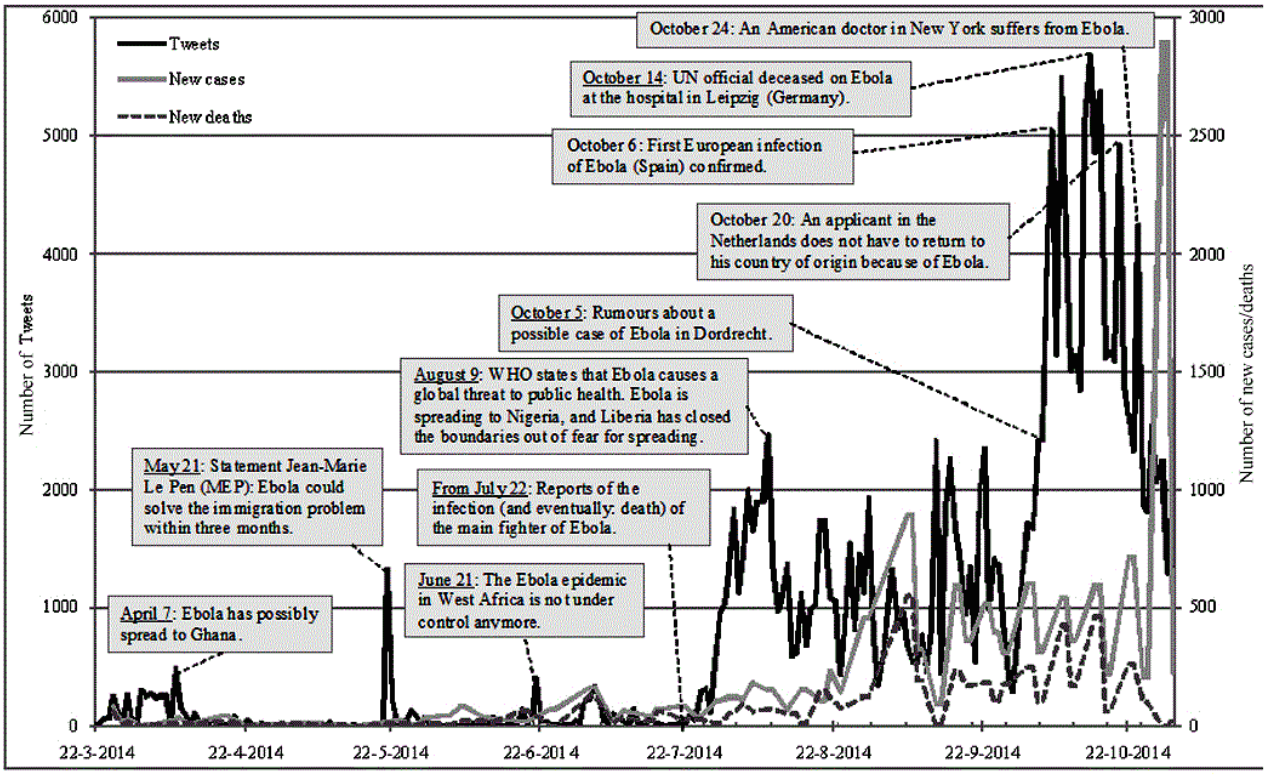


Figure 2. The daily amount of coded tweets containing fear for self or other or none plotted over time from July 22 to October 31, 2014. The primary bar lines indicate one month, the secondary one week.

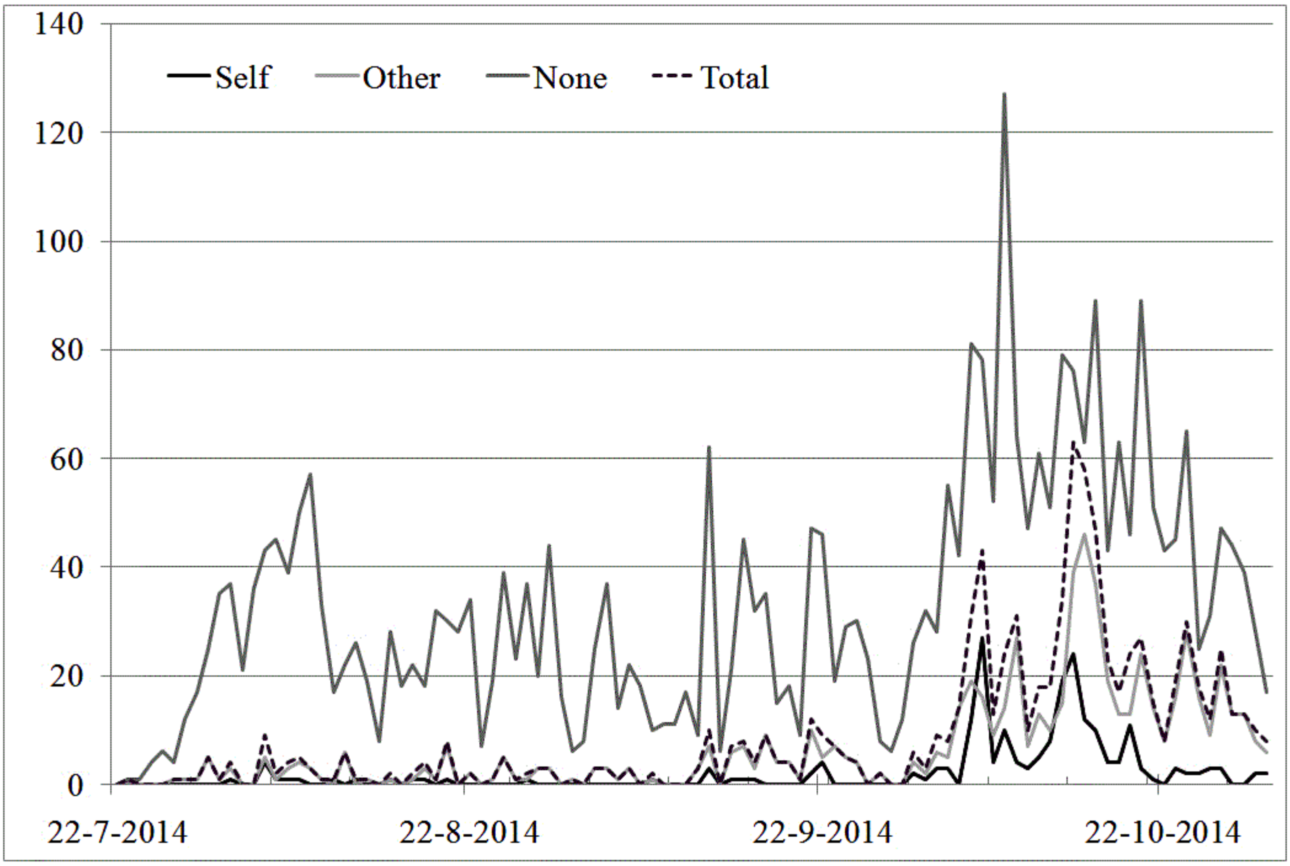
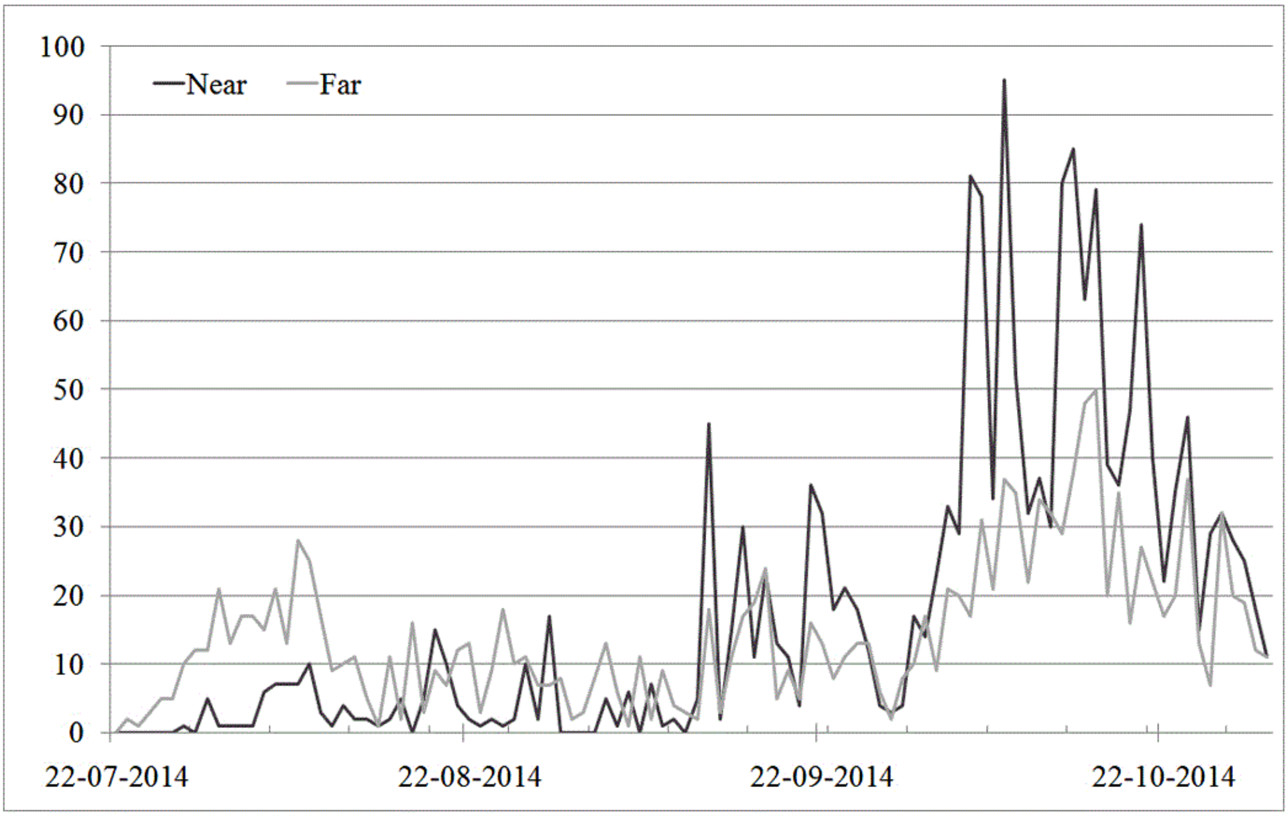


Figure 3. The daily amount of coded tweets about Ebola, related to near (Netherlands and neighboring countries) versus far (all other) locations, from July 22 to October 31, 2014. The primary bar lines indicate one month, the secondary one week.



Ebola-Related Fear as Expressed in Dutch Tweets

Figure 2 shows the daily number of fear expressions (self, other, total, and none), plotted over time. About 19.9% (n=896) of the coded tweets contained fear for Ebola, but the majority (80.1%, n=3604) did not. The figure shows two fairly even trends for fear for self and others, wherein the frequency for self (4.9%, n=222) was lower than for others (15.0%, n=674) during almost the entire period. "Total" and "none" fear did not always follow the same pattern, and the number of occurrences for no fear was the highest during the entire period. Comparing the tweets coded for fear with the overall number of Ebola related tweets, it becomes clear that "no fear" has the best correlation with the total Twitter curve. After correcting for exogenous influences of the events on the ground and the cointegration of these time-series (Johansen procedure $r=2$), "no fear" shows strongest correlation with the overall level of tweets ($r_{222}=.95$, $P<.05$). "Self" ($r_{222}=.56$, $P<0.05$) and "other" ($r_{222}=.73$, $P<0.05$) show more modest associations. Fear for self and others follow a different pattern, especially in the early phase, with correlations before August 18, 2014 of $r=.47$ and $r=.59$ after cointegration correction respectively. The first significant peak of fear for others is not seen until October 5, 2014, and although "self" increases as well, its first peak is seen the next day. On those days, there were notifications of a potential (later disconfirmed) Ebola patient in the Dutch city Dordrecht (October 5, 2014), and the first confirmed spread to Europe (October 6, 2014). On October 13, 2014, a notification was sent about a potential Ebola patient in a Belgian hospital; this news event is followed by the second peak for "self" on October 14, 2014 and for "others" on October 15, 2014.

Fear as a Function of Psychological Distance

As shown in Table 1, the percentage of fear for self was highest for the Netherlands (8.5%, n=134) and neighboring countries (18.5%, n=49). When a tweet did not specify a location, that tweet showed no fear for Ebola in almost all situations (90.1%, n=1053). Areas observed less than 100 times in total were excluded as separate values because of their possibly distorted percentages, but they were included in the "total" percentages. For the Netherlands and West Africa, percentages of tweets displaying fear for others amounted to around 16%; for the other locations (save "no location"), these percentages amounted to around 20-25%. The fairly low percentage for West Africa is particularly striking given the vast majority of infections and deadly cases reported in this area, compared with only few incidental infections in other regions.

Table 1. Location versus fear: crosstab of the percentage tweets per location (N>100) that contained fear for self, fear for others, or no fear. The excluded locations (N<100) are taken into account in the "total" percentages.

Location	Fear		
	Self	Others	None
Netherlands (n=1572)	8.5	16.2	75.3
Neighboring countries (n=265)	18.5	22.6	58.9
South Europe (n=113)	1.7	20.4	77.9
North America (n=362)	0.6	24.3	75.1
West Africa (n=632)	1.1	15.4	83.5
No location (n=1174)	1.5	8.5	90.0
Total (N=4500)	4.9	15.0	80.1

Nevertheless, it remains difficult to draw final conclusions about a possible correlation between fear and psychological distance. Therefore, a new proximity variable was created by merging the Netherlands and neighboring countries under the category of "near" and all other locations under "far." "No location" tweets were regarded missing, as they did not indicate distance (see Table 2). A chi-square test showed a significant positive relation between proximity and fear: $\chi^2_2=103.2$ ($P<.001$). Table 2 suggests that fear for self was found significantly more often than expected when linked to near locations, and significantly less when linked to far locations. The frequencies for fear for others and no fear did not differ significantly. Figure 3 further shows that the number of tweets about near locations increased particularly in the beginning of October, when the first Ebola cases were reported in Europe. Tweets on far locations are much less frequent during this time frame, in spite of significant increases in infections and deaths reported in West Africa (Figure 1). These findings confirm H1 and H2, that proposed that public attention and fear for Ebola decreased as a function of psychological distance.

Table 2. Psychological distance versus fear: crosstab for the percentage of tweets containing fear for self, fear for others, or no fear when location is near or far from the Netherlands, with the standardized residuals in brackets.

Psychological distance	Fear		
	Self	Others	None
Near (n=1837)	9.96 (6.5)	17.09 (−0.5)	72.95 (−1.6)
Far (n=1493)	1.28 (−7.4)	18.23 (0.6)	80.48 (1.8)
Total (N=3330)	6.20	17.59	76.21

Discussion

Principal Findings

This research on social media was, as far as we know, the first to examine the role of psychological distance in real-time Web-based responses to an approaching epidemic. Whereas previous studies have examined how social media relates to epidemic curves and classical media, the influence of psychological distance and severity of key events had not been explored. Current findings suggest that CLT may be a useful framework to increase understanding of public response to epidemic outbreaks. Even though the vast majority of Ebola cases occurred in West Africa and only a few suspected cases appeared near the Netherlands, findings showed that public attention for Ebola did not coincide well with the epidemic curve. As hypothesized, public attention for Ebola and expressions of fear for Ebola mostly responded to psychologically close events. Especially “fear for self” responded to increases in psychological closeness of the Ebola outbreak. Overall, findings suggest that events occurring in psychologically far regions of the world do not automatically capture public attention, even if the events are very severe.

The findings extend previous CLT findings in several ways. First, whereas previous research showed the effects of psychological distance in offline contexts [16,26,28], this has not been fully investigated in relation to Web-based communication contexts. It has been suggested that Web-based environments permit more confounded relationships for psychological distance dimensions (eg, with videoconferencing one can interact with spatially distant others in real time) compared with offline contexts [36]. Whether spatial distance is still important in Web-based contexts is a debated topic [37-39]. Therefore, this research contributes to CLT literature by demonstrating influence of two dimensions of psychological distance; namely spatial and social distance in Web-based expressions of public attention and fear. Second, our findings complement several lines of research that employ CLT framework to influence psychological distance perceptions in order to change behavioral intentions regarding important events. For instance, in climate change and distant suffering literature, psychological distance is suggested to be a barrier to engagement and behavioral intentions to act [13,29,40,41]. Our findings lend evidence to the application of CLT framework, and more specifically, the potential of rendering a situation psychologically close in order to increase attention to it.

The higher percentage of tweets expressing fear for self for neighboring countries compared with the Netherlands can be explained by the few confirmed cases of Ebola reported in, for example, Leipzig (Germany). Since both areas can be regarded as psychologically close in this study, following CLT we can reason that events in these areas were represented concretely [11], and perceived as truer [25] and more memorable [42] when compared with events taking place in distant locations. Whereas no overall relationship was found with regard to fear for others, a striking difference was observed between tweets involving North America and West Africa. Tweets about Ebola in North America (socially closer to the Netherlands) had the highest percentage fear for others, whereas tweets about Ebola in West Africa (spatially closer) had the lowest, in spite of extreme differences in reported cases between these regions. This suggests that people perceive psychologically far risks especially as more relevant for socially closer others. Social distance in this context might be more informative than spatial distance. Whereas this was not explicitly tested in this study, we can reason that people may have more knowledge about socially close others (eg, North America compared with West Africa) and can therefore imagine their situation more concretely and feel fear for those others. Subsequently, the findings extend previous studies on social media and health by showing that psychological distance not only affects perceptions of objects and events [20,21,23] but also determines (Web-based) public attention for an event and real-time expressions of fear (for self). The responses of the Twitter users might not be representative for the entire Dutch population, but the Netherlands have nevertheless a relatively large number of Twitter users compared with other countries, providing an adequate indication of the actual opinion and sentiment in the Netherlands.

In this research, and in line with previous studies [8,9], increases in attention for Ebola cooccurred with some severe real-world events, such as the World Health Organization (WHO) reports that Ebola formed a global threat. Yet, we can add that not all severe events evoked fear; public attention and fear overall responded strongest to proximity. The fact that attention and fear for Ebola in the Netherlands reached its peak around the time that Ebola started crossing the Mediterranean Sea and Atlantic Ocean suggests that the crossing of psychological boundaries may trigger sudden—rather than gradual—changes in proximity perceptions. Although severity of events played only a minor role in this research, it may play a role in and across other (closer) epidemics. Attention for Ebola was relatively minor compared with outbreaks as H1N1 [9], most likely because Ebola did not reach the Netherlands during the examined time frame. Future research should examine other epidemics that actually reached the Netherlands. The severity of an Ebola infection which takes place in a location where one has limited knowledge may be harder to visualize than an epidemic taking place in own country and thus, lead to more abstract mental representations [11]. Further research can test this possibility by examining differences in language abstraction (ie, the use of abstract vs concrete language) as a function of distance and severity [25].

Conclusions

Even though humans may care morally and rationally for the tragedies and suffering of others, psychological distance of events exerts important boundaries on what attracts people’s attention and triggers their emotions. These findings point to the limits of the human condition—limits that could be taken into account when communicating about human tragedies. The use of gripping concrete stories may be especially important to bring psychologically far news events to readers’ attention.

Acknowledgments

We thank L van Dorland and M van der Horst for their assistance in data coding.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. 2016 Dec 13. Ebola (Ebola Virus Disease) URL: https://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/previous-case-counts.html [accessed 2016-12-13] [WebCite Cache ID 6mimZMDIU]
2. Adler J. International Business Times. 2014 Dec 13. Ebola panic 2014: the fears and the real lessons learned URL: http://www.ibtimes.com/ebola-panic-2014-fears-real-lessons-learned-1766952 [accessed 2016-12-13] [WebCite Cache ID 6mimovV4m]
3. NOS. 2014 May 21. Hond ebolapatiënt wordt afgemaakt URL: http://nos.nl/artikel/707441-hond-ebolapatient-wordt-afgemaakt.html [accessed 2016-12-13] [WebCite Cache ID 6mimvLXFq]
4. Lamb A, Paul MJ, Dredze M. CMCI. 2013 Jun. Separating fact from fear: tracking flu infections on twitter URL: http://cmci.colorado.edu/~mpaul/files/naacl13flu-final.pdf [accessed 2017-05-23] [WebCite Cache ID 6qfWxtDrb]
5. 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 May;6(5):e19467. [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
6. Stewart A, Romano S, Kanhabua N, Di Martino S, Siberski W, Mazzeo A, et al. Arxiv. 2016 Nov 10. Why is it difficult to detect sudden and unexpected epidemic outbreaks in twitter? URL: https://arxiv.org/abs/1611.03426 [accessed 2017-05-23] [WebCite Cache ID 6qfRyL9f8]
7. Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease detection or public opinion reflection? content analysis of tweets, other social media, and online newspapers during the measles outbreak in the Netherlands in 2013. J Med Internet Res 2015 May;17(5):5-12. [doi: 10.2196/jmir.3863]
8. Chew C, Eysenbach G. Pandemics in the age of twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov;5(11):e14118. [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
9. Klemm C, Das E, Hartmann T. Swine flu and hype: a systematic review of media dramatization of the H1N1 pandemic. J Risk Res 2014 Jun 20;19(1):e1. [doi: 10.1080/13669877.2014.923029]
10. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in twitter: a quantitative analysis. JMIR Public Health Surveill 2016 Jan 04;2(1):e1. [doi: 10.2196/publichealth.5059]
11. Liberman N, Trope Y. The psychology of transcending the here and now. Science 2008 Nov 21;322(5905):1201-1205. [doi: 10.1126/science.1161958]
12. Soderberg CK, Callahan SP, Kochersberger AO, Amit E, Ledgerwood A. The effects of psychological distance on abstraction. Psychol Bull 2015 May;141(3):525-548. [doi: 10.1037/bul0000005]
13. Spence A, Poortinga W, Pidgeon N. The psychological distance of climate change. Risk Analysis 2012;32(6):957-972. [doi: 10.1111/j.1539-6924.2011.01695.x]
14. Trope Y, Liberman N. Construal-level theory of psychological distance. Psychol Rev 2010 Apr;117(2):440-463. [doi: 10.1037/a0018963]
15. Liberman N, Trope Y, Stephan E. Psychological distance. In: Social psychology: Handbook of basic principles. New York, NY: Guilford Press; 2007:353-384.
16. Bar-Anan Y, Liberman N, Trope Y, Algom D. Automatic processing of psychological distance: evidence from a Stroop task. J Exp Psychol-Gen 2007 Nov;136(4):610-622 [FREE Full text] [doi: 10.1037/0096-3445.136.4.610] [Medline: 17999574]
17. Bar-Anan Y, Liberman N, Trope Y. The association between psychological distance and construal level: evidence from an implicit association test. J Exp Psychol-Gen 2006 Apr;135(4):609-622. [doi: 10.1037/0096-3445.135.4.609]
18. Ellithorpe ME, Brookes SE, Ewoldsen DR. So close and yet so far: construal level moderates cultivation effects. Media Psychol 2015 Sep 11;19(1):27-48. [doi: 10.1080/15213269.2015.1040126]
19. Katz SJ, Byrne S. Construal level theory of mobile persuasion. Media Psychol 2013;16(3):245-271. [doi: 10.1080/15213269.2013.798853]
20. Lutchyn Y, Yzer M. Construal level theory and theory of planned behavior: time frame effects on salient belief generation. J Health Commun 2011 Mar 01;16(6):595-606.
21. Kim DH, Sung YH, Lee SY, Choi D, Sung Y. Are you on timeline or news feed? the roles of facebook pages and construal level in increasing ad effectiveness. Comput Human Behav 2016 Apr;57:312-320. [doi: 10.1016/j.chb.2015.12.031]
22. Williams LE, Stein R, Galguera L. The distinct affective consequences of psychological distance and construal level. J Cons Res 2014 Nov 14;40(6):1123-1138. [doi: 10.1086/674212]
23. Davis JI, Gross JJ, Ochsner KN. Psychological distance and emotional experience: what you see is what you get. Emotion 2011 Apr;11(2):438-444. [doi: 10.1037/a0021783]

24. Chandran S, Menon G. When a day means more than a year: effects of temporal framing on judgments of health risk. J Cons Res 2004 Sep 01;31(2):375-389. [doi: 10.1086/422116]
25. Hansen J, Wänke M. Truth from language and truth from fit: the impact of linguistic concreteness and level of construal on subjective truth. Pers Soc Psychol Bull 2010 Oct 14;36(11):1576-1588. [doi: 10.1177/0146167210386238]
26. Wright S, Manolis C, Brown D, Guo X, Dinsmore J, Chiu C, et al. Construal-level mind-sets and the perceived validity of marketing claims. Marketing Letters 2011 Nov 03;23(1):253-261. [doi: 10.1007/s11002-011-9151-4]
27. Sherman SJ, Cialdini RB, Schwartzman DF, Reynolds KD. Imagining can heighten or lower the perceived likelihood of contracting a disease the mediating effect of ease of imagery. Pers Soc Psychol Bull 1985 Mar 01;11(1):1-127.
28. Wakslak CJ. The where and when of likely and unlikely events. Organ Behav Hum Decis Process 2012 Jan;117(1):150-157.
29. Evans L, Milfont TL, Lawrence J. Considering local adaptation increases willingness to mitigate. Global Environmental Change 2014 Mar;25:69-75. [doi: 10.1016/j.gloenvcha.2013.12.013]
30. Doré B, Ort L, Braverman O, Ochsner KN. Sadness shifts to anxiety over time and distance from the national tragedy in Newtown, Connecticut. Psychol Sci 2015 Mar 12;26(4):363-373 [FREE Full text] [doi: 10.1177/0956797614562218] [Medline: 25767209]
31. Reintjes R, Das E, Klemm C, Richardus JH, Keßler V, Ahmad A. “Pandemic Public Health Paradox”: time series analysis of the 2009/10 Influenza A / H1N1 epidemiology, media attention, risk perception and public reactions in 5 European countries. PloS one 2016 Mar 16;11(3):e0151258. [doi: 10.1371/journal.pone.0151258]
32. Harcup T, O'Neill D. What is news? Galtung and Ruge revisited. Journal Stud 2001 Dec 12;2(2):261-280. [doi: 10.1080/14616700118449]
33. Tjong Kim Sang E, van den Bosch A. Dealing with big data: the case of twitter. CLIN J 2013;3:121-134 [FREE Full text]
34. Ji X, Chun SA, Geller J. Monitoring public health concerns using twitter sentiment classifications. : IEEE; 2013 Presented at: Healthcare Informatics (ICHI), 2013 IEEE International Conference on; September 09-11, 2013; Philadelphia, PA p. 335-344. [doi: 10.1109/ICHI.2013.47]
35. Fujita K, Henderson MD, Eng J, Trope Y, Liberman N. Spatial distance and mental construal of social events. Psychol Sci 2006 Apr 01;17(4):278-282.
36. Norman E, Tjomsland H, Huegel D. The distance between us: using construal level theory to understand interpersonal distance in a digital age. Frontiers in Digital Humanities 2016 Jun 20;3:e1. [doi: 10.3389/fdigh.2016.00005]
37. Cumiskey KM. Mobile symbiosis: A precursor to public risk-taking behavior. In: Mobile communication: Bringing us together and tearing us apart. New Brunswick and London: Transaction Publishers; 2011:17-36.
38. Guadagno RE, Muscanell NL, Rice LM, Roberts N. Social influence online: the impact of social validation and likability on compliance. Psychol Pop Media Cult 2013 Jan;2(1):51-60. [doi: 10.1037/a0030592]
39. Kaltenbrunner A, Scellato S, Volkovich Y, Laniado D, Currie D, Jutemar EJ, et al. Far from the eyes, close on the web: impact of geographic distance on online social interactions. In: Proceedings of the 2012 ACM workshop on Workshop on online social networks. New York, NY: ACM; 2012 Aug 17 Presented at: WOSN; 2012; Helsinki, Finland p. 19-24. [doi: 10.1145/2342549.2342555]
40. Lorenzoni I, Nicholson-Cole S, Whitmarsh L. Barriers perceived to engaging with climate change among the UK public and their policy implications. Global environmental change 2007;17(3-4):445-459. [doi: 10.1016/j.gloenvcha.2007.01.004]
41. McGraw AP, Warren C, Williams LE, Leonard B. Too close for comfort, or too far to care? finding humor in distant tragedies and close mishaps. Psychol Sci 2012 Aug 31;23(10):1215-1223. [doi: 10.1177/0956797612443831]
42. Sherer M, Rogers RW. The role of vivid information in fear appeals and attitude change. J Res Pers 1984 Sep;18(3):321-334. [doi: 10.1016/0092-6566(84)90016-3]

Abbreviations

ECM: error correction model
VECM: vector error correction model

Edited by G Eysenbach; submitted 22.12.16; peer-reviewed by E Tambo, C Freifeld, D Becerra; comments to author 26.01.17; revised version received 09.03.17; accepted 30.03.17; published 13.06.17.

Please cite as:
van Lent LGG, Sungur H, Kunneman FA, van de Velde B, Das E
Too Far to Care? Measuring Public Attention and Fear for Ebola Using Twitter
J Med Internet Res 2017;19(6):e193
URL: http://www.jmir.org/2017/6/e193/
doi:10.2196/jmir.7219
PMID:28611015

Original Paper

The Role of Web-Based Health Information in Help-Seeking Behavior Prior to a Diagnosis of Lung Cancer: A Mixed-Methods Study

Julia Mueller[1,2,3], BSc, MSc; Caroline Jay[3], BA (Hons), MSc, PhD; Simon Harper[3], MPhil, MEng, PhD; Chris Todd[1,2], BA (Hons), MA, PhD

[1]School of Health Sciences, University of Manchester, Manchester, United Kingdom
[2]Manchester Academic Health Science Centre, Manchester, United Kingdom
[3]School of Computer Science, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Julia Mueller, BSc, MSc
School of Health Sciences
University of Manchester
Kilburn Building, LF1
Oxford Road
Manchester, M13 PL
United Kingdom
Phone: 44 161 275 6239
Fax: 44 161 306 7894
Email: julia.mueller@manchester.ac.uk

Abstract

Background: Delays to diagnosis in lung cancer can lead to reduced chance of survival, and patients often wait for several months before presenting symptoms. The time between first symptom recognition until diagnosis has been theorized into three intervals: symptom appraisal, help-seeking, and diagnostic interval (here: “pathway to diagnosis”). Interventions are needed to reduce delays to diagnosis in lung cancer. The Web has become an important lay health information source and could potentially play a role in this pathway to diagnosis.

Objective: Our overall aim was to gain a preliminary insight into whether Web-based information plays a role in the pathway to diagnosis in lung cancer in order to assess whether it may be possible to leverage this information source to reduce delays to diagnosis.

Methods: Patients diagnosed with lung cancer in the 6 months before study entry completed a survey about whether (and how, if yes) they had used the Web to appraise their condition prior to diagnosis. Based on survey responses, we purposively sampled patients and their next-of-kin for semistructured interviews (24 interviews; 33 participants). Interview data were analyzed qualitatively using Framework Analysis in the context of the pathway to diagnosis model.

Results: A total of 113 patients completed the survey (age: mean 67.0, SD 8.8 years). In all, 20.4% (23/113) reported they or next-of-kin had researched their condition online before the diagnosis. The majority of searches (20/23, 87.0%) were conducted by or with the help of next-of-kin. Interview results suggest that patients and next-of-kin perceived an impact of the information found online on all three intervals in the time to diagnosis. In the appraisal interval, participants used online information to evaluate symptoms and possible causes. In the help-seeking interval, the Web was used to inform the decision of whether to present to health services. In the diagnostic interval, it was used to evaluate health care professionals’ advice, to support requests for further investigation of symptoms, and to understand medical jargon. Within this interval, we identified two distinct subintervals (before/after relevant diagnostic tests were initiated), in which the Web reportedly played different roles.

Conclusions: Because only 20.4% of the sample reported prediagnosis Web searches, it seems the role of the Web before diagnosis of lung cancer is at present still limited, but this proportion is likely to increase in the future, when barriers such as unfamiliarity with technology and unwillingness to be informed about one’s own health are likely to decrease. Participants’ perceptions suggest that the Web can have an impact on all three intervals in the pathway to diagnosis. Thus, the Web may hold the potential to reduce delays in the diagnostic process, and this should be explored in future research and interventions. Our results also suggest a division of the diagnostic interval into two subintervals may be useful.

(J Med Internet Res 2017;19(6):e189) doi:10.2196/jmir.6336

KEYWORDS

help seeking; online health information; health information seeking; lung cancer; symptom appraisal

Introduction

Lung cancer is the leading cause of cancer deaths worldwide [1]. Low survival rates for lung cancer have been linked to delays to diagnosis [2]; the majority of patients are diagnosed at advanced disease stages, which decreases chance of survival [3].

The route from symptom recognition to diagnosis and commencement of treatment has been theorized into four intervals by Walter et al [4] in a model of pathways to treatment (Figure 1). In the first interval, the “appraisal interval,” an individual appraises and interprets bodily changes. This is followed by the “help-seeking interval,” in which the individual decides whether to consult a health care professional about the bodily changes [5]. The following “diagnostic interval” involves appraisal by health care professionals, investigations, referrals, and appointments. In the event of a diagnosis, the “pretreatment interval” then commences, which involves planning and scheduling of treatment. The length of these intervals can be influenced by disease factors (eg, site, size, growth rate), health care provider and system factors (eg, access to resources, health care policy), and patient factors (eg, psychosocial factors).

Here we focus on the three intervals leading up to diagnosis (appraisal, help-seeking, and diagnostic; Figure 1) because low survival rates in lung cancer have been linked to delays to diagnosis [2,6]. We refer to these three intervals as the “pathway to diagnosis.”

We focus on patient factors because research has shown that people with lung cancer often experience symptoms for several months before presenting to health services [7-9]. Research suggests lack of knowledge about lung cancer symptoms is one of the biggest barriers to help-seeking [7,9-14]. Furthermore, symptoms are often masked by preexisting comorbidities that have similar symptoms, making it difficult for the patient to distinguish between existing and new symptoms [12,15]. Fear of being diagnosed with cancer and fatalistic beliefs about treatability of lung cancer may also impede help seeking [8,12].

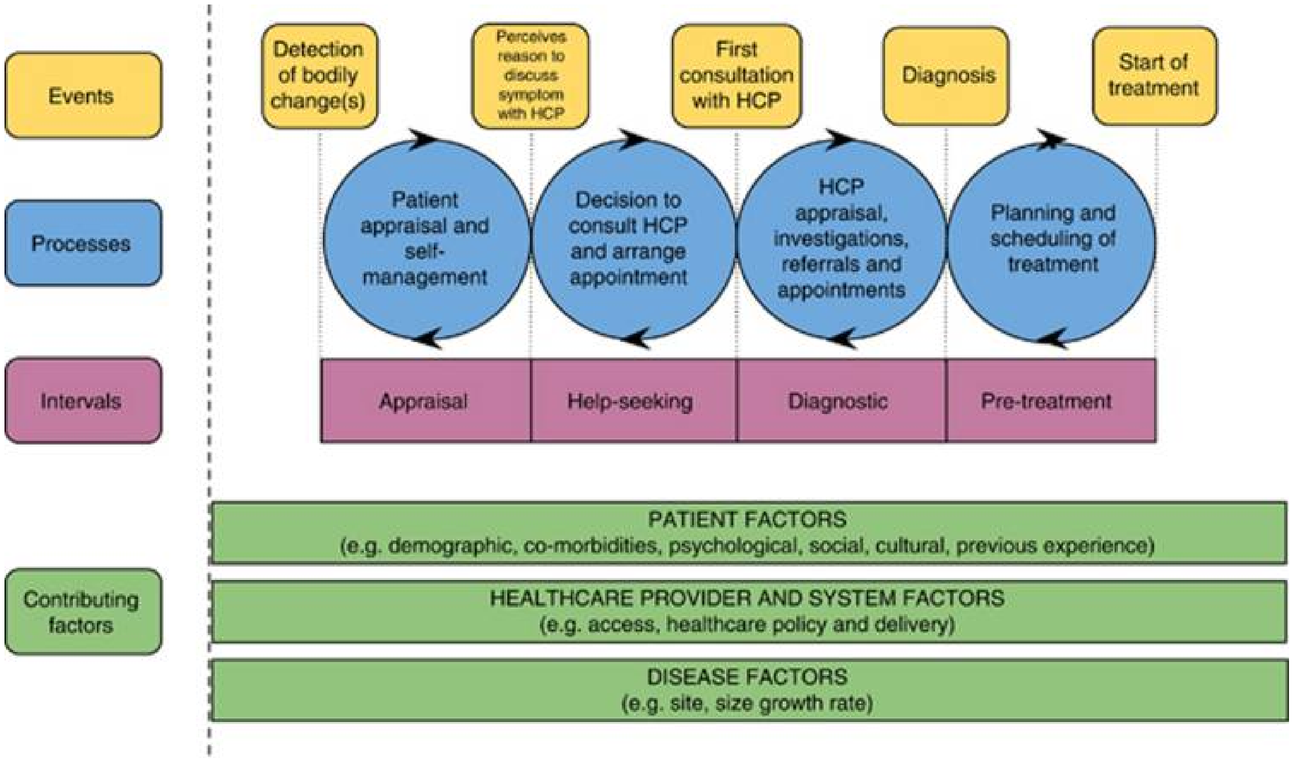
Lung and colorectal cancer patients have been shown to be proactive when appraising symptoms [14] and health-related Web use has been documented in various cancer populations, such as lung, colorectal, prostate, testicular, breast, cervical, and bowel cancer [16-18]. This suggests the Web could play a role in the time before cancer diagnosis (eg, if people with cancer search the Web for information to appraise their symptoms). Although evidence indicates that people with lung cancer do access the Web [18,19], the proportion of Web users is likely to be low because lung cancer patients tend to be older (>70 years) and to have lower education levels and socioeconomic status [1], and these factors are related to low levels of health-related Web use [20]. Overall, there is a growth of health information on the Web and an increasing tendency for individuals to seek health information online [21]. However, little is known about how people make use of this source prior to diagnosis because most research focuses on Web usage after patients have been diagnosed.

Due to the scarcity of previous research on this topic, an exploratory approach was required to gain a preliminary understanding of the potential role of online health information during the time leading up to a lung cancer diagnosis. In this study, we aimed to gain this preliminary understanding by exploring patients’ own retrospective accounts of how they remember the events leading up to their diagnosis, with particular focus on the perceived impact of Web searches on this process. Previous research has shown that family members sometimes conduct Web searches on behalf of patients [17] and that family members play an important role in lung cancer patients’ help-seeking behavior [7,8,11,22]. Therefore, we also aimed to explore accounts of next-of-kin of patients and whether they assisted patients with online searches or conducted searches on their behalf.

Our overall aim was to gain a preliminary, exploratory insight into whether Web-based information plays a role in the pathway to lung cancer diagnosis. This is important because, if the Web is found to play a role, it may be possible in the future to leverage this information source to reduce delays to diagnosis. To meet this aim, we addressed three research questions:

1. What proportion of people with lung cancer (or their family/friends) retrospectively report researching their condition online prior to diagnosis?
2. In cases in which prediagnosis Web searches take place, how do individuals perceive the impact of the information they find on their pathway to diagnosis?
3. What are possible barriers to using the Web prediagnosis?

Figure 1. Model of pathways to treatment (from Walter et al [4]).



Methods

Design

Our research questions required the combination of quantitative and qualitative methods. Quantitative methods were used to establish the proportion of lung cancer cases in which prediagnosis Web searches took place (question 1). Qualitative methods were used to explore individuals' perceptions of the impact their Web searches had on the pathway to diagnosis, as well as barriers that might prevent individuals from accessing the Web for health information prediagnosis (questions 2 and 3). Finally, mixed methods were required because a survey was needed to screen for relevant individuals for interview because we expected low levels of Web use among lung cancer patients.

Thus, this study consisted of (1) a cross-sectional, retrospective survey and (2) a qualitative interview study with a subsample of the survey participants.

Participants and Recruitment

We recruited recently diagnosed lung cancer patients to explore patients' retrospective accounts of the events leading up to their diagnosis. Participants were recruited from outpatient clinics at two large university hospitals in the northwest of England between July 2014 and March 2015. Patients were eligible if they (1) had received a lung cancer diagnosis in the 6 months prior to study entry, (2) had sufficient English language to complete the questionnaire, (3) were able to consent, and (4) reported experiencing at least one symptom before diagnosis. Patients whose diagnosis was more than 6 months before study entry were excluded to reduce recall bias [17].

Participants were sampled for interviews purposively based on questionnaire responses regarding (1) whether the Web had been used prior to diagnosis and (2) sociodemography (age, gender, smoking status) because these sociodemographic factors have been shown to be related to Web use [23]. We included both Web users and non-Web users to gain insight into reasons for and against using the Web prior to diagnosis. Next-of-kin were invited to participate in interviews because they tend to be involved in health information seeking [24] and have been shown to play an important role in lung cancer patients' help seeking [7,8,11,22]. We recruited next-of-kin who engaged in Web searches prior to diagnosis on the patient's behalf and those who did not. Data collection continued until saturation was reached (ie, when no new evidence for theoretical points emerged and we began to note similar accounts recurring) [25].

Procedure

We approached patients attending outpatient clinics, who had previously been identified from clinical notes as potentially eligible. Following informed consent, we provided participants with a questionnaire and stamped-addressed return envelope, with the option of completing the questionnaire in clinic (with the researchers, if they wished) or at home. A subset of consenting participants was selected purposively to participate in follow-up interviews, which were conducted in clinic.

Measures

Questionnaire

The paper-based questionnaire took 10 to 15 minutes to complete. Questions were standardized and assessed (1) whether the patient and/or a family member/friend had used the Web prior to diagnosis to help understand the symptoms/condition, (2) which symptoms were experienced before the diagnosis, (3) details on Web searches conducted prior to diagnosis if applicable (who conducted the search, search engine and search terms used, websites accessed), (4) information on habitual Web/technology use (whether the Internet is ever used; if yes, number of hours during a typical week), and (5) sociodemographic information (age, sex, education level, and employment status).

The development of the questionnaire was informed by previous literature on help-seeking behavior and Web-searching behavior [18,23,26], medical reference works [27,28], discussion with a Patient and Public Involvement group for cancer and palliative care, as well as brainstorming within the research group.

Interviews

Interviews were semistructured with open-ended questions and standardized prompts. The interview topic guide covered:

1. Symptom experience prior to diagnosis, with a focus on motivators and barriers to seeking help;
2. Web searches conducted prior to diagnosis (if applicable), with a focus on perceived impacts on the pathway to diagnosis (eg, the decision of whether to present to health services); and
3. Reasons for and against using the Web prior to diagnosis.

Clinical Records

Following consent, type of lung cancer and smoking status were obtained from patient records.

Analysis

Quantitative Analysis

We analyzed questionnaire data descriptively using IBM SPSS version 22 to calculate percentages, means, and standard deviations. For proportions, we calculated 95% confidence intervals using Confidence Interval Analysis (CIA) version 2.2.0 [29] as an indication of the variability of the results and to facilitate comparisons. Group differences in continuous variables were tested using the nonparametric Mann-Whitney U test, and associations between categorical variables were tested using Fisher's exact test, with Cramer's V computed to assess effect size.

Qualitative Analysis

Interviews were audio-recorded, transcribed verbatim, and organized using QSR NVivo10. Framework Analysis [25] was used to identify recurring and important themes in the data. Our analysis involved the following five stages [30,31]:

Familiarization With the Data

Familiarization was achieved by repeatedly reading all interview transcripts and noting recurring topics.

Development of a Theoretical Framework

A broad framework of topics was developed to organize the data, based on the interview protocol as well as recurring topics identified in step 1. The topics were then sorted and grouped under broad categories to create a hierarchical structure of topics and subtopics.

Indexing Data

The framework was then applied to the data by using NVivo to label transcript sections according to the topics occurring in each section. This was undertaken by at least two independent researchers and any discrepancies discussed until consensus was reached.

Summarizing Data in Thematic Charts

A matrix was created within NVivo10 [32] for each topic, with participants in the rows and subtopics in the columns. Transcript sections were then summarized into the relevant cells, keeping as close to participants' original wording as possible. To illustrate, an excerpt from a framework matrix is provided in Multimedia Appendix 1.

Synthesizing Data by Mapping and Interpreting

Matrices were next explored by comparing cells across participants and within participants to identify similarities or differences in how participants described their experiences. This facilitated identification of recurring themes and links between themes. Themes were discussed in the research group until consensus was reached. To aid interpretation, we categorized participants' reported Web searches according to the interval in the pathway to diagnosis [4] in which they occurred. Searches were assigned to the appraisal interval if they took place before the searcher perceived a reason to present the symptoms to a health care professional. They were classed as help-seeking interval if they took place between perceiving a reason to present symptoms and first consultation. Searches were assigned to the diagnostic interval if they took place after first consultation, but before a diagnosis was given.

Results

Survey

Sample Description

Between July 2014 and March 2015, 199 patients were identified as eligible and 122 consented (61.3%). Nine participants were excluded after consent because they had not experienced symptoms prior to diagnosis (it was not possible to discern this in advance from clinical records); therefore, 113 participants were included in the final sample (Figure 2). The mean age was 67.0 (SD 8.8) years and ranged from 42 to 88 years. The majority were male (56.6%, 64/113), retired (65.5%, 74/113), former smokers (69.0%, 78/113), and reported educational attainment below university level (82.3%, 93/113); 48.7% (55/113) had non-small cell lung cancer (Table 1).

Approximately half of all participants had an Internet connection (51.3%, 58/113) and 61.1% (69/113) had used an Internet device at some point; 23.0% (26/113) did not own any Internet device (Table 2).

Table 1. Participant demographic details (N=113).

Sociodemographic variables	Parameters
Age (years), mean (SD)	
Male	66.7 (8.7)
Female	67.1 (8.9)
Total	67.0 (8.8)
Gender, n (%)	
Male	64 (56.6)
Female	49 (43.4)
Employment status, n (%)	
Employed	24 (21.2)
Unemployed	12 (10.6)
Retired	74 (65.5)
Missing	3 (2.7)
Education level, n (%)	
No formal education	16 (14.2)
Secondary or high school	19 (16.8)
GCSE or equivalent	22 (19.5)
A levels or equivalent	4 (3.5)
Vocational qualification	22 (19.5)
Professional qualification	10 (8.8)
University degree	6 (5.3)
Missing	14 (12.4)
Type of lung cancer, n (%)	
Non-small cell lung cancer	55 (48.7)
Small cell lung cancer	44 (38.9)
Combined	1 (0.9)
Missing	13 (11.5)
Smoking status, n (%)	
Never-smoker	3 (2.7)
Former smoker	78 (69.0)
Current smoker	25 (22.1)
Missing	7 (6.2)

Table 2. Habitual Web/technology use among participants (N=113).

Item	n (%)
Internet connection at home	58 (51.3)
Use of a computer	60 (53.1)
Owning a computer	73 (64.6)
Use of a Web-enabled mobile phone	24 (21.2)
Owning a Web-enabled mobile phone	40 (35.4)
Use of a tablet	34 (30.1)
Owning a tablet	38 (33.6)
Ever use any Internet devices	69 (61.1)
Not owning any Internet devices	26 (23.0)

Figure 2. Participant recruitment flow diagram (STROBE diagram).

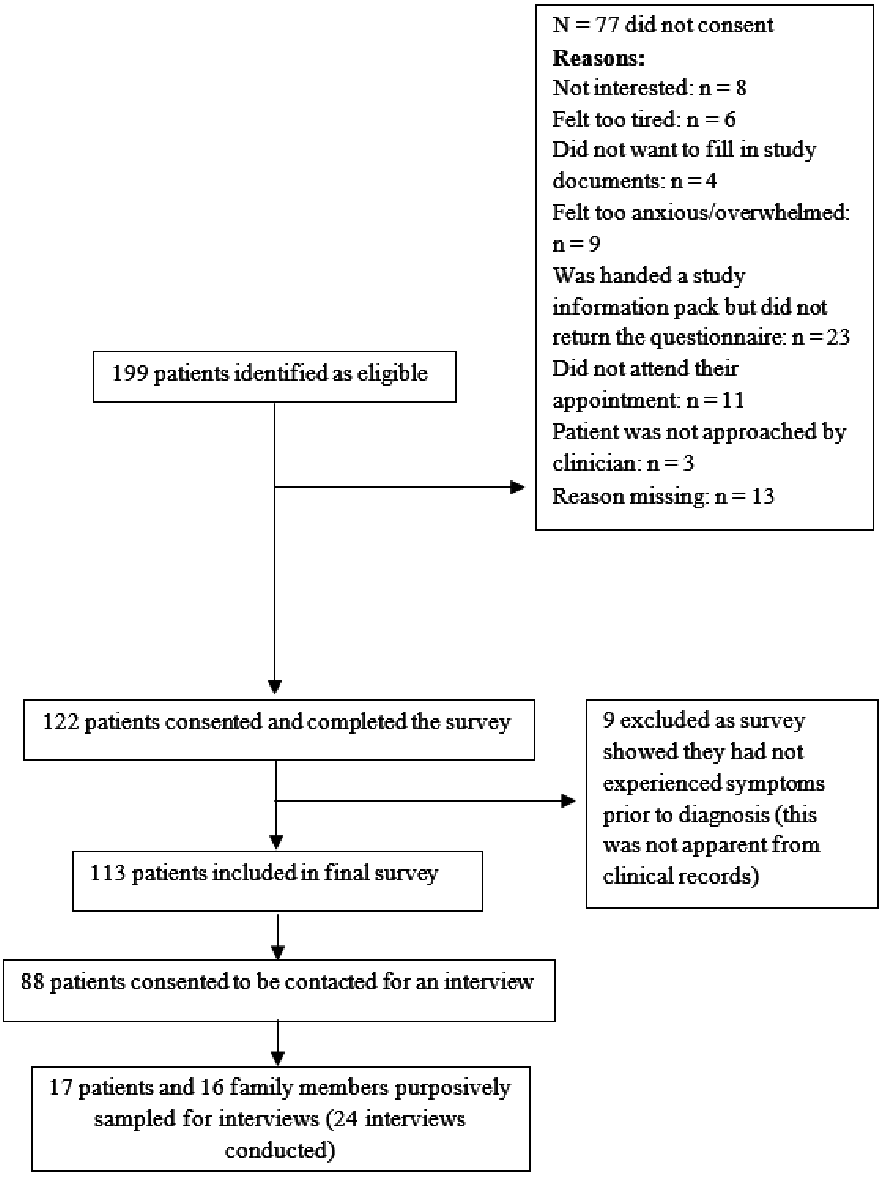
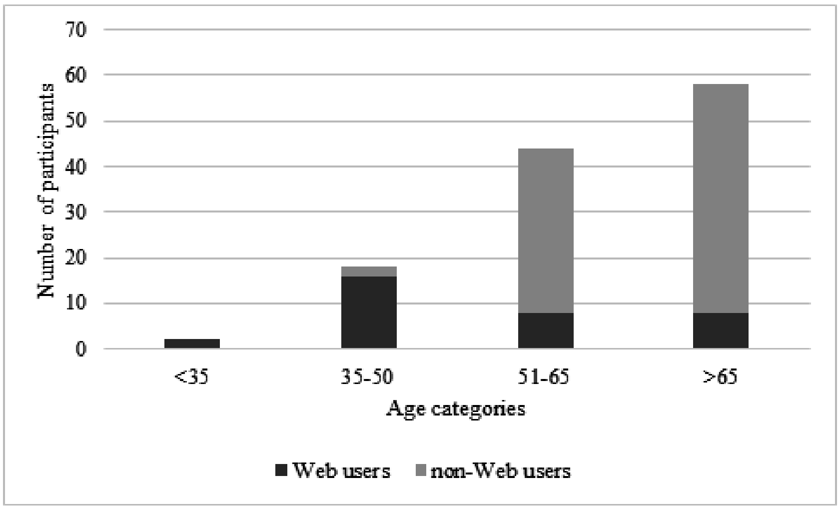


Figure 3. Age distribution of Web users (those who used the Web prior to diagnosis to help them understand symptoms) and non-Web users.



Proportion of People With Lung Cancer (or Their Family/Friends) Who Reported Web Searches Prior to Diagnosis

Of the sample, 20.4% (23/113, 95% CI 12.9%-27.8%) reported they, or a family member/friend, researched symptoms online prior to diagnosis. Seven of 113 (6.2%, 95% CI 1.8%-10.6%) stated they researched their symptoms themselves, four of these with the help of a family member/friend. Family/friends conducted online searches on the patient's behalf in another 16 cases (14.2%, 95% CI 7.7%-20.6%), thus the majority of searches (20/23, 87.0%) involved family/friends. Although 23 people reported the Web was used in their case, some reported several Web users (eg, a spouse and grandchild). Thus, in total, we identified 31 Web users: 7 patients, 7 spouses/partners, 12 sons/daughters, 2 sons-in-law/daughters-in-law, 1 grandchild, 1 nephew, and 1 friend. The age distribution of Web users is shown in Figure 3.

Description of Web Searches

Of the 23 participants who reported Web searches prior to diagnosis, 20 reported that Google was used to search; the rest did not know which search engine was used. The majority (19/23) reported using the NHS Direct website. Other websites included WebMD (7/23), patient.co.uk (5/23), Yahoo Health (2/23), and Netdoctor (1/23). Two participants reported visiting discussion forums.

Twenty-one participants reported search terms used (Table 3). Eight participants used symptoms as search terms (eg, “persistent cough”) and five used possible causes/conditions, such as “throat cancer” or “stopping smoking.” Three people used investigative test results that had been communicated to them by health professionals before they had received a final diagnosis (eg, “pleural effusion”). In these three cases, the Web searches took place before a diagnosis was given, but after some investigation of symptoms had been initiated. Five participants used combinations of these (eg, “stomach cancer and weight loss”). Overall, five participants reported that “lung cancer” was included in their search.

Web Searches and Symptoms

The majority (67.3%, 76/113) of participants reported experiencing a cough prior to diagnosis (Table 4). On average, participants reported mean 3.0 (SD 1.6) symptoms.

Table 3. Search terms used during Web searches prior to diagnosis (n=23).

Category and terms	Participants, n
Symptoms	
Persistent cough	4
Hoarse voice, hoarseness, croaky voice	3
Back pain, shoulder pain, lower back pain	2
Coughing up blood	1
Lump on neck	1
Dyspnea	1
Swollen face and neck	1
Recurrent chest infection	1
Pain in chest	1
Weight loss	1
Night sweats	1
Dry mouth	1
Conditions / possible causes	
Lung cancer, lung cancer symptoms	5
Cancer	1
Throat cancer	1
Stomach cancer	1
Myasthenia	1
Stopping smoking	1
Anxiety	1
Test results (communicated to the patient before the final diagnosis was given)	
Pleural effusion, fluid in lungs	1
Patch on vocal cords	1
Iron anemia	1
Collapsed lung	1

Table 4. Number and percentage of participants reporting respective symptoms (N=113).

Symptom	n (%)
Cough	76 (67.3)
Change in an existing cough	22 (19.5)
Hemoptysis	19 (16.8)
Dyspnea	59 (52.2)
Fatigue	51 (45.1)
Weight loss / loss of appetite	44 (38.9)
Shoulder/back pain	15 (13.3)
Chest pain	13 (11.5)

Table 5. Association between reporting symptoms and reporting Web use prior to diagnosis.

Reported symptoms	Web was used, n (expected count)	Web was not used, n (expected count)	χ^2_1	*P* (Fisher's exact test)	Cramer's *V*
Cough			0.7	.46	0.076
Yes	14 (15.6)	62 (60.4)			
No	9 (7.4)	27 (28.6)			
Change in an existing cough			0.1	>.99	0.074
Yes	3 (4.3)	18 (16.7)			
No	20 (18.7)	71 (72.3)			
Hemoptysis			0.3	.76	0.053
Yes	3 (3.9)	16 (15.1)			
No	20 (19.1)	73 (73.9)			
Dyspnea			0.7	.49	0.079
Yes	14 (12.2)	45 (46.8)			
No	9 (10.8)	43 (41.2)			
Fatigue			0.5	.64	0.065
Yes	9 (10.5)	42 (40.5)			
No	14 (12.5)	47 (48.5)			
Weight loss			3.6	.09	0.179
Yes	13 (9.0)	31 (35.0)			
No	10 (14.0)	58 (54.0)			
Shoulder/back pain			<0.1	>.99	0.005
Yes	3 (3.1)	12 (11.9)			
No	20 (19.9)	77 (77.1)			
Chest pain			0.2	>.99	0.046
Yes	2 (2.7)	11 (10.3)			
No	21 (20.3)	78 (78.7)			

The number of symptoms reported by the patient was not significantly associated with whether the Web was used prior to diagnosis (U=1041, Z= –0.19, P=.85). As Table 5 shows, having any particular symptom was not significantly associated with whether the Web was used prior to diagnosis.

Qualitative Interviews

Twenty-four interviews were conducted: 10 with patients, 7 with next-of-kin, and 7 with the patient and next-of-kin together. In total, 33 people were interviewed (n=19 female). Fourteen interviews involved Web searches prior to diagnosis, either by the patient or next-of-kin.

Perceptions of the Impact of Prediagnosis Web Searches on the Pathway to Diagnosis

In the following, we have grouped Web searches according to the time intervals in Walter et al’s model [4] during which they occurred. Within each interval, we explore participants’ perceptions of how their Web searches impacted the processes described in the model.

Appraisal Interval: Perceived Impacts on Appraisal and Self-Management

Some participants reportedly accessed the Web for information once they had perceived bodily changes, but had not yet decided to present these to a health care professional. Online information was used to identify possible causes of symptoms:

> *And [husband’s name] being how he is, he won’t go to the doctors anyway, so we did sort of self-diagnose, if you like.* [R27, wife of patient, 51-65 years]

Some participants described that the information they read online about symptoms changed the way they appraised their symptoms, causing them to view symptoms as more serious than before and, in some cases, even convincing them that the cause was lung cancer:

> *I just put in, to start with, shoulder pain, and lung cancer came up straight away...And that’s...I thought, oh, you know; I looked at it, and I thought, lung cancer? Crikey! Because I’d no idea that people got pain anywhere near there. And so I went on one website after another, after another, after another, just to try and read the symptoms to see if the*

> *symptoms were all the same on each site, and they were, basically.* [R18, daughter of patient, 51-65 years]

One participant described that the information she found online caused her to view her husband's symptoms as less serious. In this particular case, the searcher (a patient's spouse) entered the symptoms experienced by her husband as search terms, coupled with her hypothesis of what was causing the symptoms (ie, smoking cessation). For example, she reportedly searched for "stopping smoking and cough." Using this search strategy, she reportedly felt reassured by the information she found that the symptoms were caused by smoking cessation rather than a disease:

> *R27: "You see, the sweating all night and the coughing. We'd had a look online and his friends had told him, he'd stopped smoking. So that happens, you get insomnia, you can't sleep either and you're coughing a lot. And you're just bringing anything up that's been in your lungs for years. So we sort of left that at that, thinking that's what it was..."*
>
> *Interviewer: "And in your search, did you, at any point, come across any information about lung cancer?"*
>
> *R27: "I don't think I did...my specific search terms were, stopping smoking, so I kept putting stuff with smoking in. I didn't put night sweats, if you like, and get the whole...the amount of what would cause it, if you like."*
>
> *[...]*
>
> *Interviewer: "Could you just tell me a little bit more about how you went about your search?"*
>
> *R27: "So I probably put, stopping smoking and night sweats. And then what's come up about that. Stopping smoking and cough...how long after stopping smoking, will they carry on coughing?"* [R27, wife of patient, 51-65 years]

Help-Seeking Interval: Perceived Impact on the Decision to Consult a Health Care Professional

According to Walter et al's model [4], individuals in the help-seeking interval form the decision to consult a health care professional and make an appointment, and the interval is concluded when a first consultation takes place. Several participants in our interview study reportedly used online information to inform their decision on whether to present to a health care professional:

> *I kept thinking, this cough's not clearing. But like I said, it went on months and months...so that's when my son went on the Internet, and that's when he said, "Mum, Aunt [name] and Aunt [name] they're right, you need to go." And that's when I went.* [R22, patient, >65 years]

Some participants who conducted Web searches before first consultation with a health care professional perceived no impact on their decision making because they had reportedly already formed the decision to present to health services before they began their search:

> *I knew there was something wrong, that you had to go and see a doctor...The decision was made before I even googled it, yes.* [R4, patient, 51-65 years]

Diagnostic Interval: Perceived Impact on Health Care Professionals' Appraisals

In Walter et al's model [4], the diagnostic interval commences following first consultation and involves appraisal by a health care professional, investigations, referrals, and appointments. Within this interval, we identified two distinct subintervals at which Web searches took place: (1) after a health care professional had been consulted, but before relevant diagnostic tests (chest x-ray, CT scan) had been conducted, and (2) after relevant diagnostic tests had been conducted, but before a diagnosis had been communicated. Within these subintervals, Web searches had different perceived impacts.

Before Diagnostic Tests Were Underway

Participants who reported Web searches in this subinterval described presenting to health services multiple times without diagnostic tests to determine the cause of the symptoms. Participants reportedly turned to the Web because they felt dissatisfied with the advice they received from health care professionals:

> *So I went on the Internet, I think because she'd had four visits to the doctor and we weren't getting anywhere, so I went on to just see, you know, if I could find anything out really to give me an idea what else it could be other than an allergic reaction.* [R14, daughter of patient, 51-65 years]

Interviewees reportedly used the information found online to challenge their doctors' advice by suggesting other possible causes for symptoms and requesting further tests. Participants felt that their assertiveness in urging further investigation impacted on health care professionals' decisions to conduct diagnostic procedures:

> *So then I went online and I put in facial swelling and neck swelling and it said it could be an infection of the glands or the ducts. So that's why I asked the GP, when I went back with her on the fifth visit, could it be an infection in the glands or the ducts...I'm not saying he wouldn't have done it but I think the fact that I was with my mum and maybe being a little bit more assertive instigated him to maybe look a little bit further. Yeah, definitely...I wasn't rude but I was assertive, and it was only then that he investigated further and listened more closely to her chest.* [R14, daughter of patient, 51-65 years]

After Diagnostic Tests Were Underway

Participants often described a period of several weeks during which diagnostic tests were undertaken and results communicated to them, but they were not informed of what these test results could mean. During this period of uncertainty, participants reportedly conducted online searches to understand medical jargon and test results, and to identify possible causes:

> *He said, it looks like one of the lungs have collapsed, but obviously we need to go and see a specialist at the [university hospital], which is what we did. But prior to actually seeing a specialist, I started looking then, on what could cause a collapsed lung.* [R27, wife of patient, 51-65 years]

Perceived Barriers to Using the Web Prediagnosis

We also interviewed patients and next-of-kin of patients who had not accessed the Web prior to diagnosis in order to understand perceived barriers to prediagnosis Web use. This can help to understand whether the role of the Web in the pathway to lung cancer diagnosis may change in the future and if barriers are aspects that are likely to change or not.

Concern Over Unnecessary Worry and Fear

Of those who reported not researching their condition online prior to diagnosis, several reportedly avoided this because they were concerned that it could lead to unnecessary worry and fear:

> *Sometimes it can frighten the life out of you, you know what I mean? It's like when people used to buy the home medical directory and you'd got a headache and something else, when you looked it up, you've got everything under the sun.* [R8, husband of patient, >65 years]

Preferring Not to Know

Some participants stated that they wanted to know as little as possible about their health, preferring to leave decisions to health professionals:

> *No I don't like looking it up. I don't really like knowing unless I've got to.* [R21, patient, 51-65 years]

Believing Symptoms Trivial

One participant felt her symptoms were too mild and familiar to warrant further research:

> *I mean I like to know what's going on but I wouldn't research a tickly cough because I've been in that situation many times before so...* [R6 patient, patient, >65 years]

Unfamiliar With or Not Interested in Technology

Most of those who did not research their condition online were not comfortable using technology:

> *Interviewer: "Do you ever use the computer?"*
>
> *R24: "No, because I can't even, it takes me all this time to text and reply. I prefer a conversation, you know, ring somebody. You know, further than that, I just make a mess of everything."* [R24, patient, >65 years]

Some expressed disinterest in the use of technology:

> *I am computer illiterate...and I prefer it that way. Yeah. I used to use a computer when I worked, it's not that I can't, it's that I'm not interested; it's such a waste of time.* [R6, patient, 51-65]

A few participants wanted to access the Internet, but lacked the skill:

> *The lads have got it now and I think they, I think they are brilliant. I wish it's one of the things, I wish I could, but I've never gone onto the Internet.* [R21, patient, 51-65 years]

Discussion

This is the first study to explore prediagnosis Web searches among lung cancer patients. We found that approximately a fifth of the sample of lung cancer patients reported prediagnosis Web searches to research symptoms and help them understand their condition, with most searches conducted by next-of-kin. Furthermore, our analyses showed that patients and their next-of-kin perceived impacts of their prediagnosis Web searches on their pathways to diagnosis, including symptom appraisal, forming the decision to seek help, and interactions with health care professionals.

Our overall aim was to gain a preliminary, exploratory insight into whether Web-based information plays a role in the pathway to lung cancer diagnosis. To explore this role, we discuss subsequently (1) the proportion of people with lung cancer reporting prediagnosis Web searches, (2) perceived impacts of the Web searches on the pathway to diagnosis, and (3) what prevents people from accessing the Web and whether this is likely to change in future.

Proportion of People With Lung Cancer Reporting Prediagnosis Web Searches

We found that 20.4% (23/113) of our sample reported Web searches prior to diagnosis to help appraise symptoms or understand their condition. The majority of searches were conducted by or with the help of a family member. Although more than half (61.1%, 69/113) of the patients in the survey indicated having used the Internet in the past, and approximately half (51.3%, 58/113) reported having an Internet connection at home, only 6.2% (7/113, 95% CI 1.8%-10.6%) of patients reported researching their condition online themselves.

Only one other published study has examined the proportion of cancer patients who engage in prediagnosis Web searches. In a study with colorectal cancer patients, Thomson et al [17] found 25% (61/242, 95% CI 20%-31%) of patients had researched symptoms online themselves, not including patients whose family or friends searched on their behalf. Comparing our 95% confidence interval of 1.8% to 10.6% with Thomson et al's 20% to 31%, the proportion found in our sample of lung cancer patients is clearly lower. This may be due to our participants being older with lower education levels than those in the Thomson et al study [17] because these factors have been related to lower levels of health-related Web use [20].

Perceived Impact of Web Searches on the Pathway to Diagnosis

In our qualitative interview study, we explored how patients and their next-of-kin perceived the impact of prediagnosis Web searches on the events leading up to diagnosis. By mapping participants' accounts onto the model developed by Walter et al [4], we showed that participants perceived an influence of the information they found online on all three intervals leading up to diagnosis (appraisal, help-seeking, diagnostic).

Appraisal Interval

In the appraisal interval, participants reportedly used online information to assess the seriousness of their symptoms and to identify possible causes. Previous research has suggested a rising importance of online health information on symptom appraisal processes; in the United States, for example, more than a third of adults report having used online information to identify the cause of symptoms [20].

Our findings suggest that participants searching in the appraisal interval had differing experiences with Web searches, with some reporting that the information led them to believe their symptoms were serious, whereas others reported that the information reassured them that symptoms were not serious.

Our interview findings tentatively suggest that search strategies may play an important role in how online information affects appraisal of symptoms. The participant who reportedly felt reassured had conducted a hypothesis-driven search strategy by researching symptoms paired together with her hypothesized cause "stopping smoking." Therefore, search results were biased toward the hypothesized cause. Previous research [33] has suggested that searchers who use hypothesis-driven searches are prone to certain forms of bias, such as confirmation bias (starting with a hypothesis and confirming it) and premature termination bias (stopping after viewing only one topic). Further research on the differential effects of symptom-driven and hypothesis-driven searches is necessary to determine how search functions on health websites should be designed to enhance patients' ability to appropriately appraise symptoms.

Help-Seeking Interval

When the Web was used in the help-seeking interval, participants reported that online information was used to inform their decision of whether to present to health services, and several participants reported that it encouraged them to make an appointment with a health care professional. Previous research confirms that most "online diagnosers" subsequently seek a professional medical opinion [20,34]. In Thomson et al's [17] study with colorectal cancer patients, approximately a quarter of patients reportedly felt persuaded by the information found online to see a health professional. Furthermore, analyses of search engine log data have indicated that those who research symptoms online often subsequently show health care utilization intent (eg, by searching for clinics near their geographical area) [35].

Overall, previous research coupled with our results suggests that there may be a causal relationship between Web use and deciding to seeking help; this should be examined quantitatively in future research. Research in this area would be especially crucial for conditions such as lung cancer, in which earlier presentation to health services can maximize chances of survival.

Diagnostic Interval

Our analyses revealed two key findings regarding the diagnostic interval. Firstly, our results suggest that a division of this interval into two subintervals may be useful when examining the role of Web-based information. The first subinterval is the period from first consultation to the initiation of relevant diagnostic tests. The second subinterval begins with relevant diagnostic procedures and concludes with the final diagnosis. Our findings suggest the Web plays different roles in these two intervals.

When we examined searches that took place in the first sublevel, our findings suggest Web-based health information can empower patients and their families to appraise and challenge doctors' advice and request further diagnostic procedures. This is particularly interesting because efforts to reduce patient delays to diagnosis in lung cancer have focused on encouraging presentation to health services [36-38] (ie, the appraisal and help-seeking interval). Little attention has been paid to the role patients play in the diagnostic interval.

Recent years have seen a shift toward patient-driven health care, with patients increasingly interested in decision making [39]. Web-based health information has been associated with this shift by increasing patients' awareness of health professionals' fallibility and uncertainties in diagnoses [40]. Thus, with the help of Web-based health information, patients may play an increasingly important role during the diagnostic interval. This is particularly important because our results indicate that individuals turn to the Web when they are dissatisfied with advice received from health care professionals and when they experience delays in obtaining a diagnosis. If future research and interventions focus on how the Web can be leveraged to support patients in this role, delays to diagnosis may be reduced.

In the second sublevel, after diagnostic tests were initiated, the Web was used to facilitate understanding of medical terms. The use of medical jargon in consultations, dissatisfaction with doctors' communication skills, and the prevalence of low health literacy is well documented [41,42]. The majority of cancer patients prefer to be informed about their diagnosis [43]. Thus, the Web may prove a useful information resource before the diagnosis because it can facilitate understanding of medical jargon.

Barriers to Prediagnosis Web Searches: Current and Future Importance of the Web Prior to a Lung Cancer Diagnosis

We explored barriers to using the Web for health information prior to diagnosis. This can help to assess the extent of the role the Web can play in lung cancer patients' pathways to diagnosis and whether this role is likely to change in future.

The majority (79.6%, 90/113) of our sample reported not accessing the Web prior to diagnosis for a range of reasons. Although some of these barriers, such as perceived triviality of symptoms, may persist in the future, others are likely to change. For example, although some participants in our study reportedly preferred to defer to the doctor and not know details regarding their own health, engagement in health care is increasing and patients generally desire more detailed information from health professionals than they receive [43,44]. Furthermore, participants reported unfamiliarity with, or disinterest in, technology. This is unsurprising because the current cohort of those aged 65 and older, who are at highest risk for lung cancer, are less likely than any other age group to access the Internet [45]. Future lung cancer patients will be more familiar with the Web [46,47].

Recommendations and Future Research

The association between prediagnosis Web searches and length of intervals in the pathway to diagnosis should be assessed statistically with a larger sample size. However, we will first need a validated, reliable measure of time intervals leading up to diagnosis. To date, no validated measure exists, and measures used in previous research have considerable limitations [48].

Furthermore, strategies to leverage the Web to encourage early presentation to health services should be investigated in future research. For example, our analyses indicate that individuals turn to the Web when they experience difficulty communicating with health professionals. Future research should therefore explore how health websites can provide information that will help facilitate patients' communications with health professionals.

Our results suggest hypothesis-driven searches (search terms based on hypothesized conditions) and evidence-driven searches (search terms based on symptoms only) may have differing effects on how individuals evaluate symptoms. Future research should systematically investigate differential effects of hypothesis-driven and evidence-driven search strategies on symptom appraisal and subsequent help-seeking behavior. This will help to inform the development of Web-based symptom appraisal tools and search engine algorithms.

Limitations

Our findings relied on retrospective, self-reported measures of patients, pertaining to events that took place prior to diagnosis. Cancer patients' reports of the events leading up to diagnosis can be inconsistent [49]. Patients in our study had been diagnosed up to 6 months prior to study entry, and were asked to recall events before the diagnosis. Therefore, some of our measures may be subject to recall bias. For example, patients may not have been able to recall all search terms used during their prediagnosis Web searches. It was not feasible, however, to identify and recruit individuals with lung cancer prior to diagnosis.

Furthermore, patients may have been unaware of Web searches conducted by family/friends so this variable may be underreported in the survey because the survey was completed by patients. This could be addressed in future research by conducting a large-scale survey among patients and their next-of-kin.

Finally, as with all qualitative research, one must be cautious in generalizing from our purposive interview sample to the wider population of lung cancer patients.

Conclusions

Because only 20.4% of the sample reported prediagnosis Web searches, it seems that the role of the Web prior to a diagnosis of lung cancer is still limited at present, but this proportion is likely to increase in the future, when barriers such as unfamiliarity with technology and unwillingness to be informed about one's own health are likely to decrease.

Participants perceived an impact of their Web searches on symptom appraisal, the decision to present to health services, and on how they communicated with doctors and requested referrals to specialist care. This suggests using the Web prior to diagnosis may impact the appraisal, help-seeking, and diagnostic intervals referred to in Walter et al's model [4], and thus on the length of time until a diagnosis is made. Although a quantitative analysis will be required to assess the statistical association between Web usage prediagnosis and the length of time from symptom occurrence to diagnosis, our study highlights potential mechanisms of how Web-based health information may influence pathways to diagnosis and can thus help to inform design of future research. The Web as a health information source is here to stay and, if it is to be an effective tool for health care systems, websites should use evidence-based designs to help potential patients make appropriate decisions about seeking medical treatment.

Acknowledgments

This project was funded by a Medical Research Council (MRC) Doctoral Training Partnership studentship (Ref: 1354978) and by a President's Doctoral Scholar Award by the University of Manchester. Ethical approval was granted by the NHS Health Research Authority NRES Ethics Committee (Ref:14/SW/0128, IRAS Project ID 143412). We would like to thank the patients who took part and the clinicians who provided access to the clinics.

Authors' Contributions

JM, CT, SH, and CJ planned and designed the study. JM collected the data, carried out the analysis, and wrote the first draft of the paper. CT, SH, and CJ contributed to discussion of the content and analysis method. All authors reviewed and edited the manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Excerpt of a framework matrix.

[PDF File (Adobe PDF File), 268KB - jmir_v19i6e189_app1.pdf]

References

1. Torre LA, Siegel RL, Jemal A. Lung dancer statistics. Adv Exp Med Biol 2016;893:1-19. [doi: 10.1007/978-3-319-24223-1_1] [Medline: 26667336]
2. Holmberg L, Sandin F, Bray F, Richards M, Spicer J, Lambe M, et al. National comparisons of lung cancer survival in England, Norway and Sweden 2001-2004: differences occur early in follow-up. Thorax 2010 May;65(5):436-441. [doi: 10.1136/thx.2009.124222] [Medline: 20435867]
3. Zee YK, Eisen T. Survival from lung cancer in England and Wales up to 2001. Br J Cancer 2008 Sep 23;99 Suppl 1:S43-S46 [FREE Full text] [doi: 10.1038/sj.bjc.6604584] [Medline: 18813257]
4. Walter F, Webster A, Scott S, Emery J. The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. J Health Serv Res Policy 2012 Apr;17(2):110-118 [FREE Full text] [doi: 10.1258/jhsrp.2011.010113] [Medline: 22008712]
5. Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. Br J Cancer 2012 Mar 27;106(7):1262-1267 [FREE Full text] [doi: 10.1038/bjc.2012.68] [Medline: 22415239]
6. Radzikowska E, Roszkowski-Śliż K, Głaz P. The impact of timeliness of care on survival in non-small cell lung cancer patients. Pneumonol Alergol Pol 2012;80(5):422-429 [FREE Full text] [Medline: 22926903]
7. Corner J, Hopkinson J, Roffe L. Experience of health changes and reasons for delay in seeking care: a UK study of the months prior to the diagnosis of lung cancer. Soc Sci Med 2006 Mar;62(6):1381-1391. [doi: 10.1016/j.socscimed.2005.08.012] [Medline: 16168544]
8. Tod AM, Craven J, Allmark P. Diagnostic delay in lung cancer: a qualitative study. J Adv Nurs 2008 Feb;61(3):336-343. [doi: 10.1111/j.1365-2648.2007.04542.x] [Medline: 18197868]
9. Corner J, Hopkinson J, Fitzsimmons D, Barclay S, Muers M. Is late diagnosis of lung cancer inevitable? Interview study of patients' recollections of symptoms before diagnosis. Thorax 2005 Apr;60(4):314-319 [FREE Full text] [doi: 10.1136/thx.2004.029264] [Medline: 15790987]
10. McCutchan GM, Wood F, Edwards A, Richards R, Brain KE. Influences of cancer symptom knowledge, beliefs and barriers on cancer symptom presentation in relation to socioeconomic deprivation: a systematic review. BMC Cancer 2015 Dec 23;15:1000 [FREE Full text] [doi: 10.1186/s12885-015-1972-8] [Medline: 26698112]
11. Smith LK, Pope C, Botha JL. Patients' help-seeking experiences and delay in cancer presentation: a qualitative synthesis. Lancet 2005 Sep;366(9488):825-831. [doi: 10.1016/s0140-6736(05)67030-4]
12. Tod AM, Joanne R. Overcoming delay in the diagnosis of lung cancer: a qualitative study. Nurs Stand 2010 Apr;24(31):35-43. [doi: 10.7748/ns2010.04.24.31.35.c7690] [Medline: 20441033]
13. Shim J, Brindle L, Simon M, George S. A systematic review of symptomatic diagnosis of lung cancer. Fam Pract 2014 Apr;31(2):137-148 [FREE Full text] [doi: 10.1093/fampra/cmt076] [Medline: 24347594]
14. McLachlan S, Mansell G, Sanders T, Yardley S, van der Windt D, Brindle L, et al. Symptom perceptions and help-seeking behaviour prior to lung and colorectal cancer diagnoses: a qualitative study. Fam Pract 2015 Oct;32(5):568-577 [FREE Full text] [doi: 10.1093/fampra/cmv048] [Medline: 26099812]
15. Corner J, Brindle L. The influence of social processes on the timing of cancer diagnosis: a research agenda. J Epidemiol Community Health 2011 Jun;65(6):477-482. [doi: 10.1136/jech.2008.084285] [Medline: 21138896]
16. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 6;328(7439):564 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]
17. Thomson MD, Siminoff LA, Longo DR. Internet use for prediagnosis symptom appraisal by colorectal cancer patients. Health Educ Behav 2012 Oct;39(5):583-588 [FREE Full text] [doi: 10.1177/1090198111423941] [Medline: 21990571]
18. Peterson MW, Fretz PC. Patient use of the internet for information in a lung cancer clinic. Chest 2003 Feb;123(2):452-457. [Medline: 12576365]
19. White RW, Horvitz E. Evaluation of the feasibility of screening patients for early signs of lung carcinoma in web search logs. JAMA Oncol 2016 Nov 10;3(3):398-401. [doi: 10.1001/jamaoncol.2016.4911] [Medline: 27832243]
20. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health Online 2013 URL: http://pewinternet.org/Reports/2013/Health-online.aspx [accessed 2016-07-08] [WebCite Cache ID 6iqsuJyWn]
21. Chou WS, 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]
22. Braybrook DE, Witty KR, Robertson S. Men and lung cancer: a review of the barriers and facilitators to male engagement in symptom reporting and screening. J Men Health 2011 May;8(2):93-99. [doi: 10.1016/j.jomh.2011.03.002]
23. Valero-Aguilera B, Bermúdez-Tamayo C, García-Gutiérrez JF, Jiménez-Pernett J, Vázquez-Alonso F, Suárez-Charneco A, et al. Factors related to use of the Internet as a source of health information by urological cancer patients. Support Care Cancer 2012 Dec;20(12):3087-3094. [doi: 10.1007/s00520-012-1431-x] [Medline: 22415609]

24. Schook RM, Linssen C, Schramel FM, Festen J, Lammers E, Smit EF, et al. Why do patients and caregivers seek answers from the Internet and online lung specialists? A qualitative study. J Med Internet Res 2014;16(2):e37 [FREE Full text] [doi: 10.2196/jmir.2842] [Medline: 24496139]
25. Spencer L, Ritchie J, O'Connor W, Morrell G, Ormston R. Analysis in practice. In: Qualitative Research Practice: A Guide For Social Science Students and Researchers. London: Sage Publications Ltd; 2013:295-346.
26. Qureshi SA, Morkve O, Mustafa T. Patient and health system delays: health-care seeking behaviour among pulmonary tuberculosis patients in Pakistan. J Pak Med Assoc 2008 Jun;58(6):318-321 [FREE Full text] [Medline: 18988391]
27. Bourke SJ, Burns GP. Lecture Notes: Respiratory Medicine. Oxford: Wiley-Blackwell; 2011.
28. Dilworth JP, Baldwin DR. Resporatory Medicine: A Specialist Handbook. Singapore: Harwood Academic Publishers; 2001.
29. Altman D, Machin D, Bryant T, Gardner M. Statistics with Confidence: Confidence Intervals and Statistical Guidelines. London: British Medical Journal; 2000.
30. Ward DJ, Furber C, Tierney S, Swallow V. Using Framework Analysis in nursing research: a worked example. J Adv Nurs 2013 Nov;69(11):2423-2431. [doi: 10.1111/jan.12127] [Medline: 23517523]
31. Tierney S, Elwers H, Sange C, Mamas M, Rutter MK, Gibson M, et al. What influences physical activity in people with heart failure?: a qualitative study. Int J Nurs Stud 2011 Oct;48(10):1234-1243. [doi: 10.1016/j.ijnurstu.2011.03.003] [Medline: 21459380]
32. QSR International Pty Ltd. NVivo qualitative data analysis Software Version 10. 2012. URL: http://www.qsrinternational.com/nvivo-support/downloads/nvivo10-for-windows [WebCite Cache ID 6qHyLTEyp]
33. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]
34. 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]
35. White R, Horvitz E. From web search to healthcare utilization: privacy-sensitive studies from mobile data. J Am Med Inform Assoc 2013 Jan 1;20(1):61-68 [FREE Full text] [doi: 10.1136/amiajnl-2011-000765] [Medline: 22661560]
36. Ironmonger L, Ohuma E, Ormiston-Smith N, Gildea C, Thomson CS, Peake MD. An evaluation of the impact of large-scale interventions to raise public awareness of a lung cancer symptom. Br J Cancer 2015 Jan 06;112(1):207-216 [FREE Full text] [doi: 10.1038/bjc.2014.596] [Medline: 25461805]
37. Tustin N. University of Southern California Dissertations and Theses. 2012. Patient delay: an experimental study in reducing time taken to seek medical help for cancer [dissertation] URL: http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll127/id/480476 [accessed 2017-05-07] [WebCite Cache ID 6qHylJyVB]
38. Cancer Research UK. Be Clear on Cancer: lung cancer campaign URL: http://www.cancerresearchuk.org/health-professional/early-diagnosis-activities/be-clear-on-cancer/lung-cancer-campaign [accessed 2017-03-03] [WebCite Cache ID 6ogw1s5wK]
39. Pelzang R. Time to learn: understanding patient-centred care. Br J Nurs 2010;19(14):912-917. [doi: 10.12968/bjon.2010.19.14.49050] [Medline: 20647984]
40. Thompson AG. The meaning of patient involvement and participation in health care consultations: a taxonomy. Soc Sci Med 2007 Mar;64(6):1297-1310. [doi: 10.1016/j.socscimed.2006.11.002] [Medline: 17174016]
41. Thomas M, Hariharan M, Rana S, Swain S, Andrew A. Medical jargons as hindrance in doctor-patient communication. Psychol Stud 2014 Jun 24;59(4):394-400. [doi: 10.1007/s12646-014-0262-x]
42. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40245.x] [Medline: 15836552]
43. van Weert JC, Bolle S, van DS, Jansen J. Older cancer patients' information and communication needs: what they want is what they get? Patient Educ Couns 2013 Sep;92(3):388-397. [doi: 10.1016/j.pec.2013.03.011] [Medline: 23602863]
44. Al Qadire M. Jordanian cancer patients' information needs and information-seeking behaviour: a descriptive study. Eur J Oncol Nurs 2014 Feb;18(1):46-51. [doi: 10.1016/j.ejon.2013.09.007] [Medline: 24183255]
45. Zickuhr K, Madden M. Older Adults and Internet Use. Washington, DC: Pew Internet & American Life Project; 2012. URL: http://www.sainetz.at/dokumente/Older_adults_and_internet_use_2012.pdf [accessed 2016-07-08] [WebCite Cache ID 6iqt81HW0]
46. Chevalier A, Dommes A, Marquié J. Strategy and accuracy during information search on the Web: effects of age and complexity of the search questions. Comput Hum Behav 2015 Dec;53:305-315. [doi: 10.1016/j.chb.2015.07.017]
47. Couper MP, Singer E, Levin CA, Fowler FJ, Fagerlin A, Zikmund-Fisher BJ. Use of the Internet and ratings of information sources for medical decisions: results from the DECISIONS survey. Med Decis Making 2010;30(5 Suppl):106S-114S. [doi: 10.1177/0272989X10377661] [Medline: 20881159]
48. Andersen RS, Vedsted P, Olesen F, Bro F, Søndergaard J. Patient delay in cancer studies: a discussion of methods and measures. BMC Health Serv Res 2009 Oct 19;9:189 [FREE Full text] [doi: 10.1186/1472-6963-9-189] [Medline: 19840368]
49. Hess LM, Method MW, Stehman FB, Weathers TD, Gupta P, Schilder JM. Patient recall of health care events and time to diagnose a suspected ovarian cancer. Clin Ovarian Gynecol Cancer 2012 Jun;5(1):17-23. [doi: 10.1016/j.cogc.2012.04.001]

Edited by G Eysenbach; submitted 11.07.16; peer-reviewed by S Bolle, A Marcu, WC Su; comments to author 19.11.16; revised version received 22.01.17; accepted 02.03.17; published 08.06.17.

Please cite as:
Mueller J, Jay C, Harper S, Todd C
The Role of Web-Based Health Information in Help-Seeking Behavior Prior to a Diagnosis of Lung Cancer: A Mixed-Methods Study
J Med Internet Res 2017;19(6):e189
URL: http://www.jmir.org/2017/6/e189/
doi:10.2196/jmir.6336
PMID:28596146

Original Paper

Dr Google Is Here to Stay but Health Care Professionals Are Still Valued: An Analysis of Health Care Consumers' Internet Navigation Support Preferences

Kenneth Lee[1], BPharm (Hons), PhD; Kreshnik Hoti[2,3], BPharm, PhD; Jeffery David Hughes[2], BPharm, GradDipPharm, MPharm, PhD; Lynne Emmerton[2], BPharm (Hons), PhD

[1]Division of Pharmacy, School of Medicine, University of Tasmania, Hobart, Australia
[2]School of Pharmacy, Faculty of Health Sciences, Curtin University, Perth, Australia
[3]Division of Pharmacy, Faculty of Medicine, University of Prishtina, Prishtina, Albania

Corresponding Author:
Kenneth Lee, BPharm (Hons), PhD
Division of Pharmacy
School of Medicine
University of Tasmania
Private Bag 26
Hobart, 7001
Australia
Phone: 61 362262191
Fax: 61 362262870
Email: kenneth.lee@utas.edu.au

Abstract

Background: The Internet offers great opportunities for consumers to be informed about their health. However, concerns have been raised regarding its impact on the traditional health consumer-health professional relationship. Our recent survey of 400 Australian adults identified that over half of consumers required some form of navigational support in locating appropriate Web-based health information. We propose that support provided by health professionals would be preferred by consumers; this preference is regardless of whether consumers have a need for navigational support. Secondary analysis of the survey dataset is presented here to quantify consumer-reported support preferences and barriers when navigating Web-based health information.

Objective: We aimed to quantitatively identify consumers' support preferences for locating Web-based health information and their barriers when navigating Web-based health information. We also aimed to compare such preferences and barriers between consumers identified as needing and not needing support when locating Web-based health information.

Methods: Chi-square (χ^2) tests identified whether each listed support preference differed between subgroups of consumers classified as needing (n=205, 51.3%) or not needing (n=195, 48.8%) navigational support; degree of association, via phi coefficient (ϕ) tests, were also considered to ascertain the likely practical significance of any differences. This was repeated for each listed barrier. Free-text responses regarding additional support preferences were descriptively analyzed and compared with the quantitative findings to provide a richer understanding of desired support for health information searches.

Results: Of the 400 respondents, the most preferred mode of navigational support was involvement of health professionals; this was reported by participants identified as needing and not needing navigational support. While there was a significant difference between groups, the degree of association was small (χ^2_1 [N=400]=13.2; $P<.001$; $\phi=.18$). Qualitative data from the free-text responses supported consumers' desire for health professional involvement. The two most commonly reported barriers when navigating desired Web-based health information were (1) volume of available information and (2) inconsistency of information between sources; these were reported by participants with and without a need for navigational support. While participants identified with a need for navigational support were more likely to report volume (χ^2_1 [N=387]= 4.40; $P=.04$; $\phi=.11$) and inconsistency of information (χ^2_1 [N=387]= 16.10, $P<.001$, $\phi=.20$) as barriers, the degrees of association were small to moderate.

Conclusions: Despite concerns in the literature that the popularity of the Internet could compromise the health consumer-health professional relationship, our findings suggest the contrary. Our findings showed that health professionals were found to be the most commonly preferred mode of navigational support, even among consumers classified as not needing navigational support. Further research into how health professionals could assist consumers with Web-based health information seeking could strengthen the health consumer-health professional relationship amidst the growing use of “Dr Google.”

(J Med Internet Res 2017;19(6):e210) doi:10.2196/jmir.7489

KEYWORDS

health care; information seeking behavior; Internet; chronic disease; patients; surveys

Introduction

The relationships between health care consumers and health care professionals have changed over recent decades [1-4]. From the traditional one-way information transfer approach [4], education and communication between health care consumers and health care professionals is now a two-way exchange [2-4]. Similarly, health care consumer autonomy and sovereignty are arguably on the rise [5,6]. In this era, the role of health information in facilitating consumers’ health care roles appears to be of paramount importance [5], particularly given the increasing prevalence of chronic health conditions [7] that require day-to-day self-management [3].

The Internet provides consumers with an avenue to source, query, and publicly debate health information, as well as to self-diagnose and self-manage medical conditions. Concerns have been raised among health care professionals about the potentially negative impact of Web-based health information on the relationships between health care consumers and health care professionals [6,8]. Specifically, some health care professionals are concerned that consumers may make decisions about their health based on misleading or poor-quality information, overriding health care professionals’ recommendations [8,9]. By contrast, a recent study by Laugesen et al found that Web-based health information does not significantly influence treatment concordance, provided there is a good relationship between health care consumers and health care professionals [10]. While a larger body of evidence suggests that health care professionals are, overall, positive about the influence of the Internet on their relationships with health care consumers [8,9,11-14], little research explores the position of health care professionals in this evolving relationship.

Apart from concerns about the influence of the Internet on health care, the issues of variable quality and comprehensibility of available Web-based health information need to be addressed [8,15]; despite these issues, the Internet remains a popular channel for health information [16,17]. Collectively, these factors mean that consumers need to be able to find and appropriately use health information for their health management [18].

Appropriate searching and utilization of Web-based health information requires adequate health literacy [19], as well as digital or computer literacy and media literacy, collectively known as eHealth literacy [20]. While it has been established that the prevalence of limited health literacy [18,21-23] leads to poorer health outcomes and higher rates of hospital admissions and mortality [24,25], far fewer studies have investigated the prevalence of and health outcomes associated with eHealth literacy levels. For example, only one study appears to have specifically explored the prevalence of limited eHealth literacy [26], and one recent study explored associations between eHealth literacy and perceived health status and self-management skills [27]. Furthermore, little has been attempted in terms of interventions that can assist consumers in acquiring quality Web-based health information [28].

In order to inform tailored interventions to assist consumers with acquiring quality and relevant Web-based health information, our earlier research qualitatively explored consumers’ Web-based health information-seeking behaviors and the range of preferences for support with navigating Web-based health information [29]. From the manifest-level themes [30] identified in our qualitative study [29], we developed a questionnaire to survey a representative sample of 400 Australian adults living with chronic health conditions. The survey estimated that approximately half (51.3%, with a precision of 4.9% either side of the true population value) of Australian adult consumers using Web-based health information and living with chronic health conditions require support with navigating Web-based health information [31]; this survey also found that a need for navigational support was more likely among consumers with lower reported levels of eHealth literacy [31].

This study builds on our qualitative study [29] and previous survey [31] by (1) quantitatively identifying key consumer-reported navigational support preferences, (2) quantitatively identifying consumer-reported barriers to navigating desired Web-based health information, and (3) comparing findings from Objectives 1 and 2 between respondents classified as needing or not needing navigational support.

This study also aims to qualitatively explore additional navigational support preferences to provide a richer understanding of desired support for health information searches.

Methods

Overview

This study is a secondary analysis of the raw dataset from our previous survey [31], which used a Web-based questionnaire to identify health care consumers’ Web-based health information-seeking behaviors, eHealth literacy, and motivation to engage in their health care (activation). Examples of the types of questions asked in the aforementioned survey pertained to types of health information sought (eg, information about treatment options, medical conditions, diets, and exercises), why health information is sought on the Internet, what actions are taken once Web-based health information has been obtained, all 8 items from the eHealth Literacy scale developed by Norman and Skinner [20], all 13 items from the Patient Activation Measure developed by Hibbard et al [32], usual reason(s) for encountering difficulty with finding pertinent health information on the Internet, and preferred modes of support to assist with obtaining pertinent Web-based health information. The specific provider of the Web-based questionnaire platform used for the aforementioned survey research is Qualtrics, a company that also offers recruitment services via a third-party global research company, Research Now. While our previous analysis detailed an algorithm to identify the proportion of consumers with and without navigational needs when searching for Web-based health information, the secondary analysis reported here provides the first-known insight into navigational support preferences and barriers when finding the desired information and compares these findings between respondents in need of and those not in need of navigational support.

Ethical approval for the research, including the current analyses, was granted by the Curtin University Human Research Ethics Committee (HR06/2013).

Participants, Recruitment, and Sample Size

The target population of our previous survey study [31] was adults in Australia with a chronic health condition, who had sought health information for their condition on the Internet. Recruitment was contracted by Qualtrics to Research Now to fulfil the quota of 400 submitted questionnaires with a representative sample of Australian adults with the following criteria: ability to easily read and write in English, aged 18 years or older, diagnosis of at least one chronic health condition, and experience with the Internet to find information about their health conditions. These participants were chosen by Research Now from a combination of global Internet surveys or opinion panels, based on the participants' profiles and whether they would likely meet our stated eligibility criteria [33]. Screening questions that reflect our eligibility criteria were included in the questionnaire as an additional step to ensure that participants were indeed eligible to participate. The sample size of 400 participants was calculated using conservative parameters for prevalence studies such that there was 95% confidence that this study's reported values would fall within 4.9% on either side of the true population values [34].

Navigational Needs and Navigational Support Preferences

As defined in our previous survey, "navigational needs" refers to "individuals who report having difficulty finding, and would like support in locating, desired Web-based health information" [31]. This was operationally defined as participants who indicated (1) at least some level of difficulty in finding Web-based health information and (2) a desire for help to find Web-based health information [31]. Building on our definition of navigational needs, this study defines navigational support preferences as the mode by which consumers would prefer support in locating desired Web-based health information.

Analysis

Statistical analyses were conducted using IBM SPSS version 23. Descriptive statistics were used to identify preferred navigational modes of support (Objective 1) and the most commonly-reported barriers when navigating desired Web-based health information (Objective 2). Chi-square tests of independence (χ^2) were utilized to identify whether each support preference and barrier differed among the groups of consumers identified with and without a need for navigational support, with phi (ϕ)-coefficient tests of degrees of association to indicate the potential practical significance of these differences (Objective 3) [35]. Free-text responses regarding additional support preferences were also descriptively analyzed and compared with quantitative results to provide a richer understanding of support preferences.

Results

Overview

As described in our previous paper [31], 400 participants completed the questionnaire, and 51.3% (205/400) of these participants were identified with navigational needs. Specifically, "a total of 1104 individuals were invited by Research Now from their diverse participant pool. Consent was obtained from 93.03% (1027/1104) of these participants. Of these 1027 individuals, 50.05% (514/1027) met our eligibility criteria, and 77.82% (400/1027) completed the questionnaire" [31].

Navigational Support Preferences

In terms of determining participants' navigational support preferences, participants were asked, "Which of the following ideas would help you find health-related information that you need?" The most commonly reported mode of support, by both participants with and without navigational needs, was health care professionals (Table 1). While all reported modes of support (apart from the "other" option) were significantly more likely to be reported by participants with navigational needs, the ϕ coefficients suggest only a small degree of association.

Table 1. Support preferences for navigating Web-based health information (N=400).

Modes of support	No navigational needs, n (%[a,b])	Navigational needs, n (%[a,c])	Total, n (%[a])
Health care professionals to recommend websites (χ^2_1 [N=400]=13.2; *P*<.001; ϕ=.18)	112 (57.4)	153 (74.6)	265 (66.3)
An icon on each website to indicate whether it is from a trustworthy source (χ^2_1 [N=400]=8.8; *P*=.003; ϕ=.15)	96 (49.2)	131 (63.9)	227 (56.8)
Blocking of unreliable websites (χ^2_1 [N=400]=6.8; *P*=.01; ϕ=.13)	82 (42.1)	113 (55.1)	195 (48.8)
Improvement in the layout of websites (χ^2_1 [N=400]=12.8; *P*<.001; ϕ=.18)	35 (17.9)	69 (33.7)	104 (26.0)
Workshops on how to find trustworthy health-related information on the Internet (χ^2_1 [N=400]=3.9; *P*=.049; ϕ=.10)	17 (8.7)	31 (15.1)	48 (12.0)
Other	9 (4.6)	4 (2.0)	13 (3.3)

[a] Respondents could select multiple options (percentages do not total 100%).

[b] Percentages are based on the 195 participants with no navigational needs.

[c] Percentages are based the 205 participants with navigational needs.

Elaborating on the “other” option, the majority of additional modes of support suggested by participants reiterate and elaborate on one or more of the response options listed in Table 1. The below comment by one of the participants illustrates the desire for health care professionals to play a role in supporting Web-based health information seeking:

> *...maybe a handout of trusted web site addresses (specific to the condition discussed) could be given to the patient at the time of the visit to the doctor? This would make it easy for the patient to find out more specific information regarding their condition and know that the information they are reading is trustworthy and relevant.* [ID132]

Participants also suggested that a centralized health portal or database containing relevant and quality health information could be beneficial:

> *A large, centrally managed, government endorsed database of conditions, treatments, side effects and associated forums, institutes/foundations might be helpful. Unbiased, Australian, featuring advice from eminent specialists, and importantly, featured stories showing perspectives from sufferers on how they manage their conditions. I still find myself having to get my information from American sites quite often.* [ID228]

A few participants expressed a desire for medically-related search engines. One participant (ID197) said, “Google should have a doctor’s section (where you can read everything that is medically related).” Another participant (ID198) posed the question, “Search engine purely for medical research?”

Other participants expressed a desire for greater online communicative features. As one participant (ID171) expressed it, “Online chat services, just like hotline.” Another participant (ID236) felt that a “Website comprised of common information and online chat function” would be helpful.

Although few participants quantitatively indicated that they would welcome workshops to help them find health information on the Internet (Table 1), some indicated a desire to be better educated on how to use Web-based health information effectively. A notable example was this:

> *With the ability for anyone to put any information on the internet i [sic] think there is a lot of ill-informed information on the internet. It really depends on what your beliefs are about health as to what information is useful. I think there should be more information on what people should be asking their doctor or other health professionals in order to get the best information for them. Perhaps more sites on checklists on things to consider with your health and sources of information, what to ask, alternative options for health information.* [ID296]

Participant ID296’s response also suggests the need to be educated on how to have more productive interactions with health care professionals.

Perceived Barriers to Navigating Web-Based Health Information

Of the 400 participants, only 13/400 (3.3%) indicated that they never experienced difficulty finding relevant health information on the Internet. Table 2 summarizes the potential barriers to navigating Web-based health information, as reported by the remaining 387 participants. These 387 participants were asked, “When you do have difficulty finding helpful health-related information online, what is/are the usual reason(s)?” Overall, perceived barriers were similarly reported by both participants with and without navigational needs, except for one finding: the navigational needs group was moderately associated with a barrier relating to conflicting information obtained from different websites. Additionally, the two most commonly reported barriers by both groups were the volume of available information and conflicting information obtained from different websites.

Table 2. Perceived barriers to navigating Web-based health information (N=387).

Perceived barriers	No navigational needs, n (%[a,b])	Navigational needs, n (%[a,c])	Total, n (%[a])
There is a large number of websites available on the Internet (χ^2_1 [N=387]=4.4; P=.04; ϕ=.11)	80 (44.0)	112 (54.6)	192 (49.6)
Information obtained from different websites sometimes does not match up (χ^2_1[N=387]=16.1; P<.001; ϕ=.20)	65 (35.7)	115 (56.1)	180 (46.5)
The information I come across contains a lot of medical terms or jargon.	55 (30.2)	77 (37.6)	132 (34.1)
I do not know what I am looking for.	24 (13.2)	37 (18.0)	61 (15.8)
There appears to be a lack of information specific to my needs.	18 (9.9)	25 (12.2)	43 (11.1)
Other	27 (14.8)	11 (5.4)	38 (9.8)

[a] Respondents could select multiple options (percentages do not total 100%).

[b] Percentages are based on the 182 participants with no navigational needs.

[c] Percentages are based on the 205 participants with navigational needs.

Discussion

Principal Findings

Health care professionals were reported to be the most commonly selected option to assist health care consumers in finding credible information on the Internet. Specifically, participants indicated a desire for health care professionals to recommend suitable websites. Amidst some health care professionals' concerns that consumers' use of the Internet for health information could negatively impact the relationships between health care consumers and health care professionals [8,9], this study suggests that these concerns may be unfounded, at least among people with a chronic disease. The findings of this research support a study by Laugesen et al on the impact of Web-based health information on patients' treatment compliance [10].

A desire for health care professionals' support was significantly more likely to be reported by participants with navigational needs. Given that our previous study [31] found that participants with lower reported eHealth literacy levels were more likely to have navigational needs, this study suggests that health care professionals can have a role to play in supporting health information navigation on the Internet, particularly among consumers with lower levels of eHealth literacy. Nevertheless, the small degree of association between participants with and without navigational needs and their desire for health care professionals' support suggests that most lay consumers would be amenable to support from their health care professionals when navigating the Internet. Thus, this study provides reassurance that health care professionals are valued in an age where the Internet is a popular channel for obtaining health information [17]. However, we propose that the relationships between health care consumers and health care professionals would benefit from health care professionals acknowledging to their patients that Web-based information can be found and used effectively and that health care professionals can provide guidance as required. This time investment may indeed strengthen their relationship with health care consumers and assist in the eHealth literacy of their patients. However, further investigation would be required to determine the specific roles health care professionals could play in guiding lay consumers to become better informed about their health.

This study identified a number of perceived barriers when navigating Web-based health information. Specifically, the volume of available information on the Internet and the inconsistency of information obtained from different sources were identified as the two most commonly reported barriers across consumers with and without navigational needs. However, these two barriers were significantly more likely to be reported by participants with navigational needs, suggesting the need for greater efforts to be focused on addressing these barriers among consumers with navigational needs. In particular, a moderate association was found between the lack of consistency in information obtained from different sources among participants with navigational needs; this finding appears logical given the plethora of Web-based health information that can result in confusion and may not be directly comparable or may contain conflicting information or advice [15], hence the need for navigational support for this group. Consumers without navigational needs may be more competent and confident in comparing Web-based information to guide their decisions. Thankfully, as consumers with navigational needs are more likely to demonstrate a desire for navigational support from health care professionals, this study further highlights the important roles health care professionals can play in guiding consumers to Web-based health information.

When compared with previous studies, the overabundance of Web-based health information has indeed been reported as a potential issue with using the Internet as a channel for obtaining health information, as indicated in a review by Cline and Haynes [15]. Some studies suggest that information overload and ambiguity or inconsistency of information (not specific to health) can lead to patient confusion [5,36]. One study applied this concept to health care and found that information overload can influence a patient's cognition, and ambiguity or inconsistency of health information has both cognitive and affective effects [5]. While this study did not specifically assess information overload, the prevalence of this issue as a barrier suggests the need for initiatives to streamline search results and tailor suggested websites to individuals' needs and health literacy; attempts to partially address this need appear to have been considered by Google, as evidenced in the implementation of health information in their Knowledge Graphs [37]. In a Knowledge Graph, health information is presented as a summary with information drawn from credible sources, including the Mayo Clinic. Another approach to addressing barriers to navigating Web-based health information is the re-design of website user interfaces to align with consumers' information-seeking behaviors [38,39]. Overall, findings from this study highlight barriers to navigation commonly reported by consumers, thereby providing a focus for the development of future and current interventions to assist consumers in navigating Web-based health information.

The need to filter information was also recognized by participants who suggested a centralized portal for health information, provision of handouts with a list of trusted websites, and a "Google-esque" search engine that only displays medically-related health information from reputable sources. When compared with the responses listed in Table 1, one of the common themes is a desire for a reduction in information sources. For example, "An icon on each website to indicate whether it is from a trustworthy source" and "Blocking of unreliable websites" were among the most commonly rated modes of support (Table 1); both modes of support, if implemented, would arguably result in more relevant search results and more reliable content. Similarly, one could argue that by having health care professionals recommend websites, information sources searched by consumers could be reduced in volume. Collectively, this study suggests that a key perceived barrier to navigating Web-based health information is the volume of available information and that the preferred modes of support echo the desire to effectively distil available information into a smaller set of pertinent and quality sources.

Strengths and Limitations

A key strength of this study is the "within-method" triangulation [40] that was used to compare quantitative and free text (qualitative) findings from this study. The complementary nature of the collective findings serves to broaden understanding of the phenomenon and is a tactic to improve internal consistency or reliability of findings. Additionally, cross-validation, also known as "between-method" triangulation [40], was used to compare findings from similar studies with findings from this study.

Our structured, third party–administered questionnaire limited clarification and confirmation of free-text responses with participants. However, the use of a number of free-text responses and within-method data triangulation enabled a richer understanding of the phenomena beyond the understanding that would have been obtained if only quantifiable response options had been used [40].

As this study is a secondary analysis of a raw dataset, the methodological limitations recognized in our earlier paper [31] also apply here. For example, our reference to "navigation" only referred to finding (searching for or sourcing) Web-based health information. The findings of this study may not extend to consumers' preferences for assistance in interpreting and applying health information. Considering our key finding about the preference for assistance from a health care professional in finding relevant health information, it may be assumed that this type of "health coach" role would extend to assisting the health care consumer to manage the results of Internet searches, and specifically, understanding the relevance of information and how it may be applied at an individual level. Additionally, characteristics of the sample of participants, while reported by the third-party provider (Research Now) to be representative of our target population, cannot be confirmed. As such, the transferability of our findings to the wider population cannot be assured. However, given the moderately large sample size and the congruency between the quantitative and free-text responses of our participants, findings from this study may be applicable at least to the broader population of health information consumers living with chronic health conditions.

Further Research

The findings of this study suggest that future studies exploring consumers' health information-seeking behaviors should explore the concept of information overload and its effects on individuals having different levels of eHealth literacy. The concept of information overload, while investigated at length in other disciplines such as management [41], does not appear to be widely studied in the context of Web-based health information.

The likely acceptability of assistance from health care professionals is also worthy of further investigation. It is unclear whether health care professionals feel sufficiently skilled and confident to inform consumers about suitable information sources on the Internet for specific health conditions. Further research could explore health care professionals' ability to assess consumers' levels of eHealth literacy and subsequently direct them to suitable, reliable Internet sources. Given the increasing popularity of mobile-based health care and the advent of mobile apps [42], further research could also examine whether there is a need for consumers to be directed to appropriate apps and whether health care professionals are able to provide appropriate recommendations on suitable consumer health apps. To our knowledge, there has been no research into how health care professionals might integrate this type of assessment and advice into regular health consultations.

Regardless of how health care professionals might assist consumers in navigating Web-based health information, it is important to recognize that consumers' preferences for seeking or obtaining health information vary, and this difference can influence health behaviors. For example, a study by Williams-Piehota et al found that female participants with a lower desire for seeking health information were less likely to participate in a mammography after being given detailed information about mammograms as compared with when they were given concise information [43]. Conversely, female participants with a higher desire for seeking health information were less likely to participate in a mammography after being given concise information as compared with detailed information [43]. Thus, future interventions need to be cognizant of information-seeking needs at the individual level, to support desirable health behaviors.

Conclusions

This study explored consumer-perceived barriers to navigating desired Web-based health information and consumers' navigational support preferences. Our findings identified that the volume of available information and the inconsistency of information obtained from different information sources were the most commonly identified barriers. Despite concerns that the Internet and consumer sovereignty could negatively impact the relationships between health care consumers and health care professionals, our findings suggest the contrary, as health care professionals were reported as the most commonly selected option for providing navigational support. Further exploration of how health care professionals could assist consumers with their Web-based health information-seeking could see the strengthening of this relationship amidst the growing use of the Internet for obtaining health information.

Acknowledgments

The authors acknowledge the staff members of Qualtrics for coordinating the recruitment for the survey. KL was supported by an Australian Postgraduate Award during the design and data collection phases of the study.

Authors' Contributions

KL conceived and designed the study with assistance from LE, KH, and JH. KL performed the statistical analyses, interpreted the findings, and drafted the manuscript. All the authors were involved in reviewing and revising the manuscript.

Conflicts of Interest

None declared.

References

1. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
2. Ishikawa H, Hashimoto H, Kiuchi T. The evolving concept of “patient-centeredness” in patient-physician communication research. Soc Sci Med 2013 Nov;96:147-153. [doi: 10.1016/j.socscimed.2013.07.026] [Medline: 24034962]
3. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [Medline: 12435261]
4. Hoving C, Visser A, Mullen PD, van den Borne B. A history of patient education by health professionals in Europe and North America: from authority to shared decision making education. Patient Educ Couns 2010 Mar;78(3):275-281. [doi: 10.1016/j.pec.2010.01.015] [Medline: 20189746]
5. Gebele C, Tscheulin DK, Lindenmeier J, Drevs F, Seemann A. Applying the concept of consumer confusion to healthcare: development and validation of a patient confusion model. Health Serv Manage Res 2014;27(1-2):10-21. [doi: 10.1177/0951484814546959] [Medline: 25595013]
6. Hardey M. Consumers, the internet and the reconfiguration of expertise. In: Harlow E, Webb SA, editors. Information and Communication Technology in the Welfare Services. London: Jessica Kingsley Publishers; 2003.
7. Nugent R. Chronic diseases in developing countries: health and economic burdens. Ann N Y Acad Sci 2008;1136:70-79. [doi: 10.1196/annals.1425.027] [Medline: 18579877]
8. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the internet on health care and the physician-patient relationship: National U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
9. de Oliveira JF. The effect of the internet on the patient-doctor relationship in a hospital in the city of Sao Paulo. J Inf Syst Technol Manag 2014 Aug 21;11(2):327-344. [doi: 10.4301/S1807-17752014000200006]
10. Laugesen J, Hassanein K, Yuan Y. The impact of internet health information on patient compliance: a research model and an empirical study. J Med Internet Res 2015 Jun 11;17(6):e143 [FREE Full text] [doi: 10.2196/jmir.4333] [Medline: 26068214]
11. Iverson SA, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]
12. Hart A, Henwood F, Wyatt S. The role of the Internet in patient-practitioner relationships: findings from a qualitative research study. J Med Internet Res 2004 Sep 30;6(3):e36 [FREE Full text] [doi: 10.2196/jmir.6.3.e36] [Medline: 15471762]
13. Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
14. Stevenson F, Kerr C, Murray E, Nazareth I. Information from the Internet and the doctor-patient relationship: the patient perspective--a qualitative study. BMC Fam Pract 2007 Aug 16;8:47 [FREE Full text] [doi: 10.1186/1471-2296-8-47] [Medline: 17705836]
15. 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]

16. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of e-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
17. Fox S, Duggan M. Pew Research Center. 2013. Part two: sources of health information URL: http://www.pewinternet.org/2013/11/26/part-two-sources-of-health-information/ [accessed 2017-02-10] [WebCite Cache ID 6oABK7tT4]
18. Baker D, Wolf M, Feinglass J, Thompson J, Gazmararian J, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med 2007 Jul 23;167(14):1503-1509. [doi: 10.1001/archinte.167.14.1503] [Medline: 17646604]
19. Nutbeam D. Defining and measuring health literacy: what can we learn from literacy studies? Int J Public Health 2009;54(5):303-305. [doi: 10.1007/s00038-009-0050-x] [Medline: 19641847]
20. Norman C, Skinner H. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
21. Manning DM, Dickens C. Health literacy: more choice, but do cancer patients have the skills to decide? Eur J Cancer Care (Engl) 2006 Dec;15(5):448-452. [doi: 10.1111/j.1365-2354.2006.00687.x] [Medline: 17177901]
22. Australian Bureau of Statistics. ABS. Canberra (Australia); 2006. 4228.0 - Adult literacy and life skills survey, summary results, Australia, 2006 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4228.0Main+Features12006%20(Reissue) [accessed 2017-02-10] [WebCite Cache ID 6oABOyup2]
23. Kutner M, Greenberg E, Jin Y, Paulsen C. NCES. Washington, DC; 2006. The health literacy of America's adults: results from the 2003 national assessment of adult literacy URL: https://nces.ed.gov/pubs2006/2006483.pdf [accessed 2017-02-10] [WebCite Cache ID 6oABXgGIz]
24. Berkman N, Sheridan S, Donahue K, Halpern D, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
25. Bostock S, Steptoe A. Association between low functional health literacy and mortality in older adults: longitudinal cohort study. BMJ 2012;344:e1602 [FREE Full text] [Medline: 22422872]
26. Xesfingi S, Vozikis A. eHealth literacy: in the quest of the contributing factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
27. Neter E, Brainin E. Perceived and performed ehealth literacy: survey and simulated performance test. JMIR Hum Factors 2017 Jan 17;4(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.6523] [Medline: 28096068]
28. Lee K, Hoti K, Hughes JD, Emmerton LM. Interventions to assist health consumers to find reliable online health information: a comprehensive review. PLoS One 2014;9(4):e94186 [FREE Full text] [doi: 10.1371/journal.pone.0094186] [Medline: 24710348]
29. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
30. Boyatzis R. Transforming qualitative information: thematic analysis and code development. Thousand Oaks, CA: Sage Publications; 1998.
31. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of “Dr Google”: a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
32. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
33. Research Now. Research Now. London: Research Now Group; 2017. Research now: panel book URL: http://www.researchnow.com/wp-content/uploads/2017/02/Panel-book_UK_Jan17.pdf [accessed 2017-03-27] [WebCite Cache ID 6pGTk0rNS]
34. National Statistical Service. NSS.: Australian Bureau of Statistics Sample size calculator URL: http://www.nss.gov.au/nss/home.nsf/pages/Sample+size+calculator [accessed 2017-02-10] [WebCite Cache ID 6oABb5afI]
35. Field A. Discovering statistics using SPSS. Thousand Oaks, CA: Sage Publications; 2013.
36. Walsh G, Hennig-Thurau T, Mitchell V. Consumer confusion proneness: scale development, validation, and application. J Mark Manag 2007 Sep 17;23(7-8):697-721. [doi: 10.1362/026725707X230009]
37. Ramaswami P. Googleblog. 2015. A remedy for your health-related questions: health info in the knowledge graph URL: https://googleblog.blogspot.com.au/2015/02/health-info-knowledge-graph.html [accessed 2017-02-10] [WebCite Cache ID 6oABeyqmt]
38. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
39. Cui L, Carter R, Zhang G. Evaluation of a novel conjunctive exploratory navigation interface for consumer health information: a crowdsourced comparative study. J Med Internet Res 2014;16(2):e45 [FREE Full text] [doi: 10.2196/jmir.3111] [Medline: 24513593]
40. Jick TD. Mixing qualitative and quantitative methods: triangulation in action. Admin Sci Q 1979 Dec;24(4):602-611. [doi: 10.2307/2392366]

41. Eppler MJ, Mengis J. The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. Inf Soc 2004 Nov;20(5):325-344. [doi: 10.1080/01972240490507974]
42. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
43. Williams-Piehota P, Pizarro J, Schneider TR, Mowad L, Salovey P. Matching health messages to monitor-blunter coping styles to motivate screening mammography. Health Psychol 2005 Jan;24(1):58-67. [doi: 10.1037/0278-6133.24.1.58] [Medline: 15631563]

Edited by G Eysenbach; submitted 21.02.17; peer-reviewed by P Pang, R Ownby; comments to author 16.03.17; revised version received 28.03.17; accepted 10.04.17; published 14.06.17.

Please cite as:
Lee K, Hoti K, Hughes JD, Emmerton L
Dr Google Is Here to Stay but Health Care Professionals Are Still Valued: An Analysis of Health Care Consumers' Internet Navigation Support Preferences
J Med Internet Res 2017;19(6):e210
URL: http://www.jmir.org/2017/6/e210/
doi:10.2196/jmir.7489
PMID:28615156

Original Paper

Classifying Chinese Questions Related to Health Care Posted by Consumers Via the Internet

Haihong Guo[1], MSc; Xu Na[1], MSc; Li Hou[1], PhD; Jiao Li[1], PhD

[1] Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, China

Corresponding Author:
Jiao Li, PhD
Institute of Medical Information & Library, Chinese Academy of Medical Sciences
3rd Yabao Road Chaoyang District
Beijing, 100020
China
Phone: 86 01052328740
Fax: 86 01052328610
Email: li.jiao@imicams.ac.cn

Abstract

Background: In question answering (QA) system development, question classification is crucial for identifying information needs and improving the accuracy of returned answers. Although the questions are domain-specific, they are asked by non-professionals, making the question classification task more challenging.

Objective: This study aimed to classify health care–related questions posted by the general public (Chinese speakers) on the Internet.

Methods: A topic-based classification schema for health-related questions was built by manually annotating randomly selected questions. The Kappa statistic was used to measure the interrater reliability of multiple annotation results. Using the above corpus, we developed a machine-learning method to automatically classify these questions into one of the following six classes: *Condition Management*, *Healthy Lifestyle*, *Diagnosis*, *Health Provider Choice*, *Treatment*, and *Epidemiology*.

Results: The consumer health question schema was developed with a four-hierarchical-level of specificity, comprising 48 quaternary categories and 35 annotation rules. The 2000 sample questions were coded with 2000 major codes and 607 minor codes. Using natural language processing techniques, we expressed the Chinese questions as a set of lexical, grammatical, and semantic features. Furthermore, the effective features were selected to improve the question classification performance. From the 6-category classification results, we achieved an average precision of 91.41%, recall of 89.62%, and F_1 score of 90.24%.

Conclusions: In this study, we developed an automatic method to classify questions related to Chinese health care posted by the general public. It enables Artificial Intelligence (AI) agents to understand Internet users' information needs on health care.

(J Med Internet Res 2017;19(6):e220) doi:10.2196/jmir.7156

KEYWORDS

classification; natural language processing; hypertension; consumer health information

Introduction

The Internet is increasingly becoming a main resource for consumers to acquire health information. Until December 2015, there were 152 million Internet health users in China, indicating that 22.1% of Chinese Internet users have looked online for health information and services [1]. Many studies have proved that health-related information online could impact consumers' health-related attitudes and behaviors [2-4]. However, consumers have difficulty in expressing their information needs accurately using medical query terms, thus failing to retrieve relevant health information [5,6]. Automatic question answering (QA) systems are available for such users and they respond with concise and correct answers using natural language processing techniques. Thus the QA systems have become one of the most important research focuses in the field of biomedicine [7].

In general, a QA system consists of 3 modules: question analysis, information retrieval, and answer extraction. In the first module, question classification plays an important role in identifying the information needs of consumers, reducing the space of candidate answers, and further improving the accuracy of returned answers [8]. Classification schema is the basis of question classification. However, due to the difference in the information needs of health providers and consumers [9-11], the existing question classification schemas for professional health-related questions (the International Classification of Primary Care [12,13], the Taxonomies of Generic Clinical Questions (TGCQ) [14], etc) are not suitable for consumer health questions. Although some research focuses on the classification schema of consumer health questions [15,16], it has not been defined in a systematic manner yet. Therefore, it is a prerequisite to design a concise and valid classification schema.

Several studies have been conducted for automatic quesiton classification in the field of health and medicine in order to identify the general topics of clinical questions [17], distinguish between answerable and unanswerable intensive care unit (ICU) questions [18], separate consumer health questions from professional medical questions [19], and classify the types of consumer health questions [20]. Research has demonstrated that support vector machines (SVMs) performed the best among the most commonly explored algorithms, including naive bayesian, decision tree, maximum entropy, logistic regression, and conditional random fields. However, different patterns of thinking and habits of Chinese expression offen cause a mass of difference in the flexibility of word order and parse for Chinese health questions [21]. Several studies on Chinese NLP focused on clinical named entity recognition [22], diseases, or drag-related clinical information extraction [23,24] and speculation detection [25] from the free-text of pathology and operation notes. The main challenges in these tasks were word segmentation and feature representation and selection. To our knowledge, few studies have investigated consumer health question classification in Chinese.

As one of the most common chronic diseases, hypertension has become the main risk factor of cardiovascular diseases. It was estimated that China had 270 million patients with hypertension in 2012, and the incidence rate was approximately 3% per year [26]. Thus, hypertension-related questions are frequently asked with large variability on the Internet. For this reason, this study aimed at building a general topic classification schema and an automatic classification method for consumer health questions in Chinese, for the purpose of facilitating users' hypertension-related information needs analysis and answer extraction.

Methods

Data Collection

We collected questions posted by health consumers from 1st January to 10th August, 2014, with the tags “hypertension (高血压)” or “blood pressure (血压)” under the Q&A (有问必答) section on a Chinese health website with more than 35 million registered users [27] and imported the data into a MySQL (MySQL Community Edition, Oracle) database. The resulting database included 98,032 messages, from which 2000 messages were randomly selected as the sample for analysis.

In this study, “question” is defined as a request that a health consumer has posted on the website on a certain subject to elicit answers from physicians, which was identified based on meaning, not form. We focused on questions related to hypertension (高血压), which was sometimes expressed as “high blood pressure (高血压),” or simply as “high pressure (高压).” Therefore, we manually discarded messages that did not match the definition and that were irrelevant to hypertension but which contained similar words such as “high pressure oxygen (高压氧),” “hyperbaric cabin (高压舱),” “high voltage (高压电),” “pressure cooker (高压锅),” and so on. A new message was randomly selected from the database when an irrelevant message was discarded from the sample, so as to keep the sample size at 2000.

The website provides a template for users to generate questions, which includes three fields: (1) describe your health status (病情描述), (2) treatments or tests in the past (曾经的治疗或检查情况), and (3) what kinds of help do you want (想得到怎样的帮助). This template might lead to confusion in customers regarding how to post their questions. To deal with this case, we developed a rule: if the phrase “what kinds of help do you want (想得到怎样的帮助)” was found in the message, then we would take the sentence after the phrase as the “question.” Otherwise, we would take the whole message as the “question.” By doing this, we collected 2000 questions with an average length of 48 words.

Classification Schema and Corpus Construction

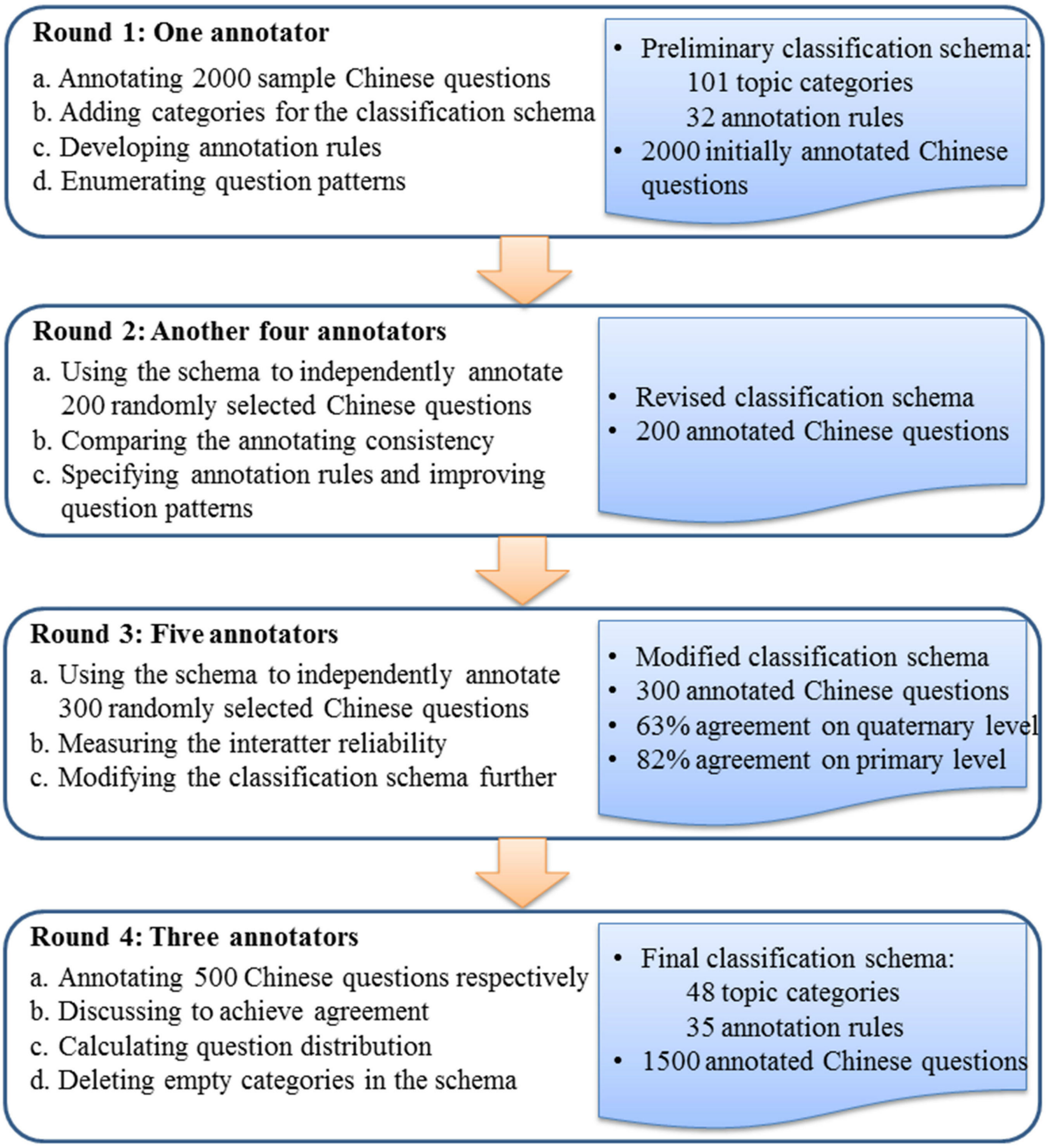
A topic-based classification schema was developed based on TGCQ [14] and the Layered Model of Context for Consumer Health Information Searching (LMCC) [15], and some categories were divided into more specific sub-categories to code the specific information needs. We produced the annotations in 4 rounds (Figure 1). In round 1, one annotator (specialized in medical informatics) annotated all the 2000 sample questions, following the classification schema. Some categories were added to accommodate questions that did not fall into any existing specialty during the process. For purposes of consistency improvement among coders, and the usability of the classification, a list of annotation rules was developed and some question patterns were enumerated for even the smallest category. As a result, the preliminary classification schema for consumer health questions included 101 topic categories and 32 annotation rules.

In round 2, four other annotators (two specialized in medicine and two specialized in informatics) independently annotated 200 questions randomly selected from the sample, using the classification schema. The authors compared the consistency of the five coding results (including the one in the first round) and categorized the 200 questions into three groups: (1) all annotators agreed (n=73), (2) only one disagreed (n=63), and (3) more than one disagreed (n=64). Then we focused on the last group. We addressed ambiguous elements by further specifying annotation rules and improving the descriptions of the question patterns.

In round 3, the revised classification was distributed to the five annotators who independently annotated another 300 questions randomly selected from the remaining sample of 1800 messages. This step was done to measure the interrater reliability of the classification schema as well as to further modify it.

In the last round, each of three annotators independently annotated 500 from the remaining 1500 messages. So each of the 2000 sample questions were annotated by at least two annotators. The authors compared the coding results and the disparities were discussed to achieve an agreement. The codes agreed upon during this step were regarded as the final schema. The number of questions in each category was calculated, and categories in which no questions were filled were deleted (such as physical characteristics of drugs, pharmacodynamics, and mechanism of drug action).

Figure 1. A four-round annotating process to construct and modify the classification schema and annotated corpus.



Automatic Classification of the First-Level Topics

The 2000 questions annotated by the above steps were used to train and test the classifiers for the primary level topics, including *Diagnosis*, *Treatment*, *Condition Management*, *Healthy Lifestyle*, and *Health Provider Choice*.

Feature Selection

We explored various features for machine-learning, including lexical, grammatical, semantic, and statistical information (Multimedia Appendix 1).

Lexical Features Such as Bag-of-Words and Part-of-Speech

The word segmentation was obtained from Rwordseg [28], a Chinese word segmentation tool under R environment based on the Chinese lexical analysis system, ICTCLAS, which reached a precision of 97.58% on the *973* experts testing [29]. The Chinese part-of-speech tag was obtained by using the Stanford Parser (version 3.3.1) [30].

Grammatical Features Such as Interrogative Words, Noun Head Chunks, Verb Head Chunks, Noun Rear Chunks, Verb Rear Chunks, Interrogative Words + Noun or Verb

Head Chunks, and Noun or Verb Rear Chunks + Interrogative Words

We manually developed a dictionary of 42 Chinese interrogative words based on baike.baidu [31] and general types of Chinese consumer questions summarized by our former research [32]. In this study, the noun or verb head chunk is the first noun or verb after the first interrogative word in a question, and the noun or verb rear chunk is the last noun or verb before the first interrogative word. They are likely to be the dependent words of the interrogative word that help to express the semantic information in the question [33].

CMeSH Concepts and Semantic Types

The controlled vocabulary of Chinese Medical Subject Headings (CMeSH) [34] was applied to recognize the medical concepts and their semantic types (*Disease*, *Drug*, and *Symptom*) in the Chinese consumer health questions.

Keywords

These were a combination of lexical and statistical features. We used three ways to extract the keywords from a question: (1) the first *k* words of maximum term frequency (TF), (2) the first *k* words of maximum inverse-document frequency (IDF), and (3) the first *k* words of maximum TF-IDF. We adopted the heuristic equation (1) developed by Cao et al [17] to calculate *k*, which was based on the observation that the number of keywords increases when the question length increases.

Statistical Features

These include question length, maximum, minimum, and average word length, maximum, minimum and average TF, maximum, minimum, and average IDF, and maximum, minimum and average TF-IDF. The corpus used to calculate the IDF of each word contained nearly 100 thousand hypertension-related messages that we had collected in our former research [32]. The TF, IDF, and TF-IDF were computed by equations (2), (3), and (4), shown in Figure 2.

Figure 2. Mathematical equations.

$$k_{q_i} = \begin{cases} \frac{n_{q_i}}{6} + 1 \\ 3, n_{q_i} > 12 \end{cases} \qquad (1)$$

where k_{q_i} is the number of keywords extracted from the i^{th} question, and n_{q_i} is the number of words in the i^{th} question.

$$\mathrm{TF}(w_{ij}) = \frac{n_{w_{ij}}}{n_{q_i}} \qquad (2)$$

$$\mathrm{IDF}(w_{ij}) = \log\left(\frac{n_q}{n_{qw_{ij}}}\right); \qquad (3)$$

$$\mathrm{TFIDF}(w_{ij}) = \mathrm{TF}(w_{ij}) \times \mathrm{IDF}(w_{ij}); \qquad (4)$$

where w_{ij} is the j^{th} word in the i^{th} question, $n_{w_{ij}}$ is the number of w_{ij} in the i^{th} question, n_q is the total number of questions in the corpus, and $n_{qw_{ij}}$ is the number of questions in which w_{ij} appears.

$$\Phi(t) = \frac{\left(\bar{x}_t^{(+)} - \bar{x}_t\right)^2 + \left(\bar{x}_t^{(-)} - \bar{x}_t\right)^2}{\frac{1}{n_+ - 1}\sum_{s=1}^{n_+}\left(x_{s,t}^{(+)} - \bar{x}_t^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{s=1}^{n_-}\left(x_{s,t}^{(-)} - \bar{x}_t^{(-)}\right)^2}; \qquad (5)$$

where $\bar{x}_t$, $\bar{x}_t^{(+)}$ and $\bar{x}_t^{(-)}$ are the average of the t^{th} feature of the whole, positive, and negative data sets respectively; $x_{s,t}^{(+)}$ is the t^{th} feature of the s^{th} positive instance, and $x_{s,t}^{(-)}$ is the t^{th} feature of the s^{th} negative instance. The numerator indicates the discrimination between the positive and negative sets, and the denominator indicates the one within each of the two sets.

As the feature space dimension was very large, and some of them could have degraded the performance of the classifiers, we adopted Φ-score to select the most discriminative features, which measures the discriminations in two sets of real numbers [35]. Given the training vectors $x_{s,}$, s=1,..., m, if the number of positive and negative instances are n+ and n-, respectively, the Φ-score of the t^{th} feature is defined in equation (5) in Figure 2. The larger the Φ(t) is, the more likely this feature is more discriminative. Therefore, we used Φ(t) as the feature selection criterion, and the implementation steps were as follows:

Calculate Φ(t) of every feature

Calculate the avg Φ of each type of feature and, further, set it as the threshold of the corresponding feature type. The avg Φ was chosen as the feature selection threshold because the distribution of Φ differs greatly between different types of features, while this method can help to keep all the useful features in different types [36].

For each type of machine-learning feature, select features with Φ ≥ avg Φ of this type.

Classifiers

Since a question can be assigned to multiple topics, the task in this paper was a multi-label classification problem, which was usually transformed into one or more single-label classification or regression problems [37]. We therefore transformed the task into six binary classification problems (one-versus-rest for each) so as to suit the SVMs [38], which were commonly used and claimed to be the best in related works [17-20]. We used machine-learning algorithms within the R project for statistical computing (version 3.3.1) for automatic question classification, including naive bayesian, SVMs, decision tree, maximum entropy, logistic regression, and conditional random fields. The results showed that SVMs performed the best among all the algorithms in 10-fold cross-validation.

Training and Testing

Due to the skewed distribution of consumer questions to different topics, an under-sampling method for the majority classes was applied to ensure that each classifier was trained and tested on the same number of “positive” and “negative” questions. We reported the classification performance using 10-fold cross-validation. The sample data for each binary classifier was equally divided into 10-folds: one of them was used as testing data, and the ramaining 9 folds as training data. The cross-validation process was repeated 10 times (equal to the folds) and the average value and standard deviation were reported. All cases in the sample data were used for both training and validation. Thus, each case was used for validation exactly once, which was the distinct advantage in this method [39].

Evaluation Metrics

The interrater reliability of the classification schema was evaluated by the kappa statistic, which could correct agreement that occurred by chance. Kappa=(P_o-P_e)/(1-P_e), where P_o is the observed agreement and P_e is the agreement expected by chance [40]. When the number of categories was large, as in this study, P_e would be close to zero, and the kappa value would be close to P_o [14]. Thus, we directly used P_o as the kappa value. The bigger the kappa value, the better the agreement. We assume that when the user asked more than one question, it was acceptable to answer any one of them. Therefore, a liberal reliability criterion was used: a match was recorded if either the main or minor topics assigned by one annotator matched the other’s assignments.

The performance of automatic classification methods was evaluated by precision (p), recall (r) and F_1 score, all of which are commonly used as evaluation metrics for text categorization, and we report the average of each metric. Precision is the number of correctly classified cases divided by the total number of cases classified for the category; recall is the number of correctly classified cases divided by the total number of cases of that class; and F_1 score is the harmonic mean of precision and recall, calculated as F_1 =(2×p×r) / (p + r).

Results

Classification Schema of Consumer Health Questions

The final classification schema was a four-hierarchical-level of specificity, consisting of 48 quaternary categories (see Multimedia Appendix 1) and 35 annotation rules. The first level included seven areas, namely diagnosis, treatment, condition management, epidemiology, healthy lifestyle, health provider choice, and other. A branching structure of secondary, tertiary, and quaternary levels describes more specific topics of the questions than its upper level. One or more closely related question patterns were listed for each quaternary category. Table 1 shows examples of consumer health questions in Chinese with their pattern and annotated tags on the topics of diagnosis and treatment, respectively.

Table1. An example of consumer health questions in Chinese with their pattern and annotated tags.

Table 1.

General Topics	Items	Contents
Diagnosis	Question	昨天不知道怎么事，突然感到心慌慌的，四肢发凉，全身冒冷汗，之后老婆扶我到小区医院那里去看，量了一下血压，血压比以往要高，之后医生叫我放松，休息了20分钟左右，又感觉没有什么事了。。 请问突然感觉到心慌，四肢发凉，血压升高，这是啥病啊? (Yesterday, my heart suddenly palpitated, my limbs became cold, and my whole body began to sweat. Then my wife accompanied me to the community hospital and checked my blood pressure; it was higher than before. The doctor told me to relax, and I feel much better after resting for about 20 minutes… suddenly felt flustered, limbs became cold, and blood pressure rose. What disease is it?)
	Pattern	临床发现X1、X2、X3、……，这是啥病？(Clinical finding X1, X2, X3,… What disease is it?)
	Tag	1.1.4.1 “诊断(Diagnosis)→病因/临床发现的解释(Interpretation of clinical finding)→不具体的发现或多种发现(Uncertain/multiple findings)”
Treatment	Question	65岁老人血压高经常不稳定，吃哪种降压药最好？(A 65-year-old man with unsteady high blood pressure… What’s the best blood pressure drug to take?)
	Pattern	病情y，吃/用/服用哪种药最好？(Condition y: What’s the best drug to take or use?)
	Tag	2.1.2.1 “治疗(Treatment)→药物治疗(Drug therapy)→效力/适应症/药物选择(efficacy/indications/drug choosing)→治疗(Treatment)”

General Topics of Questions Asked by Health Consumers

This study found that although health consumers would ask numerous health questions about themselves or their families, the general topics of the questions were limited to a small number and each category of the topics had its particular question patterns. The 2000 Chinese consumer health questions were annotated with 2000 major codes and 607 minor codes.

The distribution of the sample questions on the primary level category is shown in Table 2. 26.35% of the questions were annotated with more than one topic, which demonstrated that health consumers tend to ask more than one question at a time and the question messages usually belong to multiple topic categories [11] (Multimedia Appendix 1). These findings indicated that the various consumer health questions could be represented by limited topics and keywords, and the task to classify those topics was a multi-label problem [37].

Table 2. Distribution of the 2000 consumer health questions in Chinese on the primary level of topics.

No.	General Topics	Positive	Negative	Total
1	Diagnosis	600	1400	2000
2	Treatment	1167	833	2000
3	Condition management	136	1864	2000
4	Epidemiology	233	1767	2000
5	Healthy lifestyle	278	1722	2000
6	Health provider choice	45	1955	2000
7	Other	5		
	Total	2000	2000	2000

Interrater Reliability of the Classification Schema

The kappa statistic for the five annotators was 0.63 in the quaternary level of the classification, indicating “substantial” reliability, better than in several similar studies, such as assigning topics to general clinical questions (kappa=0.53) [14]. When only the primary and secondary levels were considered, the kappa value increased to 0.75. When only the seven broad areas in the primary level were considered, the kappa value was 0.82, slightly better than automatically classifying question types for consumer health questions in English conducted by Roberts et al [20].

Feature Selection for Automatic Question Classification

The Φ-score of each feature was calculated for each binary classifier. We found that their distribution between different types of features differed greatly. The performance of classifiers using features with $\Phi \geq$ avg Φ was not worse than that of those classifiers using all the features in the corresponding types, and some of them were even higher than the latter. Taking the topic of *Lifestyle* as an example, the average and standard deviation of Φ in each feature type are shown in the third and fourth columns in Table 3. The avg Φ of bag-of-words was 0.0016 with a standard deviation of 0.0067, while the values of keywords with maximum TF were 0.0008 and 0.0009, respectively. The average F_1 score of the classifier was 74.08% when using all the 6154 features in part-of-speech, while the performance increased to 78.84% when just taking the 1490 features with $\Phi \geq$ avg Φ (Figure 2). Similar cases can be seen in feature types of noun rear chunks, interrogative + noun or verb head chunks, verb rear chunks + interrogative, keywords with maximum TF, TF-IDF, and so on. The observations indicated that some of the features in each type either do nothing to the classifiers or have some side effects on them. Rejecting these features could not only save the computing resources so as to increase the efficiency, but also improve the performance of the classifiers.

Table 3. Number and Φ distribution of each type of feature for the Chinese consumer health question classification on the topic of *Lifestyle.*

Levels	Features Types[a]	Avg Φ	σ (Φ)	n_{AF}	$n_{(\Phi \geq \text{avg } \Phi)}$
Lexical	Bag-of-words	0.0016	0.0067	4967	1301
	Part-of-speech	0.0014	0.0060	6154	1490
Grammatical	Interrogative words	0.0039	0.0204	97	13
	Noun head chunks	0.0011	0.0010	48	14
	Verb head chunks	0.0008	0.0007	19	6
	Noun rear chunks	0.0011	0.0019	73	14
	Verb rear chunks	0.0010	0.0013	22	3
	Interrogative + noun head chunks	0.0011	0.0013	328	86
	Interrogative + verb head chunks	0.0011	0.0010	312	85
	Noun rear chunks + interrogative	0.0010	0.0013	315	67
	Verb rear chunks + interrogative	0.0012	0.0024	318	74
Semantic	CMeSH concepts	0.0016	0.0033	43	9
	CMeSH semantic types	0.0124	0.0101	3	1
Lexical & Statistical	Keywords (TF)	0.0008	0.0009	1510	282
	Keywords (IDF)	0.0007	0.0008	1137	192
	Keywords (TF-IDF)	0.0008	0.0008	1208	190
Statistical	Statistical features	0.0073	0.0060	13	5
	Total with duplicates replaced			15349	3656

[a] For each type of feature, σ (Φ) is the standard deviation of Φ, n_{AF} is the total number of features, n ($\Phi \geq$ avg Φ) is the number of features with $\Phi \geq$ avg Φ.

Therefore, the features with $\Phi \geq$ avg Φ in every feature type were selected as input features for machine-learning, in order to keep all the useful features in different types and to improve the performance of the classifiers. Thus, each classifier received a different feature set, and the number of features within them are showed in the third column in Table 4. For example, words such as "*drinking* (饮酒)," "*eat a meal* (吃饭)," "*breakfast* (早餐)," "*stay up late* (熬夜)," "*weight* (体重)," "*daily life* (平时)," "*nurse one's health* (调理)," and so on were the effective bag-of-words features for the classifier for *Healthy Lifestyle* but not effective for the classifier for *Diagnosis*. On the contrary, words such as "*diagnose* (诊断)," "*judge* (判断)," "*indicate* (提示)," "*physical examination* (查体)," "*cardiac murmur* (杂音)," "*head rush* (脑充血)," "*dazed* (昏沉沉)," and so on were the effective bag-of-words features for the classifier for *Diagnosis* but not effective for the classifier for *Healthy Lifestyle*.

Table 4. Feature reduction and the performance of each classifier.

General topics	N (all features)	N (selected features)	Feature reduction proportion	Avg F_1	σ (F_1)
Diagnosis	15349	5311	0.6540	0.9855	0.0164
Treatment	15349	4216	0.7253	0.7602	0.0482
Condition management	15349	3150	0.7948	0.9963	0.0117
Epidemiology	15349	4194	0.7268	0.7177	0.0798
Healthy lifestyle	15349	3656	0.7618	0.9913	0.0166
Health provider choice	15349	2282	0.8513	0.9635	0.0594

Figure 3. Performance of each feature type for Chinese consumer health question classification on the topic of Lifestyle.

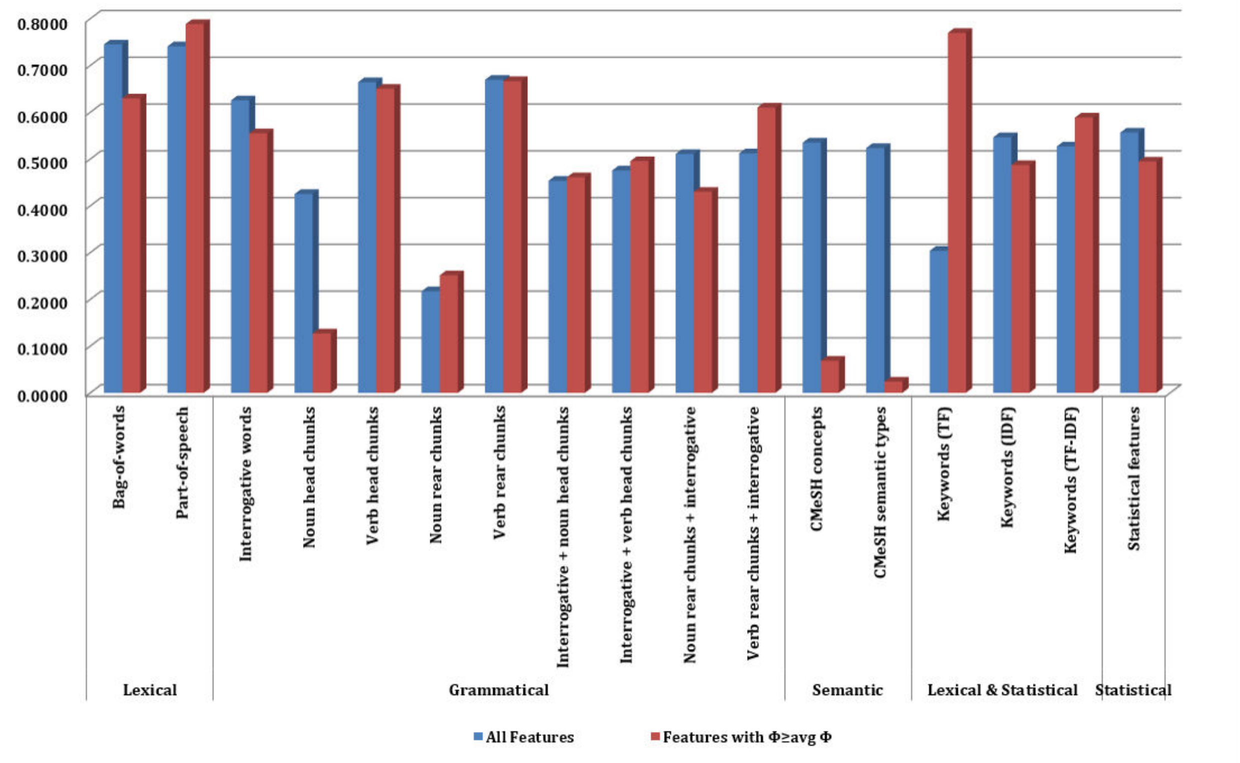
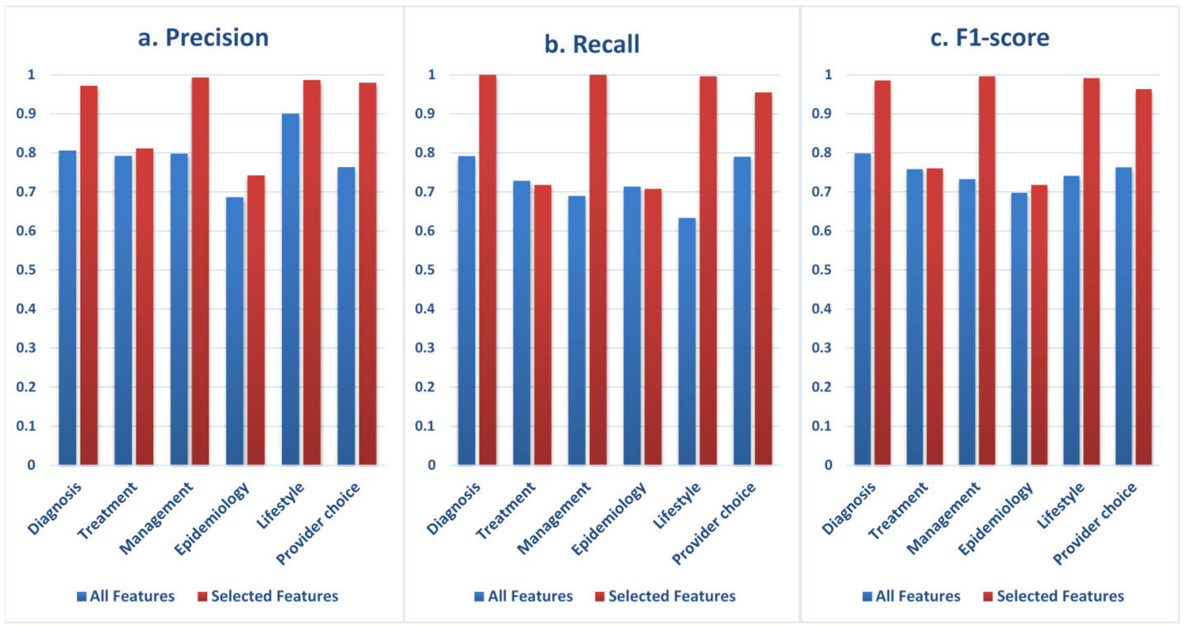


Figure 4. Performance improvement of each classifier by selecting features above the threshold.



Performance of the Automatic Classification Methods

The results were obtained from SVMs in the kernlab package because it performed the best among all the classification algorithms available in the R project. The research findings showed that the feature spaces were reduced from 65.40% to 85.13% by dropping features under the threshold (Table 4). Furthermore, the performance of each classifier also improved significantly (Figure 3). For example, the average F_1 score of the classifier for *Diagnosis* dramatically increased from 79.84% to 98.55%, while the feature space reduced from 15349 to 5311 (reduced 65.40%). The feature space of the classifier for Treatment reduced greatly (Down from 15349 to 4216, reduced 72.53%), although the improvement of performance was not so obvious (F1 score increased from 75.82% to 76.02%). The results of 10-fold cross validation on each binary classifier affirmed the hypothesis we proposed in the Feature Selection sections.

The performance on the classification of most topics of consumer health questions in Chinese was high. The evaluation metrics (average precision, recall, and F_1 score) of *Diagnosis, Condition Management, Healthy Lifestyle, and Health Provider Choice* were above 95%, while, the metrics of *Treatment* and *Epidemiology* were 76.02% and 71.77%, which were relatively weaker than the others. The standard deviation indicated that the performance of each binary classifier was relatively robust. As of 6-category classification results, we achieved an average precision of 91.41%, recall of 89.62%, and F_1 score of 90.24%.

Discussion

Principal Findings

A classification schema of consumer health questions was built in this study and 2000 hypertension-related consumer health questions in Chinese were manually annotated based on this schema. The research findings demonstrated that health consumers were mainly concerned about what was wrong with their health (or the health of someone they cared about), why it was wrong, how to treat it (including choosing which provider to treat), whether the drugs they used had adverse effects or would do harm in some conditions (eg, pregnancy, breast feeding), whether they could recover from the illness, and what they could do to improve their health in everyday life (mainly diet suggestions).

We explored a machine-learning method to automatically classify these Chinese consumer health questions into one of the six primary level topics, with a novel scoring metric to select the most effective features from the abundant feature types we had explored. The results proved that selecting the features with $\Phi \geq \text{avg } \Phi$ in each feature type as input features for machine-learning not only increased the efficiency, but also improved the performance of the classifiers successfully. From the 6-category classification results, we achieved an average precision of 91.41%, recall of 89.62%, and F_1 score of 90.24%.

Comparison With Prior Work

Similarities and Differences in Questions Asked by Health Consumers and Providers

Compared with the 1396 clinical questions annotated by Ely et al [14,41], we found that while health consumers and providers both asked questions about diagnosis, treatment, condition management, and epidemiology, the questions posted by consumers were much more ambiguous. For instance, the frequency of questions with multiple findings was twice that of health providers' inquiries under the category of interpretation of clinical findings. It might be because consumers could not identify the most important findings, so they tended to list all the findings they knew. Although the frequency of treatment questions was almost equal in the two groups, health providers' questions were more specific to drug therapy (37.2% vs 22.1%), and they sometimes asked these questions on very specialized topics, such as composition, pharmacodynamics, action mechanism, and serum levels of drugs [14,41]. Such questions were rarely asked by health consumers. Moreover, health consumers would ask how to keep healthy or help in recovery in daily life, because many of them have recognized that lifestyle factors, such as diet, exercise, weight loss, and mood control, would impact their health status as well [42]. However, physicians seldom asked these questions during a patient encounter, possibly because they mainly focused on medical service rather than lifestyle advice [10]. Similarly, health consumers never asked questions about coordination with other providers, doctor-patient communication, doctor and patient education, administrative rules, ethics, and legal issues, because these tasks were usually regarded as health providers' responsibility. These findings affirmed again that, health consumers' information needs differed significantly from those of providers. Therefore, the existing classification schemas and automatic classification methods for clinical questions cannot be applied to consumer health questions directly [11].

Features Explored for Automatic Classifiers

Compared with other related studies on automatic question classification in the domain of health and medicine, we explored an abundant number of feature types for automatic classifiers. For example, Cao et al [17] explored the features of bag-of-words, n-grams, part-of-speech, UMLS concepts, and semantic types, as well as IDF to identify general topics of clinical questions. Patrick et al [18] used bag-of-words, Bigram, interrogative words, SNOMED category, verb and its subject, and verb and its object as feature sets to distinguish answerable and unanswerable ICU questions. Liu et al [19] picked bag-of-words, word length, question length, IDF, interrogative words, personal pronouns, indefinite pronouns, and auxiliary verbs as learning features to separate consumer questions from professional questions in the health domain. Roberts et al [20] explored the features of bag-of-words, part-of-speech, UMLS concepts, named entity, word length, IDF, and noun and verb head chunks to classify question types for consumer health questions in English, while Conway et al [43] used bag-of-words, n-grams, semantic UMLS types, and named entity as features to classify disease outbreak reports. In other words, bag-of-words, part-of-speech, and semantic types were the most commonly used features for question classification. Our work adopted all the effective features in the prior works with the UMLS concepts and semantic types replaced by CMeSH concepts and semantic types. We also explored three ways to extract the keywords from a question using the machine-learning features, that is, we took the first k words with maximum TF, IDF, or TF-IDF as keywords. In addition, we added noun or verb rear chunks + interrogative words according to the specialties of word order in the Chinese language. The results showed that it worked better than the interrogative words + noun or verb head chunks, which was commonly used for question classification in English.

Feature Selection for Automatic Question Classification

The feature selection methods in our work were quite different from other relative works, and it has proved that our methods were much more effective and easy. Cao et al [17], Liu et al [19], and Roberts et al [20]employed a method of combining different types of features without considering the threshold, in which they explored different combinations of different feature types and selected the best combination with the maximum F_1

score of the classifier. Thus, a feature type would be either picked up or rejected in their approach, which may have caused the loss of some effective features. Another disadvantage of this method was the difficulty in exploring all the possible combinations of different feature types. On the contrary, we adopted a much more efficient method to combine all the effective features from each feature type with $\Phi \geq \text{avg}\ \Phi$, which was also proved to be very effective (as described in the results section).

Performance of the Classifiers

The performance of the classifiers trained by our study was quite satisfying. The average F_1 scores for the four classifiers for *Condition management, Health lifestyle, Diagnosis, and Health provider choice* were 99.63%, 99.13%, 98.55%, and 96.35%, respectively. The results were more significant than those of other similar studies. For example, the classification of 13 general topics of clinical questions conducted by Cao et al [17], which reached the highest F_1 score (89.3%) on the classifier for *Pharmacology*, while the majority of the scores were between 70% and 80%, and the classification of 13 question types of consumer questions carried out by Roberts et al [20], which achieved the highest F_1 score (90.6%) on the classifier for *Management*, with 5 between 80% and 90%, and 5 below 70%. It is worth noting that the methods proposed in this paper and those in the related works were experimented on different datasets in different languages. Further, there were two main reasons for the differences between this study and others, although the same algorithm of SVMs was used. On the one hand, the feature types and the feature selection methods applied in this study were more efficient and effective (as discussed above); on the other hand, the classification schema used in this study was more distinguishable. For example, questions about *Etiology* or *Cause* and *Diagnosis* were very similar and always asked together. Thus, they were annotated as one topic (*Diagnosis)* in our study. However, Cao et al [17] and Roberts et al [20] annotated them as different types.

Limitations

One of the limitations of this work is that the sample questions we used to build the classification schema and to train the automatic classifiers were from only one Chinese health website and defined to be hypertension or blood pressure related. Therefore, the applicability of the classification schema and the validity of the automatic classifiers for the vast majority of questions from other websites and other diseases remain to be tested. Another limitation of this work is that some types of features, such as keywords and bag-of-words, might be correlated. However, our feature selection algorithm did not take the impact of correlation into consideration. We only reached moderate performances on the automatic classifiers for the general topics of Treatment and Epidemiology, whereas the reasons for this remain to be explored in the future.

Conclusions

One of the specialties of this research was that Chinese consumer health questions were chosen as the research object. We built a classification schema of consumer health questions which consisted of 48 quaternary categories and 35 annotation rules, and we annotated 2000 questions in Chinese that were randomly selected from nearly 100 thousand messages about hypertension. Then, by using these annotated questions as the corpus, we explored a machine-learning method to automatically classify Chinese consumer health questions into six general topics to facilitate users' information needs analysis and answer extraction. We explored an abundant number of feature types and adopted a novel method to select all the effective features with $\Phi \geq \text{avg}\ \Phi$. The results proved that our classification approach was relatively more efficient and effective as compared with similar studies.

Acknowledgments

This research was supported by the Chinese Academy of Medical Sciences (Grant No. 2016ZX330011) and the National Social Science Foundation of China (Grant No. 14BTQ032). The authors would like to thank Dr Chao Xu for his helpful suggestions on data processing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables.

[PDF File (Adobe PDF File), 53KB - jmir_v19i6e220_app1.pdf]

References

1. China Internet Network Information Center. cnnic. 2016. China internet network development state statistic report URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/201601/P020160122469130059846.pdf
2. Fox S. Pew Research Center. 2011. The social life of health information URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf
3. Liu Y. d.wanfangdata. 2011. Empirical research on impact of internet use on individual health behaviors URL: http://d.wanfangdata.com.cn/Thesis/D186454 [accessed 2017-05-27] [WebCite Cache ID 6qlZYIitE]

4. Chen AT. The Relationship between health management and information behavior over time: a study of the illness journeys of people living with fibromyalgia. J Med Internet Res 2016 Oct 25;18(10):e269 [FREE Full text] [doi: 10.2196/jmir.5309] [Medline: 27780794]
5. Fiksdal AS, Kumbamu A, Jadhav AS, Cocos C, Nelsen LA, Pathak J, et al. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. J Med Internet Res 2014;16(10):e224 [FREE Full text] [doi: 10.2196/jmir.3341] [Medline: 25348028]
6. Zeng QT, Kogan S, Plovnick RM, Crowell J, Lacroix E, Greenes RA. Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. Int J Med Inform 2004 Feb;73(1):45-55. [doi: 10.1016/j.ijmedinf.2003.12.015] [Medline: 15036078]
7. Mao XL, LI XM. A survey on question and answering systems. Journal of Frontiers of Computer Science and Technology 2012;6(3):193-207.
8. Zhen L, Wang X, Yang S. Overview on question classification in question-answering system. Journal of Anhui University of Technology (Natural Science) 2015;32(1):48-54.
9. Boot Cécile RL, Meijman FJ. Classifying health questions asked by the public using the ICPC-2 classification and a taxonomy of generic clinical questions: an empirical exploration of the feasibility. Health Commun 2010 Mar;25(2):175-181. [doi: 10.1080/10410230903544969] [Medline: 20390683]
10. Reeder B, Le T, Thompson HJ, Demiris G. Comparing information needs of health care providers and older adults: findings from a wellness study. Stud Health Technol Inform 2013;192:18-22. [Medline: 23920507]
11. Roberts K, Demner-Fushman D. Interactive use of online health resources: a comparison of consumer and professional questions. J Am Med Inform Assoc 2016 Jul;23(4):802-811. [doi: 10.1093/jamia/ocw024] [Medline: 27147494]
12. Gebel R, Okkes I. International classification of primary care (ICPC-2-NL). United Kingdom: Oxford University Press; 2000.
13. Schnall R, Cimino JJ, Currie LM, Bakken S. Information needs of case managers caring for persons living with HIV. J Am Med Inform Assoc 2011 May 01;18(3):305-308 [FREE Full text] [doi: 10.1136/jamia.2010.006668] [Medline: 21270131]
14. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, et al. A taxonomy of generic clinical questions: classification study. Br Med J 2000 Aug 12;321(7258):429-432 [FREE Full text] [Medline: 10938054]
15. Yan Z. Toward a layered model of context for health information searching: an analysis of consumer-generated questions. J Am Soc Inf Sci Technol 2013;64(6):1158-1172. [doi: 10.1002/asi.22821]
16. Roberts K, Masterton K, Fiszman M. Citeseerx.ist. 2014. Annotating question types for consumer health questions URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=9DE05B71F8A3E344512A3366E516A1B8?doi=10.1.1.673.6987&rep=rep1&type=pdf [accessed 2017-05-27] [WebCite Cache ID 6qlbO0DdO]
17. Cao Y, Cimino JJ, Ely J, Yu H. Automatically extracting information needs from complex clinical questions. J Biomed Inform 2010 Dec;43(6):962-971 [FREE Full text] [doi: 10.1016/j.jbi.2010.07.007] [Medline: 20670693]
18. Patrick J, Li M. An ontology for clinical questions about the contents of patient notes. J Biomed Inform 2012 Apr;45(2):292-306 [FREE Full text] [doi: 10.1016/j.jbi.2011.11.008] [Medline: 22142949]
19. Liu F, Antieau LD, Yu H. Toward automated consumer question answering: automatically separating consumer questions from professional questions in the healthcare domain. J Biomed Inform 2011 Dec;44(6):1032-1038 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.008] [Medline: 21856442]
20. Roberts K, Kilicoglu H, Fiszman M, Demner-Fushman D. Automatically classifying question types for consumer health questions. AMIA Annu Symp Proc 2014;2014:1018-1027 [FREE Full text] [Medline: 25954411]
21. Zhang N, Zhu L. A review of Chinese Q & A system questions. Technology Intelligence Engineering 2016:01-42.
22. Wu Y, Jiang M, Lei J, Xu H. Named entity recognition in Chinese clinical text using deep neural network. Stud Health Technol Inform 2015;216:624-628 [FREE Full text] [Medline: 26262126]
23. Zeng Q, Zhang X, Li Z. Extracting clinical information from free-text of pathology and operation notes via Chinese natural language processing. 2010 Dec 1 Presented at: IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW); 2010; Hong Kong p. 593-597.
24. Wang H, Zhang W, Zeng Q, Li Z, Feng K, Liu L. Extracting important information from Chinese Operation Notes with natural language processing methods. J Biomed Inform 2014 Apr;48:130-136 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.017] [Medline: 24486562]
25. Zhang S, Kang T, Zhang X, Wen D, Elhadad N, Lei J. Speculation detection for Chinese clinical notes: impacts of word segmentation and embedding models. J Biomed Inform 2016 Apr;60:334-341 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.011] [Medline: 26923634]
26. National Center for Cardiovascular Diseases of China. Report on cardiovascular diseases in China (2013). Beijing: Beijing Encyclopedia of China Publishing House; 2014.
27. XYWY. XYWY. 2016. Trustworthy Internet Healthcare Service Platform URL: http://www.xywy.com/about/index.html [accessed 2017-05-27] [WebCite Cache ID 6qldIeXre]
28. Li J. Csdn.net. 2016. Introduction of Rwordseg URL: http://download.csdn.net/detail/cl1143015961/8436741 [accessed 2017-05-27] [WebCite Cache ID 6qleBRsFs]

29. sewm.pku. ICTCLAS: Institute of Computing Technology, Chinese Lexical Analysis System URL: http://sewm.pku.edu.cn/QA/reference/ICTCLAS/FreeICTCLAS/
30. nlp.stanford. Stanford Log-linear Part-Of-Speech Tagger URL: https://nlp.stanford.edu/software/tagger.shtml
31. baike.baidu. Question words URL: http://baike.baidu.com/item/%E7%96%91%E9%97%AE%E8%AF%8D [accessed 2017-05-27] [WebCite Cache ID 6qle8GlKD]
32. Guo H, Li J, Dai T. Consumer health information needs and question classification: analysis of hypertension related questions asked by consumers on a Chinese health website. Stud Health Technol Inform 2015;216:810-814. [Medline: 26262164]
33. Niu YQ, Chen J, Duan LG. Study on classification features of Chinese interrogatives. Computer Application and Software 2012;29(3):108-111.
34. Sun HX, Li JL, Li DY. The study on semantic mapping from free word to subject headings based on semantic system of CMeSH. Data analysis and knowledge discovery 2013;29(11):46-51. [doi: 10.11925/infotech.1003-3513.2013.11.07]
35. Chen YW, Lin CJ. Combining SVMs with various feature selection strategies. In: Studies in Fuzziness and Soft Computing. Berlin, Heidelberg: Springer; 2006:315-324.
36. Yang L, Li J, Yao Y, Wu X. [A P300 detection algorithm based on F-score feature selection and support vector machines]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2008 Feb;25(1):23-6, 52. [Medline: 18435249]
37. Tsoumakas G, Katakis I. lpis.csd. 2016. Multi-label classification: an overview URL: http://lpis.csd.auth.gr/publications/tsoumakas-ijdwm.pdf
38. Gou B, Huang X. SVM multi-class classification. Journal of Data Acquisition & Processing 2006;21(3):334-339.
39. Wikipedia. Cross-validation (statistics) URL: https://en.wikipedia.org/wiki/Cross-validation_(statistics).pdf
40. Elliott AC, Woodward WA. Statistical analysis quick reference guidebook: with SPSS examples. United States: SAGE Publications, Inc; 2006.
41. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, et al. Analysis of questions asked by family doctors regarding patient care. Br Med J 1999 Aug 07;319(7206):358-361 [FREE Full text] [Medline: 10435959]
42. Publicity Department of National HealthFamily Planning Commission of the People?s Republic of China, Chinese Health Education Center. moh.gov. 2012 Chinese residents health literacy monitoring report URL: http://www.moh.gov.cn/zhuz/xzqq/201412/e09a26a97b914abb80172e85f46c5d84.shtml [accessed 2017-05-27] [WebCite Cache ID 6qlfPXCbQ]
43. Conway M, Doan S, Kawazoe A, Collier N. Classifying disease outbreak reports using n-grams and semantic features. Int J Med Inform 2009 Dec;78(12):e47-e58. [doi: 10.1016/j.ijmedinf.2009.03.010] [Medline: 19447070]

Abbreviations

TGCQ: Taxonomies of Generic Clinical Questions
LMCC: Layered Model of Context for Consumer Health Information Searching

Edited by G Eysenbach; submitted 14.12.16; peer-reviewed by S Zheng, Y Peng; comments to author 23.02.17; revised version received 20.04.17; accepted 20.04.17; published 20.06.17.

Please cite as:
Guo H, Na X, Hou L, Li J
Classifying Chinese Questions Related to Health Care Posted by Consumers Via the Internet
J Med Internet Res 2017;19(6):e220
URL: http://www.jmir.org/2017/6/e220/
doi:10.2196/jmir.7156
PMID:28634156

Original Paper

Interactive Voice Response System: Data Considerations and Lessons Learned During a Rectal Microbicide Placebo Adherence Trial for Young Men Who Have Sex With Men

Jose Bauermeister[1], MPH, PhD; Rebecca Giguere[2], MPH; Cheng-Shiun Leu[2,3], PhD; Irma Febo[4], MD; Ross Cranston[5], MBChB, MD, FRCP; Kenneth Mayer[6], MD; Alex Carballo-Diéguez[2], PhD

[1]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
[2]HIV Center for Clinical and Behavioral Studies, Columbia University and New York State Psychiatric Institute, New York, NY, United States
[3]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, United States
[4]Gama Project, Department of Pediatrics, University of Puerto Rico Medical Campus, San Juan, Puerto Rico
[5]Division of Infectious Diseases, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[6]Fenway Institute, Fenway Health, Boston, MA, United States

Corresponding Author:
Jose Bauermeister, MPH, PhD
Department of Family and Community Health
School of Nursing
University of Pennsylvania
418 Curie Blvd, Suite 402
Philadelphia, PA, 19104
United States
Phone: 1 2158989993
Fax: 1 2158989993
Email: bjose@nursing.upenn.edu

Abstract

Background: Rectal microbicides, if proven effective, may aid in reducing human immunodeficiency virus (HIV) incidence; however, demonstration of efficacy and effectiveness is contingent on accurate measurement of product adherence. Delays in self-report, in particular, may affect the accuracy of behavioral data.

Objective: The aim of this study was to capitalize on mobile phone use by young men who have sex with men (YMSM), and examine the use of an interactive voice response system (IVRS) by YMSM aged 18-30 years enrolled in a multisite, 12-week microbicide safety and acceptability trial.

Methods: YMSM (N=95) enrolled across 3 sites (Boston, Pittsburgh, and San Juan) were asked to report their use of an applicator applied placebo rectal gel product during receptive anal intercourse (RAI) using the IVRS. IVRS was available in Spanish and English. After the 12-week trial, we examined whether IVRS problems were associated with YMSM's sociodemographic characteristics (eg, age, race and ethnicity, and education), sexual behavior, or recruitment site. We used a multinomial logistic regression to compare YMSM who experienced no IVRS problems (n=40) with those who reported one IVRS problem (n=25) or two or more IVRS problems (n=30).

Results: We recorded 1494 IVRS calls over 12 weeks. Over half of the participants (55/95; 58%) experienced challenges using the IVRS during the 12-week trial. YMSM reporting greater RAI occasions during the trial were more likely to experience one (odds ratio [OR]=1.08, 95% CI (1.02-1.14); $P \leq .01$) or more (OR=1.10, 95% CI (1.03-1.16); $P \leq .001$) IVRS challenges. Greater educational attainment was associated with multiple IVRS challenges (OR=7.08, 95% CI (1.6-31.6); $P \leq .01$). Participants in the Puerto Rico site were most likely to report multiple IVRS problems.

Conclusions: Although IVRS was a useful data collection technology in our trial, several challenges experienced by English and Spanish speaking YMSM diminish its overall acceptability. We discuss strategies to optimize future development of IVRS data quality protocols based on lessons learned.

(J Med Internet Res 2017;19(6):e207) doi:10.2196/jmir.7682

KEYWORDS

user-computer interface; speech recognition software; HIV; survey methodology

Introduction

The interactive voice reporting system (IVRS) has gained popularity given its potential to collect time-stamped, prospective behavioral data, and to reduce participants' recall bias during data collection [1-3] For example, participants can call into the system and report their data, rather than waiting until their next scheduled face-to-face visit with the research team [4,5]. IVRS has several additional methodological benefits to face-to-face approaches [1]. These benefits include making data collection accessible at any time and from any place, easing individuals' ability to participate in research if they experience scheduling or transportation-related barriers, and having greater cost-savings than face-to-face approaches. IVRS can also reduce literacy concerns by allowing participants to hear the questions and respond verbally and/or using a numeric keypad, allowing for assessments across multiple languages, and deploying complex, tailored skip-patterns based on participants' answers [3]. From a technological standpoint, IVRS also reduces compatibility issues as participants can use their phone of choice to report their behaviors and reduce data entry errors through automation [6].

Researchers have used IVRS to monitor behaviors over short, intensive periods (eg, multiple assessments a day over a week) and to follow participants for longer study periods. Within HIV prevention and care, researchers have noted that IVRS is an acceptable method for participants to share sensitive information and might encourage greater privacy than face-to-face methods [7-11]. In a study comparing Hispanic college students' use of IVRS with a Self-Administered Questionnaire (SAQ) and a timeline follow-back (TLFB) over a 3-month interval, Schroeder and colleagues [9] found that participants' sexual and substance use behaviors were underreported in the TLFB and over-reported in the SAQ when compared with the daily IVRS reporting. In their analyses, however, Schroeder and colleagues noted that IVRS use varied based on participants' sociodemographic characteristics (eg, age and sexual orientation), with younger and sexual minority participants trending toward a greater likelihood of placing IVRS calls than older and heterosexual counterparts [9]. These findings suggest that participants' characteristics might result in differential acceptability and use of IVRS.

Building on prior research examining the use of IVRS in HIV prevention studies, we examined the use of an IVRS to collect behavioral data from a sample of YMSM recruited to participate in a multisite rectal microbicide acceptability and adherence trial. Beyond overall use, however, we examined the prevalence of IVRS-related problems experienced by YMSM during the 12-week trial. We then examined whether YMSM's sociodemographic characteristics (eg, age, sexual orientation, and educational attainment), sexual behaviors during the trial, and trial site (eg, Boston, Pittsburgh, and Puerto Rico) were associated with IVRS problems. We used our findings to discuss strategies to optimize future development of IVRS data quality protocols based on lessons learned in our trial.

Methods

Study Participants

Study data came from a larger project called Microbicide Safety and Acceptability in Young Men [12,13]. The study received institutional review board (IRB) approval from all participating institutions, and all participants signed informed consent. After screening (stage 1A), YMSM participated in a run-in period in which they were asked to apply a rectal placebo gel using a rectal-specific applicator (stage 1B), followed by a safety trial in which participants applied tenofovir 1% gel using a vaginal applicator for rectal delivery of the gel (stage 2). The study took place in 3 sites: Pittsburgh, PA; Boston, MA; and San Juan, PR. Study candidates were recruited from clinics, bars, clubs, newspaper advertisements, and social networks. Recruitment materials indicated that the investigators were looking for YMSM (aged 18-30 years) for a study about their sexual health and their feelings about rectally inserting a placebo gel resembling a microbicide gel currently under development before receptive anal intercourse (RAI). Full protocol description (Clinicaltrials.gov NCT01283360) is presented in detail elsewhere [12]. We focused our attention on stage 1B where the IVRS was used (December 2010 to October 2012).

Among participants who received medical clearance in stage 1A, we selected those fulfilling the more stringent eligibility criterion of having had condomless RAI within the prior 3 months to participate in stage 1B. This allowed us to focus on those with more recent potential risk and invite them to enroll in stage 1B. After undergoing an informed consent process, receiving risk reduction counseling and provision of condoms, and updating their medical history, participants received 20 rectal applicators filled with a placebo gel and instructions to insert the entire content of 1 applicator rectally within 90 min before each RAI episode. We used an applicator specifically designed for the delivery of a rectal microbicide [12] filled with hydroxyethylcellulose (HEC) gel. Hydroxyethylcellulose is also known as the "universal placebo" because of its use as placebo in most gel microbicide trials. Six weeks after their first stage 1B visit, participants returned for the mid-trial follow-up visit and were dispensed up to 20 additional applicators to ensure they had 20 on hand for the next 6 weeks. Six weeks after the mid-trial follow-up visit, participants returned for the final follow-up visit of stage 1B in which they completed a Web-based computer assisted self-interview (CASI) and semistructured interview that included questions on gel and applicator use. Participants received US $50 for each study visit (4 visits for stage 1A and 1B) and US $50 for a completed video teleconference interview. They also received US $1 per used applicator returned at visits 2 and 3.

Interactive Voice Response System (IVRS)

Over the 12 weeks of stage 1B, participants were instructed to call an IVRS after each instance of RAI and/or applicator insertion, or at least once a week if they did not have RAI. Participants could respond by voice or use their keypad. The

IVRS system was available in English or Spanish. During visit 2, participants generated a 4-digit password to identify themselves within the IVRS and were trained on how to enter data into the IVRS, including completing a mock call into the IVRS. At each call, participants were asked to report the number of times they had used the gel, whether they had RAI or inserted anything other than the gel, and whether they experienced any problems with the IVRS. Participants who did not call into the system at least once a week were contacted by the IVRS and reminded to log their behaviors. Furthermore, to encourage use of the IVRS as a data collection tool, we incentivized participants with US $1 per call with a maximum of US $30 and an additional US $10 bonus a month if they called at least once a week. In total, participants could earn up to US $60 in stage 1B for reporting their product use via IVRS.

IVRS data was downloaded into an excel (Microsoft) spreadsheet at the end of the trial. We examined the frequency with which participants reported a problem when entering their data within the IVRS system, as well as the number of call entries that were incomplete (eg, hung up call) or sequential (eg, participant called into IVRS and had to call back again to revise [eg, mistyping] or finalize their entry). For this analysis, we coded an IVRS problem as a call where participants reported experiencing an IVRS problem or when an IVRS call was incomplete (eg, hung up call) or duplicative (eg, participant called again to complete their IVRS entry). Participants who indicated that they had an IVRS problem since their last completed call were given the option to leave a voicemail indicating what went wrong. A member of the study team transcribed these voicemails in English or Spanish. Spanish comments were subsequently translated into English. We include examples from these transcriptions in the Results section to illustrate participants' problems when using the IVRS.

Participant Measures

Sociodemographic data were collected via a Web-based CASI. Demographic information included age, race or ethnicity (white, African American, Latino, mixed or other), sexual orientation, and highest educational attainment (1=8th grade or lower, 2=some high school, 3=high school or general educational development [GED], 4=partial college, 5=college graduate, 6=some graduate school, and 7=graduate school degree). Participants also indicated whether they were currently in school (0=no; 1=yes) and employed (0=no; 1=yes). Participants were also asked to report the number of receptive anal intercourse occasions in the prior 3 months. Participants were identified by their site's location (1=Pittsburgh, PA; 2=Boston, MA; and 3=San Juan, PR).

Data Analytic Strategy

After examining the descriptive statistics for our variables of interest, we used SPSS (version 23; IBM Corporation, New York) to test whether YMSM's sociodemographic characteristics, sexual behavior, or site of recruitment were associated with the odds of experiencing IVRS problems during the 12-week trial. We used a multinomial logistic regression to compare the odds of experiencing a single or multiple IVRS problems during the 12-week trial. Due to limited observations in several categorical variables (eg, race or ethnicity and sexual orientation), we collapsed these indicators into dummy variables (eg, white: yes or no and gay-identified: yes or no) in our regression analyses.

Results

Sample Description

Study participants (N=95) had a mean age of 23 years. The racial or ethnic composition of the sample was predominantly Latino (46/95, 48%) and white (33/95, 36%), followed by a fewer number of African American (9/95, 10%) and mixed or other race (7/95, 6%) participants. Most participants (86/95, 91%) identified as gay. Most of the sample (84/95, 88%) reported having at least some college education (see Table 1).

Table 1. Sociodemographic characteristics of young men who have sex with men (YMSM; N=95).

Variable		n (%)
Age (years), mean (SD[a])		23.2 (3.2)
Race or ethnicity		
	White or European American	33 (34.7)
	Black or African American	9 (9.5)
	Latino or Hispanic	46 (48.4)
	Mixed or other	7 (7.4)
Highest educational attainment		
	8th grade or lower	0 (0)
	Some high school	1 (1.1)
	High school graduate or GED[b]	10 (10.5)
	Some college	44 (46.3)
	College graduate	27 (28.4)
	Partial graduate school	2 (2.1)
	Graduate school degrees	11 (11.6)
	In school	47 (49.5)
	Currently employed	60 (63.2)
Sexual orientation		
	Gay	86 (90.5)
	Bisexual	9 (9.5)
Recruitment site		
	Pittsburgh	28 (29.5)
	Boston	26 (27.4)
	San Juan	41 (43.2)
Receptive anal intercourse occasions (before 3 months), mean (SD)		0.71)5.51(

[a] SD: standard deviation.

[b] GED: general educational development.

IVRS Calls

The IVRS recorded 1494 calls from 95 participants over 12 weeks. We flagged 162 calls (162/1494, 10.8%) reflecting an IVRS problem. The most common problems resulted from mistyping a numeric answer (74/95 participants, 46%), experiencing challenges entering their answers into the system (27/95 participants, 17%), being disconnected midway through the call and having to call back (58/95 participants, 36%), or the system being inaudible (3/95 participants, 2%). We include examples of these IVRS problems using participants' voicemail transcriptions in Table 2. Overall, 40 YMSM (42%, 40/95) did not have any trouble using the IVRS during the 12-week trial, 25 YMSM (26%, 25/95) had one problematic event across the 12-week trial, and the remaining 30 participants (32%, 30/95) reported two or more problems across the 12-week period.

Table 2. Examples of participants' comments regarding problems with the interactive voice response system (IVRS) during the trial.

Type of problem	Frequency n (%)	Exemplary quotes
Mistyping a number	74 (46)	"I mistakenly typed the wrong number; I pressed 1 when I should have pressed 2." "I had to repeat questions several times because it wasn't getting the button presses." "Had to call 5-6 times because every time I tried entering my User ID, it would say 'I'm sorry'." "(Problem with) entering my passcode." "The automated entry froze." "I mistakenly dialed that I didn't use the gel during sex, but I did use it."
Being disconnected midway through the call	58 (36)	"I've called twice already and the call drops". "Listen, I've tried to call 4 times and every time that I select an option, it indicates that I have nothing to say and hangs up on me. I've called 4 consecutive times. This time I pressed 1 to describe the situation." "Had to call back." "I was trying to record a message at the last time I called in and it stopped the message recording before I could say anything and it's happened more than once and I forget what I want to say the next time I call in because it doesn't let me re-record or anything."
Challenges reporting their answers into the system	27 (17)	"It didn't let me log into the system. Four days later it called me to indicate that I hadn't called in for that week". "The only problem with the system is that it kinda takes a bit long to go through it and you have to wait for each question to finish makes it difficult to use after you call you know what buttons to press and basically you have to wait and the call takes longer." "I did not say why I did not use the gel it was probably because my roommate was in the room and I did not want to give a voice command but it was because I didn't have sex that I didn't use the gel." "When I gave my answer to several questions, it skipped and said: 'I'm sorry, the answer is invalid'". "If there is any background noise while the phone system is waiting for me to press 1 or 2 it would say "I'm sorry that was not a valid-I'm sorry that was not a valid-I'm sorry that was not a valid" and if it was completely quiet then it would stop doing that and I was able to press the buttons."
Inaudible system	3 (2)	"Your voice is not very clear at all, it's kinda blurry at some points." "It was just cutting in and out."

Multivariable Regression

Using multinomial logistic regression (see Table 3), we examined whether IVRS problems were associated with participants' sociodemographic characteristics, sexual activity, or study site. Compared with YMSM who did not experience any IVRS problems, YMSM who experienced one IVRS problem during the trial were more likely to have a greater number of RAI occasions during the 12-week trial. No other differences were observed between YMSM who experienced a single IVRS problem and those who had no IVRS problems.

YMSM who experienced two or more IVRS problems were more likely than YMSM without IVRS problems to report a greater number of RAI occasions during the 12-week trial, to report greater educational attainment, and to be currently in school. Participants in Puerto Rico were more likely to report two or more IVRS problems than peers in the Boston site. There were no differences between Boston and Pittsburgh. No other differences were observed between YMSM who experienced multiple IVRS problem and those with no IVRS problems.

Table 3. Multinomial logistic regression of interactive voice response system (IVRS) problems over a 12-week period (N=95). Racial ethnic minorities serve as race or ethnicity referent group. Bisexual men serve as referent group for sexual orientation. Puerto Rico serves as referent group for recruitment sites.

Characteristics	1 IVRS problem during 12-week period (N=25)			2+ IVRS problems during 12-week period (N=30)		
	OR	95% CI	Significance	OR	95% CI	Significance
Intercept			.33			.04
Age	1.03	0.83-1.29	.77	1.04	0.81-1.32	.77
White	2.59	0.48-14.06	.27	4.69	0.48-45.29	.18
Educational attainment	1.31	0.64-2.65	.46	2.48	1.15-5.35	.02
In School	1.96	0.52-7.39	.32	7.08	1.59-31.60	.01
Is employed	0.92	0.23-3.72	.90	0.78	0.17-3.54	.75
Gay-identified	0.53	0.08-3.36	.50	0.52	0.05-5.33	.58
RAI occasions (3 months)	1.08	1.02-1.14	.01	1.10	1.03-1.16	.001
Pittsburgh	0.48	0.07-3.22	.45	0.09	0.01-1.25	.07
Boston	0.20	0.03-1.32	.09	0.05	0.01-0.55	.01
−2LL=144.33; X^2_{18}=42.5, P<.001		Nagelkerke Pseudo R^2=44.0%				

Discussion

Principal Findings

Technological advances continue to increase researchers' ability to capture behavioral data in experimental trials and observational studies. These data collection technologies (eg, IVRS, SMS text messages [short message service, SMS], and activity trackers) may be deployed to monitor participants' behaviors and contexts outside of the clinic setting and have the potential to be synchronized with other systems (eg, electronic medical records and e-based applications that support behavior change) [1]. In this study, we examined the use of an IVRS as a data collection tool in a rectal placebo gel acceptability and adherence trial with English and Spanish speaking YMSM. The IVRS recorded nearly 1500 calls over a 12-week trial period, highlighting its overall acceptability and feasibility among YMSM in HIV prevention studies.

Similar to other data collection methods, the use of IVRS as a data collection tool had its challenges. Ten percent of IVRS entries recorded during the trial were classified as having errors due to user (eg, user entering the wrong number to indicate their answer), system (eg, the IVRS not recognizing users' voice responses when there is background noise), or connectivity (eg, bad cellphone signal causing a dropped call) issues. As such, researchers employing IVRS as a data collection tool should ensure that time and attention is placed on data quality assurances before protocol implementation and during data cleaning. Furthermore, most of the IVRS challenges observed in our study might decrease over time as new innovations emerge. Improvements in wireless infrastructure (eg, better signal strength across the globe), capabilities and programing of IVRS (eg, better voice-response accuracy and automated data cleaning clarification questions), and devices (eg, mobile phone features) may alleviate participants' challenges when using an IVRS. For example, advances in IVRS programing could allow users to continue where they left off if they experience a problem (eg, call failure) and reduce frustrations stemming from having to restart their entry from the beginning. In situations when participants might feel uncomfortable verbalizing their answers (eg, someone walks into a room), designing opportunities for users to toggle their reporting through a multimodal response system (eg, switch from voice to text without interruption) may also improve IVRS data collection in real time.

Participants who reported greater RAI occasions during the trial were also more likely to experience IVRS challenges. Given that YMSM were instructed to call the IVRS every time they had RAI over the 12-week period, it is not surprising that those who contributed to a greater number of calls would have a greater likelihood of experiencing IVRS problems. After adjusting for participants' sexual activity, however, we noted that the probability of experiencing one or more IVRS challenges differed across participants' educational attainment and study site. YMSM with greater educational attainment reported more IVRS problems. These educational differences may be attributable to greater assertiveness to note problems with the system, and/or may be a proxy for socioeconomic differences regarding type of device (eg, mobile vs smart phone) owned. They may also have higher expectations about how efficient systems should work (they have more money, use high end devices, are more critical, and know what they can expect). In addition, participants in Puerto Rico were most likely to report having multiple IVRS challenges than peers in Boston. Although the IVRS had a Spanish version, it is possible that differences in participants' speed and enunciation when speaking Spanish made it harder for the system to capture their data accurately. Compared with the mainland United States, optimal cell phone signal and connectivity in Puerto Rico is lower—particularly in the more rural areas of the island. Taken together, our results support Schroeder and colleagues' findings [9] that participant sociodemographic characteristics may affect their use of the IVRS as a data collection tool.

Limitations

Our study has several limitations that deserve a mention. First, we did not ascertain participants' cellular plans or type of mobile phone during the study. Future research should consider how technological aspects (eg, type of phone and signal reliability) may affect data collection before IVRS implementation. In areas where signal strength or connectivity are a challenge, for instance, it may be better to rely on other methods (eg, an app notification system) that collects the information and stores it in the device in real time and subsequently transfers them to a central server once a connection is secured may be warranted. Third, our study focused on a sample of young men (ages 18-30 years) in 3 regions of the United States, limiting our ability to ascertain whether IVRS problems are similar or heightened in other age groups or contexts. Finally, our study took place from December 2010 to October 2012. The quality of IVRS is likely to have improved given the fast-pace of telecommunication innovations in society. Nevertheless, our study points to key data collection issues that must be considered as researchers plan and implement studies that rely on innovative data collection systems.

Overall, our study findings support the notion that IVRS is a feasible and acceptable method to collect time-stamped, prospective behavioral data from YMSM. As with other data collection methods, we encourage researchers to devote time and attention to the adequacy of IVRS for their populations of interest. Data quality assurances before protocol implementation, including considering how varying connectivity may create data collection challenges, are warranted. Nevertheless, even though some individuals might experience challenges, the interest and perseverance of YMSM to use this technology forecasts interesting possibilities that need to be explored in future research.

Acknowledgments

We greatly appreciate the hard work of the study staff at the sites and are indebted to the study participants for volunteering their time. This research was sponsored by the US National Institutes of Health (NIH), including National Institute of Child Health and Human Development (NICHD) and National Institute of Mental Health (NIMH), under R01 HD59533 (Carballo-Diéguez and McGowan, Co-PIs) and cosponsored by CONRAD. Additional support came from the NIHM to the HIV Center for Clinical and Behavioral Studies at NY State Psychiatric Institute and Columbia University (P30-MH43520; Principal Investigator: Robert Remien). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

Conflicts of Interest

None declared.

References

1. Kennedy CM, Powell J, Payne TH, Ainsworth J, Boyd A, Buchan I. Active assistance technology for health-related behavior change: an interdisciplinary review. J Med Internet Res 2012 Jun 14;14(3):e80 [FREE Full text] [doi: 10.2196/jmir.1893] [Medline: 22698679]
2. Lee H, Friedman ME, Cukor P, Ahern D. Interactive voice response system (IVRS) in health care services. Nursing Outlook 2003 Dec;51(6):277-283. [doi: 10.1016/S0029-6554(03)00161-1] [Medline: 14688763]
3. Ruikar V. Interactive voice/web response system in clinical research. Perspect Clin Res 2016;7(1):15-20 [FREE Full text] [doi: 10.4103/2229-3485.173781] [Medline: 26952178]
4. Fink JC, Doerfler RM, Yoffe MR, Diamantidis CJ, Blumenthal JB, Siddiqui T, et al. Patient-reported safety events in chronic kidney disease recorded with an interactive voice-inquiry dial-response system: monthly report analysis. J Med Internet Res 2016 May 26;18(5):e125 [FREE Full text] [doi: 10.2196/jmir.5203] [Medline: 27230267]
5. Steinberg DM, Levine EL, Lane I, Askew S, Foley PB, Puleo E, et al. Adherence to self-monitoring via interactive voice response technology in an eHealth intervention targeting weight gain prevention among Black women: randomized controlled trial. J Med Internet Res 2014 Apr 29;16(4):e114 [FREE Full text] [doi: 10.2196/jmir.2996] [Medline: 24780934]
6. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. J Med Internet Res 2006 Dec 11;8(4):e30 [FREE Full text] [doi: 10.2196/jmir.8.4.e30] [Medline: 17213048]
7. Brinkel J, Dako-Gyeke P, Krämer A, May J, Fobil JN. An investigation of users' attitudes, requirements and willingness to use mobile phone-based interactive voice response systems for seeking healthcare in Ghana: a qualitative study. Public Health 2017 Mar;144:125-133. [doi: 10.1016/j.puhe.2016.11.017] [Medline: 28274374]
8. Daftary A, Hirsch-Moverman Y, Kassie GM, Melaku Z, Gadisa T, Saito S, et al. A qualitative evaluation of the acceptability of an interactive voice response system to enhance adherence to isoniazid preventive therapy among people living with HIV in Ethiopia. AIDS Behav 2016 May 24:- Epub ahead of print. [doi: 10.1007/s10461-016-1432-8] [Medline: 27221743]
9. Schroder KE, Johnson CJ, Wiebe JS. Interactive voice response technology applied to sexual behavior self-reports: a comparison of three methods. AIDS Behav 2007 Mar;11(2):313-323. [doi: 10.1007/s10461-006-9145-z] [Medline: 16841191]
10. Schroder KE, Johnson CJ. Interactive voice response technology to measure HIV-related behavior. Curr HIV/AIDS Rep 2009 Oct 14;6(4):210-216. [doi: 10.1007/s11904-009-0028-6] [Medline: 19849964]

11. Tucker JA, Blum ER, Xie L, Roth DL, Simpson CA. Interactive voice response self-monitoring to assess risk behaviors in rural substance users living with HIV/AIDS. AIDS Behav 2012 Feb;16(2):432-440 [FREE Full text] [doi: 10.1007/s10461-011-9889-y] [Medline: 21311964]
12. Carballo-Diéguez A, Giguere R, Dolezal C, Bauermeister J, Leu C, Valladares J, et al. Adherence to rectal gel use among mainly ethnic minority young men who have sex with men during a 3-month placebo gel trial: implications for microbicide research. AIDS Behav 2014 Sep;18(9):1726-1733 [FREE Full text] [doi: 10.1007/s10461-014-0768-1] [Medline: 24719200]
13. McGowan I, Cranston RD, Mayer KH, Febo I, Duffill K, Siegel A, et al. Project gel a randomized rectal microbicide safety and acceptability study in young men and transgender women. PLoS One 2016;11(6):e0158310 [FREE Full text] [doi: 10.1371/journal.pone.0158310] [Medline: 27362788]

Abbreviations

GED: general educational developement
HEC: hydroxyethylcellulose
IRB: institutional review board
IVRS: interactive voice recognition system
RAI: receptive anal intercourse
SAQ: Self-Administered Questionnaire
TLFB: timeline follow-back
YMSM: young men who have sex with men

Edited by G Eysenbach; submitted 10.03.17; peer-reviewed by J Galea, D Matthews, N Bragazzi; comments to author 19.04.17; revised version received 23.04.17; accepted 24.04.17; published 09.06.17.

Please cite as:
Bauermeister J, Giguere R, Leu CS, Febo I, Cranston R, Mayer K, Carballo-Diéguez A
Interactive Voice Response System: Data Considerations and Lessons Learned During a Rectal Microbicide Placebo Adherence Trial for Young Men Who Have Sex With Men
J Med Internet Res 2017;19(6):e207
URL: http://www.jmir.org/2017/6/e207/
doi:10.2196/jmir.7682
PMID:28600275

Original Paper

Social Media Use and Access to Digital Technology in US Young Adults in 2016

Andrea C Villanti[1,2,3], MPH, PhD; Amanda L Johnson[1], MHS; Vinu Ilakkuvan[4], MS; Megan A Jacobs[5], MPH; Amanda L Graham[1,5,6], PhD; Jessica M Rath[3,7,8], MPH, PhD

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

[2]Department of Psychiatry, University of Vermont, Burlington, VT, United States

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

[4]Milken Institute School of Public Health, Department of Prevention and Community Health, George Washington University, Washington, DC, United States

[5]Innovations, Truth Initiative, Washington, DC, United States

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

[7]Evaluation Science and Research, Truth Initiative, Washington, DC, United States

[8]Department of Behavioral and Community Health, University of Maryland School of Public Health, College Park, MD, United States

Corresponding Author:
Andrea C Villanti, MPH, PhD
Schroeder Institute for Tobacco Research and Policy Studies at Truth Initiative
900 G Street NW, Fourth floor
Washington, DC, 20001
United States
Phone: 1 202 454 5751
Fax: 1 202 454 5785
Email: avillanti@truthinitiative.org

Abstract

Background: In 2015, 90% of US young adults with Internet access used social media. Digital and social media are highly prevalent modalities through which young adults explore identity formation, and by extension, learn and transmit norms about health and risk behaviors during this developmental life stage.

Objective: The purpose of this study was to provide updated estimates of social media use from 2014 to 2016 and correlates of social media use and access to digital technology in data collected from a national sample of US young adults in 2016.

Methods: Young adult participants aged 18-24 years in Wave 7 (October 2014, N=1259) and Wave 9 (February 2016, N=989) of the Truth Initiative Young Adult Cohort Study were asked about use frequency for 11 social media sites and access to digital devices, in addition to sociodemographic characteristics. Regular use was defined as using a given social media site at least weekly. Weighted analyses estimated the prevalence of use of each social media site, overlap between regular use of specific sites, and correlates of using a greater number of social media sites regularly. Bivariate analyses identified sociodemographic correlates of access to specific digital devices.

Results: In 2014, 89.42% (weighted n, 1126/1298) of young adults reported regular use of at least one social media site. This increased to 97.5% (weighted n, 965/989) of young adults in 2016. Among regular users of social media sites in 2016, the top five sites were Tumblr (85.5%), Vine (84.7%), Snapchat (81.7%), Instagram (80.7%), and LinkedIn (78.9%). Respondents reported regularly using an average of 7.6 social media sites, with 85% using 6 or more sites regularly. Overall, 87% of young adults reported access or use of a smartphone with Internet access, 74% a desktop or laptop computer with Internet access, 41% a tablet with Internet access, 29% a smart TV or video game console with Internet access, 11% a cell phone without Internet access, and 3% none of these. Access to all digital devices with Internet was lower in those reporting a lower subjective financial situation; there were also significant differences in access to specific digital devices with Internet by race, ethnicity, and education.

Conclusions: The high mean number of social media sites used regularly and the substantial overlap in use of multiple social media sites reflect the rapidly changing social media environment. Mobile devices are a primary channel for social media, and our study highlights disparities in access to digital technologies with Internet access among US young adults by race/ethnicity, education, and subjective financial status. Findings from this study may guide the development and implementation of future health interventions for young adults delivered via the Internet or social media sites.

(J Med Internet Res 2017;19(6):e196) doi:10.2196/jmir.7303

KEYWORDS

social media; technology; young adults

Introduction

Young adulthood, frequently defined as the period from the late teens through the 20s, is a critical time for self-identity construction [1,2]. According to Arnett's theory of emerging adulthood, identity exploration centers on love, work, and worldviews [2]. Through experimentation with various life possibilities, young people move toward enduring decisions and values during this developmental period. Social media contributes to identity formation by facilitating reflexivity, or the process of reflecting on how we see ourselves and how others see us [3]. In 2015, 90% of US young adults with Internet access used social media [4]. Social media plays an important role in shaping self-beliefs and perceived social norms [5] by inviting other people's feedback and shaping further modification of self-representations. Social media gives users an opportunity to control their public personae [6] and to experiment with more desirable versions of themselves. This can be done through a number of aspects of social media, including text, images, video, music, affinities ("likes"), visual customization of social media, use of profanity or slang, and group membership or affiliation, to name just a few.

Behavior change theories that address social norms, including social cognitive theory, provide a theoretical framework in which to understand the influence of social media on identity formation [7,8]. Through social observation and interaction, people gather important information from others that can be used to direct their own behavior and beliefs [9] through "electronic acculturation." Digital and social media are significant modalities through which young adults explore identity formation [10-13], and by extension, learn and transmit norms about health and risk behaviors during this developmental period. Recent studies from a sample of US young adults have also highlighted potentially harmful effects of high social media use on mental health, including lower perceived emotional support [14], greater perceived social isolation [15], anxiety [16], and depression [16,17].

Annual data from the Pew Research Center on Internet, Science & Tech provide the only national benchmarks of the uptake and prevalence of social media and technology use among US teenagers and adults. These reports frequently present a breakdown of the demographic characteristics associated with use of digital technologies, including by age. However, there is limited detail on the characteristics of technology users within a specific age group, except for teens. Young adults have been characterized as hard-to-reach with regard to health promotion and disease prevention, but are widely accessing social media. Given young adults' unique relationship with technology [18], social media could be a powerful venue for health intervention delivery, including mental health and substance use interventions, in this group. The purpose of this study was to provide updated estimates of social media use from 2014 to 2016, and correlates of social media use, including psychosocial measures, common risk behaviors, and access to digital technology in data collected from a national sample of US young adults aged 18-24 years in 2016. The broader goal of this work is to disseminate these data to inform health intervention efforts for young adults.

Methods

Study Sample

The Truth Initiative Young Adult Cohort Study was designed to understand the trajectories of tobacco use in a young adult population using a longitudinal cohort sample. Details of the cohort have been described elsewhere [19]. Briefly, the cohort comprises a nationally representative sample of young adults of ages 18-34 years drawn from GfK (*Gesellschaft für Konsumforschung* [Society for Consumer Research])'s KnowledgePanel. KnowledgePanel is a Web-based panel of adults of ages 18 years and older that covers those who do and do not use the Internet in the United States [20]. The panel was recruited via address-based sampling, a probability-based random sampling method that provides a statistically valid representation of the US population, including cell phone-only households; at the time of the baseline survey, the KnowledgePanel was one of the only providers that used this methodology. African American and Hispanic young adults were oversampled to ensure sufficient sample sizes for subgroup analyses. The validity of this methodology has been reported previously [21,22], and it has been used broadly in the peer-reviewed medical literature [23-27]. The baseline survey (Wave 1; n=4201) was conducted in July 2011, with subsequent assessments occurring approximately every six months; the study is ongoing with the most recent Wave 10 collected in October 2016. The cohort is refreshed at each wave to maintain the initial sample size.

This study used data from responses of participants aged 18-24 years to the Wave 7 survey (October 2014, N=1259) and the Wave 9 survey (February 2016, N=1023), as these two waves included the relevant items on social media use. In Wave 9, 34 participants (3.3%, 34/1023) were missing data on covariates and excluded, yielding an analytic sample of N=989. For the full sample of respondents of ages 18-34 years, the panel recruitment rate was 13.9% in Wave 7 and 13.2% in Wave 9 [28]. In 64% of the identified households, one member completed a core profile survey in which the key demographic information was collected (profile rate—PROR). For this particular study, only one panel member per household was selected at random to be part of the study sample and no members outside the panel were recruited. The response rate

(RR6) for Wave 7 was 61.1% and 60.7% for Wave 9. Thus, the cumulative response rate (CUMRR1; the product of these three rates) was 5.4% for Wave 7 and 5.1% for Wave 9. This study was approved by the Chesapeake Institutional Review Board, Inc, and consent was collected from participants before survey self-administration.

Measures

Social Media Site Use

Participants were asked about frequency of use for 10 social media sites: Twitter, Facebook, Instagram, Snapchat, Vine, Pinterest, Tumblr, Google+, LinkedIn, YouTube, and other sites not otherwise specified. Response options were “never,” “less than 1 time per month,” “monthly,” “weekly,” “daily,” or “multiple times per day.” Participants who reported that they used a social media site at least weekly were defined as regular users of that site. Participants who reported regular use of at least one social media site were classified as regular social media users. In 2014, participants were also asked to enumerate, in an open-ended response, approximate counts of their Facebook friends, Twitter followers, people they follow on Twitter, Instagram followers, and people they follow on Instagram.

Access to Digital Technology

All participants were asked about their access to digital devices, using the following item: “Which of the following types of digital devices do you have access to or use? Select all that apply” with the following response choices: “A smartphone with Internet access (for example, iPhone, Android, Blackberry etc)”; “A tablet with Internet access (for example, iPad, Samsung Galaxy Tab etc)”; “A desktop or laptop computer with Internet access”; “A Smart TV or video game console that has Internet access”; “A cell phone without Internet access (to talk and text).” Participants could also endorse the single response choice, “None of these.”

Other Covariates

As part of KnowledgePanel routine baseline data collection, participants provided information on age at study entry, gender, race/ethnicity (white, non-Hispanic; black, non-Hispanic; other, non-Hispanic; and Hispanic), and highest education completed (less than high school, high school, and some college or greater). GfK conducted hot deck imputation to handle missing data on age, gender, race/ethnicity, and education level. A measure of subjective financial status, developed and validated in the Truth Initiative Young Adult Cohort Study [29], was included as it has been shown to be a more robust measure of socioeconomic status for young adults than income or educational attainment. The item asked: “Considering your own income and the income from any other people who help you, how would you describe your overall personal financial situation?” with response options “don’t meet basic needs,” “just meet basic needs,” “meet needs with a little left,” and “live comfortably.” Two separate variables to assess for symptoms consistent with current depression and current anxiety were created from the two-item Patient Health Questionnaire (PHQ-2) [30] or the two-item Generalized Anxiety Disorder (GAD) scale [31], respectively. Using a four-point scale (0 = “not at all” and 3 = “nearly every day”), the two-item PHQ queries about the frequency of depressed mood and loss of interest in pleasant activities, whereas the two-item GAD queries about the frequency of uncontrollable worry and feelings of anxiety in the past two weeks. Per the scoring rubric for each scale, individuals who received a score at or above the cut-off (>3) were coded as having symptoms of depression or anxiety, respectively. Past 30-day alcohol and marijuana use were determined by two items. For alcohol, the first item asked about frequency of drinking alcohol in the past year (“never,” “monthly or less,” “2-4 times per month,” “2-3 times per week,” “4 or more times per week”); for marijuana use, participants were asked about the frequency of their current use, with response options “every day,” “some days,” and “not at all.” Those who reported any use of alcohol or marijuana were then queried about the frequency of use in the past 30 days, with respondents who reported using ≥1 day in the past month defined as current users.

Analytic Plan

All analyses were performed using Stata/SE 14.0 (StataCorp, 2014) in August 2016. Poststratification weights were used to offset any nonresponse or noncoverage bias and produce nationally representative estimates specific to each wave of data collection. Missing data were handled with list-wise deletion per Stata’s survey procedures as there was a low proportion of missing data on relevant covariates.

Bivariate analyses were conducted using the survey commands in Stata to provide the estimates of use of each social media site and overlap between use of sites, and to assess the correlates of using a greater number of social media sites regularly. Differences in prevalence estimates were assessed using *P* values from the crude log-binomial regressions of regular use of each social media site and crude linear regressions of number of social media sites on demographics (Wald tests, $P<.05$). Adjusted log-binomial regressions were used to examine the associations of sociodemographic variables on regular use of each social media site and on the number of social media sites regularly used ($P<.05$). Analyses conducted for the items on access to digital technology comprised univariate and bivariate analyses examining the overall prevalence of access to specific digital devices and differences in use of various devices by sociodemographic characteristics.

Results

Participant Characteristics

In 2014, 89.42% (1126/1298) of young adults aged 18-24 years reported regular use of at least one social media site (Table 1). This increased to 97.5% (965/989) of young adults in 2016. In 2016, there were 989 (weighted) young adults aged 18-24 years who were regular social media users (Table 2). Of these, 49.7% (492/989) were male, 52.3% (517/989) white, non-Hispanic, 58.1% (575/989) had at least some college education, and 63.4% (628/989) reported a financial situation that at least met their needs with a little left. Also, 6.0% (60/989) of the sample reported symptoms of depression and 7.0% (70/989) reported symptoms of anxiety.

Table 1. Percentage of US young adults aged 18-24 years who regularly use social media sites.

Social media sites	Percentage among all young adults (18-24 years of age)				Percentage among young adults who regularly^a use social media			
	Wave 7; N=1259	Wave 9; N=989	Relative percent difference compared with Wave 7	Absolute percent difference compared with Wave 7	Wave 7; N=1126; 89.44%	Wave 9; N=965; 97.57%	Relative percent difference compared with Wave 7	Absolute percent difference compared with Wave 7
Facebook	947 (75.24)	758 (76.6)	102	1	947 (84.15)	758 (78.5)	93	−6
YouTube	814 (64.68)	541 (54.6)	84	−10	814 (72.33)	541 (56.0)	77	−16
Instagram	465 (36.89)	779 (78.7)	213	42	465 (41.26)	779 (80.7)	196	39
Snapchat	387 (30.75)	789 (79.7)	259	49	387 (34.39)	789 (81.7)	238	47
Twitter	347 (27.58)	754 (76.2)	276	49	347 (30.85)	754 (78.1)	253	47
Google+	278 (22.06)	757 (76.5)	347	54	278 (24.67)	757 (78.5)	318	54
Pinterest	220 (17.51)	700 (70.8)	404	53	220 (19.58)	700 (72.6)	371	53
Tumblr	140 (11.08)	825 (83.4)	752	72	140 (12.39)	825 (85.5)	690	73
Vine	108 (8.56)	818 (82.7)	966	74	108 (9.57)	818 (84.8)	885	75
LinkedIn	106 (8.44)	761 (76.9)	911	68	106 (9.44)	761 (78.9)	836	69
Other sites	49 (3.86)	808 (81.7)	2119	78	49 (4.31)	808 (83.7)	1942	79

[a] Regular use is defined as using a site multiple times a day, daily, or weekly.

Social Media Use

Table 1 compares the popularity of 10 social media sites available in 2014 and 2016. Among regular social media users (weighted n=1126) in 2014, the top 5 social media sites were Facebook (947/1126, 84.15%), YouTube (814/1126, 72.33%), Instagram (465/1126, 41.26%), Snapchat (387/1126, 34.39%), and Twitter (347/1126, 30.85%). Among regular social media users in 2016 (weighted n=965) the top 5 sites were Tumblr (825/965, 85.5%), Vine (818/965, 84.8%), Snapchat (789/965, 81.7%), Instagram (779/965, 80.7%), and LinkedIn (761/965, 78.9%). Google+ (757/965, 78.5%), Facebook (758/965, 78.5%), and Twitter (754/965, 78.1%) had similar levels of regular use.

In 2014, respondents listed the greatest number of friends/followers on Facebook (n=292), followed by Instagram (followers: n=167; people you follow on Instagram: n=160) and Twitter (followers: n=111; people you follow on Twitter: n=128).

Correlates of regular use among the top 5 social media sites in 2016 are presented in Table 2. Compared with males, females were slightly less likely to regularly use Tumblr (n=426 vs 399; 86.6% vs 80.2%). There were no significant differences in regular use of a site by race/ethnicity or education, with the exception of LinkedIn. A greater proportion of regular LinkedIn users were black, non-Hispanic (PR 1.12, 95% CI 1.02-1.23) than white, non-Hispanic. Prevalence of LinkedIn use was also higher among those with less than high school education (PR 1.25, 95% CI 1.12-1.39) or a high school education (PR 1.24, 95% CI 1.15-1.33) compared with those with some college education. Respondents who reported subjective financial situation as “just meeting basic needs” were less likely to report regular use of Tumblr compared with those who “meet needs with a little left” (PR 0.91, 95% CI 0.85-0.98). Young adults who “don’t meet basic needs” had a significantly higher prevalence of using Snapchat regularly compared with those who “meet needs with a little left” (PR 1.12, 95% CI 1.02-1.24). There were no differences in regular use of any of the top 5 most popular social media sites by mental health covariates, but past 30-day alcohol users reported a lower prevalence of Snapchat use (PR 0.88, 95% CI 0.82-0.94).

Respondents reported regularly using an average of 7.6 social media sites (range 0-11 sites), with 3.0% (30/989) reporting regular use of 0-1 social media sites, 2.8% (27/989) using 2 or 3 sites, 9.1% (90/989) using 4 or 5 sites, and 85.1% (843/989) using 6 or more sites regularly. Table 3 presents the mean number of social media sites used regularly and correlates of the number of sites used regularly. In bivariate and multivariable analyses, having a high school education and past 30-day alcohol use were positively associated with using a greater number of social media sites compared with those with some college education and no past 30-day alcohol use, respectively.

Table 2. Percentages and correlates of regular social media use, by site, among US young adults aged 18-24 years (weighted n=989).

Variable	Total	Tumblr		Vine		Snapchat		Instagram		LinkedIn	
	(N[a]=989)	(N[a]=825)		(N[a]=818)		(N[a]=789)		(N[a]=779)		(N[a]=761)	
	%	%	PR[b] (95% CI)	%	PR (95% CI)	%	PR (95% CI)	%	PR (95% CI)	%	PR (95% CI)
Overall		83.4		82.7		79.7		78.7		76.9	
Sex											
Male	49.8	86.6	Ref[c]	84.2	Ref	77.7	Ref	78.7	Ref	78.9	Ref
Female	50.3	80.2	0.93 (0.87-0.98)[d]	81.2	0.96 (0.91-1.03)	81.8	1.05 (0.98-1.13)	78.7	1.00 (0.93-1.08)	75.0	0.95 (0.88-1.02)
Race/ethnicity											
White, NH[e]	52.3	84.3	Ref	81.0	Ref	79.9	Ref	79.9	Ref	75.3	Ref
Black, NH	12.9	87.2	1.03 (0.95-1.13)	84.6	1.04 (0.94-1.16)	79.3	0.99 (0.88-1.11)	79.9	1.00 (0.89-1.12)	84.6	1.12 (1.02-1.23)[d]
Other, NH	8.8	83.6	0.99 (0.89-1.11)	82.9	1.02 (0.90-1.17)	77.9	0.97 (0.84-1.13)	76.8	0.96 (0.82-1.13)	70.4	0.93 (0.77-1.13)
Hispanic	26.0	79.6	0.94 (0.87-1.02)	85.0	1.05 (0.98-1.13)	80.2	1.00 (0.92-1.09)	76.3	0.96 (0.87-1.05)	78.6	1.04 (0.96-1.14)
Education											
Less than high school	12.1	83.1	1.00 (0.89-1.12)	86.0	1.06 (0.95-1.18)	85.1	1.09 (0.97-1.22)	80.5	1.03 (0.91-1.16)	87.5	1.25 (1.12-1.39)[d]
High school	29.8	84.4	1.02 (0.95-1.09)	84.0	1.03 (0.96-1.11)	80.6	1.03 (0.95-1.12)	78.5	1.00 (0.92-1.09)	86.4	1.24 (1.15-1.33)[d]
Some college	58.1	82.9	Ref	81.3	Ref	78.2	Ref	78.4	Ref	69.9	Ref
Financial Situation											
Don't meet basic needs	6.6	80.0	0.91 (0.79-1.05)	88.0	1.07 (0.96-1.18)	90.6	1.12 (1.02-1.24)[d]	77.2	0.96 (0.81-1.13)	80.6	1.03 (0.88-1.20)
Just meet basic needs	30.0	80.1	0.91 (0.85-0.98)[d]	83.0	1.00 (0.93-1.08)	75.9	0.94 (0.86-1.03)	74.5	0.92 (0.84-1.01)	76.1	0.97 (0.89-1.06)
Meet needs with a little left	38.0	87.9	Ref	82.6	Ref	80.6	Ref	80.6	Ref	78.5	Ref
Live comfortably	25.5	81.5	0.93 (0.86-1.00)	81.0	0.98 (0.90-1.06)	80.1	0.99 (0.91-1.08)	81.3	1.01 (0.93-1.10)	74.5	0.95 (0.87-1.04)
Depression											
No	94.0	83.6	Ref	82.7	Ref	79.5	Ref	79.0	Ref	76.5	Ref
Yes	6.0	79.2	0.95 (0.82-1.09)	82.1	0.99 (0.87-1.13)	82.6	1.04 (0.91-1.18)	73.7	0.93 (0.78-1.11)	82.8	1.08 (0.96-1.22)
Anxiety											

Variable	Total	Tumblr		Vine		Snapchat		Instagram		LinkedIn	
	(N[a]=989)	(N[a]=825)		(N[a]=818)		(N[a]=789)		(N[a]=779)		(N[a]=761)	
	%	%	PR[b] (95% CI)	%	PR (95% CI)	%	PR (95% CI)	%	PR (95% CI)	%	PR (95% CI)
No	93.0	83.4	Ref	83.5	Ref	79.7	Ref	79.0	Ref	76.9	Ref
Yes	7.0	82.8	0.99 (0.88-1.12)	71.9	0.86 (0.73-1.01)	80.5	1.01 (0.89-1.15)	74.9	0.95 (0.81-1.11)	77.8	1.01 (0.88-1.16)
Past 30-day alcohol use											
No	53.0	83.2	Ref	85.0	Ref	84.5	Ref	81.2	Ref	79.8	Ref
Yes	47.0	83.6	1.00 (0.95-1.07)	80.1	0.94 (0.89-1.00)	74.3	0.88 (0.82-0.94)[b]	75.9	0.93 (0.87-1.00)	73.7	0.92 (0.86-0.99)[d]
Past 30-day marijuana use											
No	88.0	83.2	Ref	82.8	Ref	80.0	Ref	79.0	Ref	77.1	Ref
Yes	12.0	85.0	1.02 (0.93-1.12)	81.4	0.98 (0.89-1.08)	77.9	0.97 (0.87-1.09)	76.3	0.96 (0.85-1.09)	75.3	0.98 (0.87-1.10)

[a] Weighted.

[b] PR: prevalence ratio.

[c] Ref: Reference category.

[d] Denote statistical significance at *P*<.05.

[e] NH: non-Hispanic.

Table 4 presents a matrix of the overlap between use of the top 5 social media sites among young adults. These findings highlight significant overlap (81%-90%) in use of multiple social media sites.

Access to Digital Technology

Access to specific types of digital technologies among young adults was as follows: a smartphone with Internet access, 86.9% (860/989); a desktop or laptop computer with Internet access, 74.3% (736/989); a tablet with Internet access, 40.6% (401/989); a smart TV or video game console with Internet access, 29.0% (287/989); a cell phone without Internet access, 11.5% (114/989); none of these, 3.0% (30/989; Table 5). Females were significantly less likely to report having a smart TV or video game console with Internet access compared with males (PR 0.76, 95% CI 0.62-0.94). Females, however, reported greater access to a smartphone with Internet access (PR 1.07, 95% CI 1.01-1.14) compared with males. Compared with non-Hispanic whites, non-Hispanic black young adults had a significantly lower prevalence of access to a smartphone with Internet access (77.4%, 99/128 vs 89.6%, 463/517; PR 0.86, 95% CI 0.75-0.99), a desktop or laptop with Internet access (59.6%, 76/128 vs 80.4%, 416/517; PR 0.74, 95% CI 0.60-0.91), and a smart TV or video game console with Internet access (16.2%, 21/128 vs 30.2%, 156/517; PR 0.54, 95% CI 0.32-0.89). Hispanic young adults also reported lower prevalence of access to a desktop or laptop with Internet access (68.6%, 176/257 vs 80.4%, 416/517; PR 0.85, 95% CI 0.78-0.96). Respondents with a high school education or less reported significantly lower prevalence of access to a smartphone (80.3%, 333/414 vs 91.6%, 527/575; PR 0.88, 95% CI 0.82-0.94), a tablet (33.6%, 139/414 vs 45.6%, 262/575; PR 0.74, 95% CI 0.61-0.89), and a desktop or laptop with Internet (65.0%, 269/414 vs 81.1%, 466/575; PR 0.80, 95% CI 0.72-0.89) compared with those with at least some college education. Past 30-day alcohol users reported a higher prevalence of access to a smartphone with Internet (PR 1.12), a desktop or laptop with Internet (PR 1.09), and a smart TV or video game console with Internet (PR 1.31) compared with those who did not report past 30-day alcohol use. Past 30-day marijuana users reported a significantly higher prevalence of access to a smart TV or video game console with Internet access (PR 1.52) compared with nonusers.

Table 3. Correlates of number of social media sites regularly used (defined as using a site multiple times a day, daily, or weekly) among US young adults aged 18-24 years (weighted n=989).

Variable	Mean (SE)[a]	LR[b]	95% CI	aLR[b]	95% CI
Sex					
Male	7.74 (0.13)	Ref		Ref	
Female	7.38 (0.10)	−0.05	−0.09 to 0.00	−0.04	−0.09 to 0.00
Race					
White, non-Hispanic	7.53 (0.10)	Ref		Ref	
Black, non-Hispanic	7.80 (0.26)	0.04	−0.03 to 0.11	0.02	−0.05 to 0.09
Other, non-Hispanic	7.37 (0.32)	−0.02	−0.11 to 0.07	−0.04	−0.13 to 0.05
Hispanic	7.57 (0.18)	0.01	−0.05 to 0.06	0.00	−0.06 to 0.05
Education					
Less than high school	7.99 (0.34)	0.08	0.00-0.17	0.07	−0.01 to 0.16
High school	7.80 (0.16)	0.06	0.01-0.11[c]	0.05	0.00-0.10[c]
Some college	7.35 (0.09)	Ref		Ref	
Financial situation					
Don't meet basic needs	7.83 (0.26)	0.03	−0.04 to 0.10	0.01	−0.06 to 0.08
Just meet basic needs	7.40 (0.17)	−0.03	−0.08 to 0.02	−0.04	−0.09 to 0.02
Meet needs with a little left	7.62 (0.12)	Ref		Ref	
Live comfortably	7.59 (0.17)	0.00	−0.06 to 0.05	−0.01	−0.06 to 0.04
Depression					
No	7.57 (0.09)	Ref		Ref	
Yes	7.41 (0.33)	−0.02	−0.11 to 0.07	−0.02	−0.16 to 0.12
Anxiety					
No	7.57 (0.08)	Ref		Ref	
Yes	7.38 (0.38)	−0.03	−0.13 to 0.08	−0.02	−0.17 to 0.13
Past 30-day alcohol use					
No	7.80 (0.11)	Ref		Ref	
Yes	7.30 (0.12)	−0.07	−0.11 to −0.02[c]	−0.06	−0.10 to −0.01[c]
Past 30-day marijuana use					
No	7.57 (0.09)	Ref		Ref	
Yes	7.48 (0.27)	−0.01	−0.09 to 0.06	0.00	−0.07 to 0.08

[a] SE: standard error.

[b] Crude (LR) and adjusted linear regressions (aLR) with significance at *P*<.05.

[c] Denote statistical significance at *P*<.05.

Table 4. Use of multiple social media sites among regular (defined as using a site multiple times a day, daily, or weekly) users (US young adults aged 18-24 years, weighted n=989). The table presents the % of regular users (ages 18-24 years) of each particular site who use another particular site (eg, 89% of regular users of Tumblr also regularly use Vine).

Other social media use	Use Tumblr, n (%)	Use Vine, n (%)	Use Snapchat, n (%)	Use Instagram, n (%)	Use LinkedIn, n (%)
% of Tumblr users (n=825) who…	-	734 (89)	691 (84)	679 (82)	678 (82)
% of Vine users (n=818) who…	734 (90)	-	697 (85)	687 (84)	666 (81)
% of Snapchat users (n=789) who…	691 (88)	697 (88)	-	684 (87)	647 (82)
% of Instagram users (n=779) who…	679 (87)	687 (88)	684 (88)	-	631 (81)
% of LinkedIn users (n=761) who…	678 (89)	666 (88)	647 (85)	631 (83)	-

Table 5. Correlates of access to specific types of digital technologies among US young adults aged 18-24 years (weighted n=989).

Participant characteristics	A smartphone with Internet access (87%)	A tablet with Internet access (41%)	A desktop or laptop with Internet access (74%)	A smart TV or video game console with Internet access (29%)	A cell phone without Internet access (11%)
	PR[a] (95% CI)	PR (95% CI)	PR (95% CI)	PR (95% CI)	PR (95% CI)
Sex					
Male	Ref[b]	Ref	Ref	Ref	Ref
Female	1.07 (1.01-1.14)[c]	1.11 (0.94-1.31)	0.98 (0.90-1.07)	0.76 (0.62-0.94)[c]	0.67 (0.45-1.00)
Race					
White, NH[d]	Ref	Ref	Ref	Ref	Ref
Black, NH	0.86 (0.75-0.99)[c]	0.74 (0.51-1.07)	0.74 (0.60-0.91)[c]	0.54 (0.32-0.89)[c]	1.57 (0.86-2.85)
Other, NH	0.96 (0.84-1.09)	1.19 (0.88-1.60)	0.96 (0.82-1.12)	0.60 (0.34-1.08)	0.37 (0.12-1.11)
Hispanic	0.97 (0.90-1.03)	1.07 (0.89-1.28)	0.85 (0.77-0.95)[c]	1.21 (0.97-1.51)	1.02 (0.63-1.64)
Education					
Less than high school	0.81 (0.69-0.94)[c]	0.65 (0.43-0.97)[c]	0.65 (0.52-0.83)[c]	1.01 (0.68-1.51)	1.47 (0.73-2.94)
High school	0.91 (0.84-0.98)[c]	0.77 (0.63-0.94)[c]	0.86 (0.78-0.96)[c]	1.04 (0.81-1.33)	1.44 (0.93-2.25)
Some college	Ref	Ref	Ref	Ref	Ref
Financial Situation					
Don’t meet basic expenses	0.74 (0.60-0.92)[c]	0.67 (0.44-1.02)	0.71 (0.54-0.93)[c]	0.42 (0.20-0.88)[c]	1.27 (0.60-2.67)
Just meet basic expenses	0.87 (0.80-0.94)[c]	0.65 (0.52-0.82)[c]	0.90 (0.80-1.00)	0.84 (0.63-1.10)	1.38 (0.81-2.35)
Meet needs with a little left	Ref	Ref	Ref	Ref	Ref
Live comfortably	1.00 (0.94-1.06)	1.12 (0.93-1.34)	1.05 (0.96-1.16)	1.11 (0.87-1.42)	1.22 (0.72-2.07)
Depression					
No	Ref	Ref	Ref	Ref	Ref
Yes	0.97 (0.84-1.11)	0.91 (0.63-1.32)	0.87 (0.71-1.08)	0.90 (0.55-1.47)	1.74 (0.94-3.21)
Anxiety					
No	Ref	Ref	Ref	Ref	Ref
Yes	0.97 (0.85-1.11)	1.04 (0.76-1.41)	0.89 (0.73-1.08)	0.97 (0.64-1.47)	1.41 (0.75-2.63)
Past 30-day alcohol use					
No	Ref	Ref	Ref	Ref	Ref
Yes	1.12 (1.05-1.19)[c]	1.01 (0.86-1.19)	1.09 (1.00-1.19)[c]	1.31 (1.05-1.62)[c]	0.78 (0.51-1.20)

Participant characteristics	A smartphone with Internet access (87%)	A tablet with Internet access (41%)	A desktop or laptop with Internet access (74%)	A smart TV or video game console with Internet access (29%)	A cell phone without Internet access (11%)
	PR[a] (95% CI)	PR (95% CI)	PR (95% CI)	PR (95% CI)	PR (95% CI)
Past 30-day marijuana use					
No	Ref	Ref	Ref	Ref	Ref
Yes	1.02 (0.93-1.12)	0.88 (0.67-1.15)	0.97 (0.85-1.11)	1.52 (1.17-1.98)[c]	0.81 (0.45-1.47)

[a] PR: prevalence ratio.

[b] Ref: Reference category.

[c] Denote statistical significance at $P<.05$.

[d] NH: non-Hispanic.

Access to all Internet-enabled devices varied by subjective financial situation was as follows. Those who reported meeting needs comfortably or meeting needs with a little left reported greater access to a smartphone, a tablet with Internet access, a desktop or laptop with Internet access, and a smart TV or video game console with Internet access than those who "don't meet" or "just meet" basic expenses ($P<.05$). Access to a cell phone without Internet access was low overall (11.5%, 114/989) and did not vary by subjective financial situation. Those who "don't meet basic expenses" reported lower prevalence of access to a smartphone, a desktop or laptop, and a smart TV or video game console with Internet, and those who "just meet basic expenses" reported lower prevalence of access to a smartphone and tablet with Internet access, compared with those who "meet needs with a little left." In the group at greatest socioeconomic disadvantage ("don't meet basic needs") (n=65), 68.4% (n=45) reported access to a smartphone, 30.0% (n=20) to a tablet, 54.7% (n=36) to a desktop or laptop, 13.0% (n=8) to a smart TV or video game console with Internet access, and 12.3% (n=8) to a cell phone without Internet access. Compared with those who "don't meet basic needs," those who "just meet basic needs" (n=297) had similar levels of access to a smartphone (79.8%, n=237), tablet (29.4%, n=87), desktop or laptop (69.2%, n=205), smart TV or video game console (25.8%, n=76), and cell phone without Internet (13.4%, n=40).

Discussion

Principal Findings

According to Pew Research, 99% of US young adults aged 18-29 years in 2016 used the Internet [32] and 90% of these young adults used a social networking site as of 2015 [4]. As of September 2014, 87% of Internet-using young adults reported using Facebook, 53% reported using Instagram, 37% reported using Twitter, and 34% reported using Pinterest [33]. This Pew Research survey did not report on YouTube use. Our study shows that social media use has continued to increase and regular social media use was nearly ubiquitous in 2016 in a national sample of young adults aged 18-24 years. Additionally, young adults surveyed in 2016 reported regular use of different sites than in 2014 (ie, Tumblr, Vine, Snapchat, Instagram, and LinkedIn). The fact that most sites allow for multimedia content reflect the rising importance of visual content in Internet communication [34], particularly among young people. Lower use of YouTube in 2016 suggests that young adults may be accessing YouTube video content secondarily via other social media sites (eg, links in Twitter) and attributing the content to the primary site used. Participants used a high mean number of social media sites regularly and there was substantial overlap in use of social media sites.

In addition to providing updated estimates of social media use, this study found that while there were no consistent correlates of use of particular social media sites, there were a few relationships that deserve further exploration. First, there was a higher prevalence of LinkedIn use by black young adults and those with less than a college education. LinkedIn is an employment-oriented social media site that offers professional networking opportunities; higher LinkedIn use among young adults with lower education may reflect job-seeking among young people not enrolled in college. Given the professional focus of this site, it is unlikely to yield much information about the health behaviors of young people. Two other sites may provide more insight into the depiction of health behaviors on the Internet: Snapchat and Tumblr. Our study found there was a higher prevalence of Snapchat use among young adults reporting the greatest socioeconomic disadvantage. There was also a lower prevalence of Tumblr use among females and those who "just meet basic needs." Tumblr, a social blogging platform, allows users to share and discuss multimedia content (text, photos, quotes, links, music, videos) and customize their blog using embedded tools. Approximately 32% of Tumblr bloggers are 18-24 years of age, and 67% are under age 35 [35]. Tumblr users can remain anonymous, which facilitates a relatively high degree of disclosure and sharing, particularly around sensitive topics [36-38]. Snapchat enables users to share photos and short videos with closed networks of friends or to broader unknown networks that disappear in 24 hours or less. It features tools for customizing photos/videos with filters, stickers, and drawings. High use of these sites reflects the potential importance of anonymity, creativity, and ephemerality to social media users. Emerging research examines how people use these sites to portray their engagement in health-risk behaviors (eg, sex, alcohol use, tobacco use) [39-48]. In line with social cognitive theory [9], we would expect social observation and interaction via social media to influence beliefs about health-risk behaviors or the behaviors themselves. To date, limited research exists on young adult exposure to risk behaviors via social media and the impact of such exposure on subsequent risk behavior [49-54].

Mobile devices are a primary channel for social media: Pew data indicate that in 2015, 85% of young adults aged 18-29 years had a smartphone and 91% of these individuals use a social networking site on the phone [55]. Our findings are consistent, showing that in 2016, 86.9% (860/989) of young adults aged 18-24 years had a smartphone. Novel findings from our study, however, highlight disparities in access to digital technologies among US young adults by race/ethnicity, education, and subjective financial status. Black young adults, those with less than a college education, and those who “don’t meet” or “just meet” basic expenses were significantly less likely to have access to a smartphone with Internet access compared with whites, those with a college education, and those who “meet needs with a little left,” respectively. A similar pattern emerged regarding access to a desktop or laptop with Internet access. Interestingly, this did not result in a greater proportion of these respondents reporting access to a cell phone without Internet. Smartphone access remained relatively high, despite these differences, with the lowest prevalence seen in those reporting that they “don’t meet basic needs” (68.4%). Disparities in ongoing social media and Internet access may have important implications for Web-based health interventions seeking to target groups that may be at highest risk.

Limitations

Limitations of this study include the self-reported nature of social media use and the social media sites identified in the survey. We may have missed other popular social media sites and data on some of the sites used may be erroneous; for example, the high prevalence of use of Google+ likely reflects use of Google’s search engine given the failure of the Google+ social networking site [56]. Similarly, our question regarding access to and use of digital technology did not ask about device ownership as in the Pew studies. Deviations from the Pew Center data may reflect differences in the timing of the surveys, items used, sample-specific differences in survey measurement, including differences in sample sizes, and different sampling and weighting strategies. The study sample’s completion rates and cumulative response rates are similar to that of other health studies that have relied on KnowledgePanel [24-26,57]. The internal validity of our results is not compromised by the panel’s cumulative response rate, and other work suggests that surveys with a low response rate can still be representative of the sample population, even though the risk of nonresponse bias is higher [58,59]. Studies assessing nonresponse to panel recruitment in KnowledgePanel have found little indication of nonresponse bias on core demographic and socioeconomic variables [60,61], and previous estimates from this cohort for key outcomes of interest, such as ever and current cigarette use, are consistent with national survey data [19]. Although missing data were relatively low for our covariates of interest (3.3%), our analytic approach that used list-wise deletion may have introduced bias to our results if missingness is not random.

Conclusions

There are several mechanisms through which social media interventions can influence health behavior; however, few studies to date have used social media to facilitate health behavior change in young adults [62-68]. Intervention and user-generated content in social media can be a powerful source of influence through peer modeling [69]. Additionally, content delivered via social media may correct misperceptions, offer resources to assist behavior change, and provide opportunities to recruit peer support for behavior change via one’s existing social networks. The findings from the current study may guide the development and implementation of future health interventions for young adults delivered via the Internet or social media sites. Our study highlights that young people are using multiple social media sites regularly and that these sites may provide an accessible venue for delivering health messaging. These messages will need to be tailored for the top social media sites used, including creative use of images, videos, hashtags, and other content to be relevant to the target audience. Intervention exposure in the target population may be maximized through coordinated dissemination of health messages across multiple social media sites. Lower access to Internet-enabled mobile devices among black young adults and socioeconomically disadvantaged young adults does not discount the utility of Web-based health behavior interventions in this group, but highlights that other channels may be needed to complement a Web-based approach in these subgroups. This research also highlights that the top social media sites change rapidly and any social media intervention approach for young adults must be flexible and nimble enough to adapt to new sites, new patterns of use, and new ways of delivering content via social media.

Acknowledgments

This study was funded by Truth Initiative. Dr Villanti was supported in part by Truth Initiative and the Centers of Biomedical Research Excellence P20GM103644 award from the National Institute on General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Erikson E. Identity: youth, and crisis. New York: W.W. Norton; 1968.
2. Arnett JJ. Emerging adulthood. a theory of development from the late teens through the twenties. Am Psychol 2000 May;55(5):469-480. [Medline: 10842426]
3. Weber S, Mitchell C. Imaging, keyboarding, and posting identities: young people and new media technologies. In: Buckingham D, editor. Youth, Identity, and Digital Media. Cambridge, MA: MIT Press; 2008:25-48.
4. Perrin A. Pew Research Center. 2015. Social media usage: 2005-2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [accessed 2016-12-06] [WebCite Cache ID 6mYnUrF5N]
5. DeAndrea D, Ellison N, LaRose R, Steinfield C, Fiore A. Serious social media: on the use of social media for improving students' adjustment to college. Internet High Educ 2012;15(1):15-23.
6. Mazur E, Kozarian L. Self-presentation and interaction in blogs of adolescents and young emerging adults. J Adolesc Res 2009 Nov 11;25(1):124-144. [doi: 10.1177/0743558409350498]
7. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
8. Bandura A. Social learning theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
9. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001;3(3):265-299.
10. Buckingham D. Introducing identity. In: Buckingham D, editor. Youth, Identity, and Digital Media. Cambridge, MA: MIT Press; 2008:1-24.
11. This American Life. 2015. 573: status update URL: https://www.thisamericanlife.org/radio-archives/episode/573/status-update [accessed 2017-03-28] [WebCite Cache ID 6pI4omFhW]
12. Northwestern CTD. The self in selfie: identity in the age of social media URL: https://www.ctd.northwestern.edu/blog/self-selfie-identity-age-social-media [accessed 2017-03-28] [WebCite Cache ID 6pI4zhfYk]
13. Spies Shapiro LA, Margolin G. Growing up wired: social networking sites and adolescent psychosocial development. Clin Child Fam Psychol Rev 2014 Mar;17(1):1-18 [FREE Full text] [doi: 10.1007/s10567-013-0135-1] [Medline: 23645343]
14. Shensa A, Sidani JE, Lin LY, Bowman ND, Primack BA. Social media use and perceived emotional support among US young adults. J Community Health 2016 Jun;41(3):541-549. [doi: 10.1007/s10900-015-0128-8] [Medline: 26613936]
15. Primack BA, Shensa A, Sidani JE, Whaite EO, Lin LY, Rosen D, et al. Social media use and perceived social isolation among young adults in the U.S. Am J Prev Med 2017 Feb 24:- Epub ahead of print. [doi: 10.1016/j.amepre.2017.01.010] [Medline: 28279545]
16. Primack B, Shensa A, Escobar-Viera C, Barrett E, Sidani J, Colditz J, et al. Use of multiple social media platforms and symptoms of depression and anxiety: a nationally-representative study among U.S. young adults. Comput Human Behav 2017 Apr;69:1-9. [doi: 10.1016/j.chb.2016.11.013]
17. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among U.S. young adults. Depress Anxiety 2016 Apr;33(4):323-331. [doi: 10.1002/da.22466] [Medline: 26783723]
18. Pew Research Center. 2007. How young people view their lives, futures and politics: a portrait of “generation next” URL: http://www.people-press.org/files/legacy-pdf/300.pdf [accessed 2017-05-11] [WebCite Cache ID 6qNGDYj8a]
19. Rath JM, Villanti AC, Abrams DB, Vallone DM. Patterns of tobacco use and dual use in US young adults: the missing link between youth prevention and adult cessation. J Environ Public Health 2012;2012:679134 [FREE Full text] [doi: 10.1155/2012/679134] [Medline: 22666279]
20. Knowledgenetworks. GfK KnowledgePanel® URL: http://www.knowledgenetworks.com/knpanel/index.html [accessed 2017-04-27] [WebCite Cache ID 6q2p5aKY0]
21. Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the internet: comparing sample representativeness and response quality. Public Opin Q 2009 Dec 01;73(4):641-678. [doi: 10.1093/poq/nfp075]
22. Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, et al. Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. Public Opin Q 2011 Oct 05;75(4):709-747. [doi: 10.1093/poq/nfr020]
23. Rhodes DJ, Radecki BC, Ziegenfuss JY, Jenkins SM, Vachon CM. Awareness of breast density and its impact on breast cancer detection and risk. J Clin Oncol 2015 Apr 01;33(10):1143-1150 [FREE Full text] [doi: 10.1200/JCO.2014.57.0325] [Medline: 25732156]
24. Grande D, Mitra N, Shah A, Wan F, Asch DA. Public preferences about secondary uses of electronic health information. JAMA Intern Med 2013 Oct 28;173(19):1798-1806 [FREE Full text] [doi: 10.1001/jamainternmed.2013.9166] [Medline: 23958803]
25. Kumar S, Quinn SC, Kim KH, Daniel LH, Freimuth VS. The impact of workplace policies and other social factors on self-reported influenza-like illness incidence during the 2009 H1N1 pandemic. Am J Public Health 2012 Jan;102(1):134-140. [doi: 10.2105/AJPH.2011.300307] [Medline: 22095353]
26. Fowler FJ, Gerstein BS, Barry MJ. How patient centered are medical decisions?: results of a national survey. JAMA Intern Med 2013 Jul 08;173(13):1215-1221. [doi: 10.1001/jamainternmed.2013.6172] [Medline: 23712194]
27. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an informed consumer of health information and assessment of electronic health literacy in a national sample of internet users: validity and reliability of the e-HLS instrument. J Med Internet Res 2016 Jul 11;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]

28. American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8th edition. Oakbrook Terrace, IL: AAPOR; 2015.
29. Williams VF, Smith AA, Villanti AC, Rath JM, Hair EC, Cantrell J, et al. Validity of a subjective financial situation measure to assess socioeconomic status in us young adults. J Public Health Manag Pract 2016 Oct 28:- Epub ahead of print. [doi: 10.1097/PHH.0000000000000468] [Medline: 27798530]
30. Kroenke K, Spitzer RL, Williams JB. The patient health questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
31. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [Medline: 17339617]
32. Anderson M, Perrin A. Pew Research Center. 2016. 13% of Americans don’t use the internet. Who are they? URL: http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internet-who-are-they/ [accessed 2016-12-06] [WebCite Cache ID 6mYoSZcmc]
33. Duggan M, Ellison NB, Lampe C, Lenhart A, Madden M. Pew Research Center. 2015. Social media update 2014: frequency of social media use URL: http://www.pewinternet.org/2015/01/09/frequency-of-social-media-use-2 [accessed 2016-12-06] [WebCite Cache ID 6mYodm44I]
34. Walter E. Fastcompany. 2012. The Rise of Visual Social Media URL: https://www.fastcompany.com/3000794/rise-visual-social-media [accessed 2016-12-06] [WebCite Cache ID 6mYpC0XVJ]
35. Dougherty J. Leaderswest. 2013. Infographic: average time per visit on tumblr is 34 minutes URL: http://leaderswest.com/2013/06/13/infographic-average-time-per-visit-on-tumblr-is-34-minutes/ [WebCite Cache ID 6mYpnE5AI]
36. Tumblr Staff. Tumblr. A new policy against self-harm blogs URL: https://staff.tumblr.com/post/18132624829/self-harm-blogs [accessed 2016-12-06] [WebCite Cache ID 6mYq1EJKr]
37. OK2Talk. 2015. OK2Talk URL: http://ok2talk.org/ [accessed 2016-12-06] [WebCite Cache ID 6mYqIL4Bd]
38. Tumblr. 2014. Planned parenthood URL: http://plannedparenthood.tumblr.com/ [accessed 2016-12-06] [WebCite Cache ID 6mYqSQiwU]
39. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? electronic cigarettes and smoking cessation on Twitter. Int J Public Health 2016 Mar;61(2):249-256. [doi: 10.1007/s00038-016-0791-2] [Medline: 26841895]
40. Seidenberg AB, Pagoto SL, Vickey TA, Linos E, Wehner MR, Costa RD, et al. Tanning bed burns reported on Twitter: over 15,000 in 2013. Transl Behav Med 2016 Jun;6(2):271-276 [FREE Full text] [doi: 10.1007/s13142-016-0388-6] [Medline: 27356997]
41. Moreno MA, Arseniev-Koehler A, Litt D, Christakis D. Evaluating college students' displayed alcohol references on facebook and twitter. J Adolesc Health 2016 May;58(5):527-532. [doi: 10.1016/j.jadohealth.2016.01.005] [Medline: 26995291]
42. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. “Those edibles hit hard”: exploration of twitter data on cannabis edibles in the U.S. Drug Alcohol Depend 2016 Jul 1;164:64-70. [doi: 10.1016/j.drugalcdep.2016.04.029] [Medline: 27185160]
43. Huang J, Kornfield R, Emery SL. 100 Million views of electronic cigarette youtube videos and counting: quantification, content evaluation, and engagement levels of videos. J Med Internet Res 2016 Mar 18;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
44. Hilton C. Unveiling self-harm behaviour: what can social media site twitter tell us about self-harm? a qualitative exploration. J Clin Nurs 2016 Sep 07;26(11-12):1690-1704. [doi: 10.1111/jocn.13575] [Medline: 27604049]
45. Grant A, O'Mahoney H. Portrayal of waterpipe (shisha, hookah, nargile) smoking on twitter: a qualitative exploration. Public Health 2016 Nov;140:128-135. [doi: 10.1016/j.puhe.2016.07.007] [Medline: 27520707]
46. Chu K, Allem J, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Sep 22:- Epub ahead of print. [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
47. Primack BA, Carroll MV, Shensa A, Davis W, Levine MD. Sex differences in hookah-related images posted on tumblr: a content analysis. J Health Commun 2016;21(3):366-375. [doi: 10.1080/10810730.2015.1095814] [Medline: 26890733]
48. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Connolly S, Rosas C, Bharadwaj M, et al. An analysis of depression, self-harm, and suicidal ideation content on tumblr. Crisis 2016 Jul 22:1-9. [doi: 10.1027/0227-5910/a000409] [Medline: 27445014]
49. Boyle SC, LaBrie JW, Froidevaux NM, Witkovic YD. Different digital paths to the keg? how exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students. Addict Behav 2016 Jun;57:21-29. [doi: 10.1016/j.addbeh.2016.01.011] [Medline: 26835604]
50. Sidani JE, Shensa A, Hoffman B, Hanmer J, Primack BA. The association between social media use and eating concerns among US young adults. J Acad Nutr Diet 2016 Sep;116(9):1465-1472. [doi: 10.1016/j.jand.2016.03.021] [Medline: 27161027]
51. Cabrera-Nguyen EP, Cavazos-Rehg P, Krauss M, Bierut LJ, Moreno MA. Young adults' exposure to alcohol- and marijuana-related content on twitter. J Stud Alcohol Drugs 2016 Mar;77(2):349-353 [FREE Full text] [Medline: 26997194]
52. Moreno MA, Whitehill JM. Influence of social media on alcohol use in adolescents and young adults. Alcohol Res 2014;36(1):91-100 [FREE Full text] [Medline: 26259003]

53. Moreno MA, Cox ED, Young HN, Haaland W. Underage college students' alcohol displays on facebook and real-time alcohol behaviors. J Adolesc Health 2015 Jun;56(6):646-651 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.02.020] [Medline: 26003580]
54. Miller J, Prichard I, Hutchinson A, Wilson C. The relationship between exposure to alcohol-related content on facebook and predictors of alcohol consumption among female emerging adults. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):735-741. [doi: 10.1089/cyber.2014.0337] [Medline: 25489875]
55. Smith A. Pew Research Center. Washington, DC; 2015. Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [WebCite Cache ID 6mYqg5y82]
56. Fiegerman S. Mashable. 2015. Inside the failure of Google+, a very expensive attempt to unseat facebook URL: http://mashable.com/2015/08/02/google-plus-history/ [accessed 2016-12-06] [WebCite Cache ID 6mYqohwA1]
57. Kelly B, Squiers L, Bann C, Stine A, Hansen H, Lynch M. Perceptions and plans for prevention of Ebola: results from a national survey. BMC Public Health 2015;15(1):1136. [doi: 10.1186/s12889-015-2441-7] [Medline: 26572610]
58. Brick JM. The future of survey sampling. Public Opin Q 2011 Dec 15;75(5):872-888. [doi: 10.1093/poq/nfr045]
59. Halbesleben JR, Whitman MV. Evaluating survey quality in health services research: a decision framework for assessing nonresponse bias. Health Serv Res 2013 Jun;48(3):913-930 [FREE Full text] [doi: 10.1111/1475-6773.12002] [Medline: 23046097]
60. Heeren T, Edwards EM, Dennis JM, Rodkin S, Hingson RW, Rosenbloom DL. A comparison of results from an alcohol survey of a prerecruited Internet panel and the national epidemiologic survey on alcohol and related conditions. Alcohol Clin Exp Res 2008 Feb;32(2):222-229. [doi: 10.1111/j.1530-0277.2007.00571.x] [Medline: 18162076]
61. Garrett J, Dennis J, DiSogra C. Non-response bias: recent findings from address-based panel recruitment. 2010 Presented at: Annual Conference of the American Association for Public Opinion Research; May 13-16, 2010; Chicago, IL.
62. Struik LL, Baskerville NB. The role of facebook in crush the crave, a mobile- and social media-based smoking cessation intervention: qualitative framework analysis of posts. J Med Internet Res 2014;16(7):e170 [FREE Full text] [doi: 10.2196/jmir.3189] [Medline: 25016998]
63. Ramo DE, Liu H, Prochaska JJ. A mixed-methods study of young adults' receptivity to using facebook for smoking cessation: if you build it, will they come? Am J Health Promot 2015 Apr;29(4):e126-e135. [doi: 10.4278/ajhp.130326-QUAL-128] [Medline: 24575728]
64. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31. [doi: 10.1002/oby.20232] [Medline: 23505165]
65. Patrick K, Marshall SJ, Davila EP, Kolodziejczyk JK, Fowler JH, Calfas KJ, et al. Design and implementation of a randomized controlled social and mobile weight loss trial for young adults (project SMART). Contemp Clin Trials 2014 Jan;37(1):10-18 [FREE Full text] [doi: 10.1016/j.cct.2013.11.001] [Medline: 24215774]
66. Haines-Saah RJ, Kelly MT, Oliffe JL, Bottorff JL. Picture Me Smokefree: a qualitative study using social media and digital photography to engage young adults in tobacco reduction and cessation. J Med Internet Res 2015;17(1):e27 [FREE Full text] [doi: 10.2196/jmir.4061] [Medline: 25624064]
67. 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]
68. Ramo DE, Thrul J, Chavez K, Delucchi KL, Prochaska JJ. Feasibility and quit rates of the tobacco status project: a facebook smoking cessation intervention for young adults. J Med Internet Res 2015;17(12):e291 [FREE Full text] [doi: 10.2196/jmir.5209] [Medline: 26721211]
69. Huang GC, Unger JB, Soto D, Fujimoto K, Pentz MA, Jordan-Marsh M, et al. Peer influences: the impact of online and offline friendship networks on adolescent smoking and alcohol use. J Adolesc Health 2014 May;54(5):508-514. [doi: 10.1016/j.jadohealth.2013.07.001] [Medline: 24012065]

Abbreviations

GAD: generalized anxiety disorder
GfK: Gesellschaft für Konsumforschung (Society for Consumer Research)
PHQ: patient health questionnaire
SE: standard error

Edited by G Eysenbach; submitted 11.01.17; peer-reviewed by E Warner, F Modave, T van Mierlo, JP Allem; comments to author 26.02.17; revised version received 28.03.17; accepted 28.03.17; published 07.06.17.

Please cite as:
Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM
Social Media Use and Access to Digital Technology in US Young Adults in 2016
J Med Internet Res 2017;19(6):e196
URL: http://www.jmir.org/2017/6/e196/
doi:10.2196/jmir.7303
PMID:28592394

Original Paper

What Predicts Online Health Information-Seeking Behavior Among Egyptian Adults? A Cross-Sectional Study

Mayada Ghweeba[1,2], MPH; Antje Lindenmeyer[1], PhD; Sobhi Shishi[2], MD; Mostafa Abbas[2], MD; Amani Waheed[2], MD; Shaymaa Amer[2], MD

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom

[2]Community Medicine Department, Faculty of Medicine, Suez Canal University, Ismailia, Egypt

Corresponding Author:
Antje Lindenmeyer, PhD
Institute of Applied Health Research
University of Birmingham
Edgbaston
Birmingham, B15 2TT
United Kingdom
Phone: 44 0 121 4145390
Fax: 44 0 121 4145390
Email: a.lindenmeyer@bham.ac.uk

Abstract

Background: Over the last decade, the Internet has become an important source of health-related information for a wide range of users worldwide. Yet, little is known about the personal characteristics of Egyptian Internet users who search for online health information (OHI).

Objective: The aim of the study was to identify the personal characteristics of Egyptian OHI seekers and to determine any associations between their personal characteristics and their health information-seeking behavior.

Methods: This cross-sectional questionnaire study was conducted from June to October 2015. A Web-based questionnaire was sent to Egyptian users aged 18 years and older (N=1400) of a popular Arabic-language health information website. The questionnaire included (1) demographic characteristics; (2) self-reported general health status; and (3) OHI-seeking behavior that included frequency of use, different topics sought, and self-reported impact of obtained OHI on health behaviors. Data were analyzed using descriptive statistics and multiple regression analysis.

Results: A total of 490 participants completed the electronic questionnaire with a response rate equivalent to 35.0% (490/1400). Regarding personal characteristics, 57.1% (280/490) of participants were females, 63.4% (311/490) had a university level qualification, and 37.1% (182/490) had a chronic health problem. The most commonly sought OHI by the participants was nutrition-related. Results of the multiple regression analysis showed that 31.0% of the variance in frequency of seeking OHI among Egyptian adults can be predicted by personal characteristics. Participants who sought OHI more frequently were likely to be female, of younger age, had higher education levels, and good self-reported general health.

Conclusions: Our results provide insights into personal characteristics and OHI-seeking behaviors of Egyptian OHI users. This will contribute to better recognize their needs, highlight ways to increase the availability of appropriate OHI, and may lead to the provision of tools allowing Egyptian OHI users to navigate to the highest-quality health information.

(J Med Internet Res 2017;19(6):e216) doi:10.2196/jmir.6855

KEYWORDS

Internet; information-seeking behavior; computer literacy; surveys and questionnaires; Egypt

Introduction

The use of the Internet as a source of health information has increased considerably over the last decade worldwide [1]. Many Internet users have switched to the Internet to obtain information about health and health care [2,3]. The Internet World Stats report indicates that the number of Internet users in Egypt has increased from 29.8 million in June 2015 (32.6% of the population) to 33.3 million in June 2016 (37.0% of the population) [4]. Most of this growth in Internet use has been around social activities (social media and email), educational activities, and playing games. However, the number of Egyptian Internet users who search for health-related information has increased from 29.7% in 2012 to 32.4% in 2013 from the total Egyptian Internet users as reported by the Egyptian Ministry of Communication and Information Technology (MCIT) [5]. Whereas most health-related websites funded by the Egyptian Government are targeting professional users, there is now a range of private Arabic language websites providing general health information to users; however, their quality tends to be variable [6].

Online health information (OHI) seekers are defined as “Internet users who search online for information on health topics, whether they are acting as consumers, caregivers or patients aiming to take more control of their health” [7]. The variability of personal characteristics such as age, gender, race or ethnicity, and income levels can be a predictor for using the Internet as a source of health information [8-10].

Using the Internet to search for health information has many advantages: easy access, availability of a wide range of illustrated and audiovisual health resources, as well as providing an opportunity to ask experts’ opinion [11]. Furthermore, its anonymity enables users to express their health problems and share experiences privately with other people in a similar situation [12]. The Internet can contribute to the improvement of social interaction, better coping with life situations, more knowledge about particular diseases, emotional relief, and improvement in clinical outcomes for health information seekers [13]. Despite the advantages of using OHI, the Internet may also present a great risk for consumers when they are not aware of the quality standards of the obtained health information [14,15]. Studies have reported that health information available on the Internet is sometimes scientifically incomplete or inadequate [16,17]. When the health information obtained from the Internet is of questionable quality, it may negatively influence the doctor-patient relationship [18,19]. The wide variability of resources on the Internet and the commercial interest of some providers can cause confusion to the user and make it difficult to locate accurate and reliable health information [17]. Additionally, there are some barriers to access the Internet such as lack of computer skills, lower levels of education, and geographic location [20,21].

Public awareness of the usefulness of OHI has risen in low- and middle-income countries as well as in the Arab world [22,23]. Although the number of Egyptian OHI seekers has increased, knowledge of personal characteristics of OHI seekers and their motives behind searching for OHI is still limited [5]. This study aimed to identify the association between personal characteristics of Egyptian OHI seekers and their health information-seeking behavior. This may help to improve ways of promoting efficient and appropriate OHI for users and harnessing the benefits of the Internet as a source of health information.

Methods

Data for this cross-sectional questionnaire study was collected from June to October 2015. An electronic questionnaire was sent to registered users meeting the inclusion criteria (currently living in Egypt and aged 18 years or older; N=1400) of a widely used Arabic language health information website (6abibak.com) by the website administrator. Two reminders were sent to the selected participants in order to improve the response rate. The invitation included information about the aim of the study and reassurance about the anonymity of collected data. After completing the consent form, participants progressed to complete the Web-based questionnaire. The questionnaire was developed based on previous studies[16,17,24,25]. The questionnaire was divided into the following three sections:

1. Sociodemographic characteristics: this included age, gender, levels of education, employment status (employed or not employed), marital status, and whether participants were covered by a health insurance scheme or not
2. Health status: participants were asked whether they had any chronic health problems and to rate their self-reported general health status on a scale ranging from poor to excellent
3. Health information-seeking behavior: participants were asked about their main source of health information, the frequency of using OHI (hours/ week), rating the quality of the OHI obtained, health topics they had searched for on the Internet, and whether they had made any behavioral change resulting from the obtained OHI.

The study was approved by the Ethics and Research Committee of The Faculty of Medicine, Suez Canal University, Egypt.

Statistical Analysis

Statistical analysis was performed using SPSS statistical software (version 22; SPSS Inc). Descriptive statistics (eg, means and standard deviations) were used in the initial data analysis. Chi-square test was used to identify associations between different personal characteristics of the study participants and the use of Internet as the main source of health information. The level of statistical significance was set at *P* value<.05.

Standard multiple regression analysis was carried out to assess the contribution of personal characteristics (the independent variables) to the frequency of using OHI among the participants as a dependent variable. Independent variables were presented as a sequence of nested models. Model 1 included age, gender, and educational level. In model 2, we added self-reported general health and having a chronic health problem. In model 3, we added using the Internet as the main source of health information. In model 4 we added any self-reported behavioral changes due to the obtained OHI.

Table 1. Sociodemographic and health-related characteristics of study participants (N=490).

Sociodemographic characteristics		n (%)
Age (years)		
	18-20	36 (7.3)
	20-35	242 (49.4)
	35-50	181 (36.9)
	≥50	31 (6.4)
Gender		
	Male	210 (42.9)
	Female	280 (57.1)
Highest level of education		
	Primary education	69 (14.2)
	Secondary education	110 (22.4)
	University degree (undergraduate or postgraduate)	311(63.4)
Social status		
	Never married before	187 (38.2)
	Married	251 (51.2)
	Divorced or widow or widower	52 (10.6)
Employment		
	Employed	302 (61.6)
	Unemployed[a]	96 (19.6)
	Student	92 (18.8)
Self-reported general health		
	Excellent	95 (19.3)
	Good	282 (57.6)
	Poor	113 (23.1)
Having a chronic health problem	Yes[b]	182 (37.1)
Having health insurance cover	Yes	330 (67.3)

[a] Unemployed also included housewife or retired.

[b] Having one or more chronic health problems.

Results

A total of 490 participants completed the Web-based questionnaire with a response rate of 35.00% (490/1400).

Sociodemographic Characteristics

The main sociodemographic characteristics of study participants are summarized in Table 1. The mean age of the participants was 36.8 years. The number of female participants (57.1%, 280/490) was marginally higher than the male participants (42.9%, 210/490); 63.4% (311/490) were university graduates, and 61.6% (302/490) were employed.

Self-Reported Health Status

About 76.9% (377/490) participants rated their general health as good or excellent; meanwhile, only 23.1% (113/490) rated it as poor. Of all the participants, 67.3% (330/490) reported that they had health insurance coverage (either private or government-based coverage). Additionally, 37.1% (182/490) of the participants reported one or more chronic health problems.

Health Information-Seeking Behavior

The mean time spent searching for OHI was 1.58 (SD 2.13) hours/ week (Table 2). Of all the respondents, more than half (55.4%, 271/490) considered the Internet as a main source for health information, followed by consulting physicians (30.7%, 151/490), whereas TV and radio were the least commonly used sources for health information (3.9%, 19/490). The main reasons for preferring OHI were easy access (47.2%, 231/490) and getting more information about specific health problems (30.8%, 151/490),whereas only 11.6% (57/490) participants preferred OHI because of its ability to preserve the user’s anonymity.

Table 2. Distribution of the online health information (OHI)-seeking behavior of study participants (N=490).

OHI[a]-seeking behavior of the participants		n (%)
Main source of health-related information[b]		
	Internet	271 (55.4)
	Physicians	151 (30.7)
	Family members or friends or colleagues	49 (10.0)
	Radio or TV	19 (3.9)
	Others	79 (16.1)
Average time spent on health information Web pages (hour/week)		
	<2	226 (46.1)
	2-5	150 (30.7)
	>5	114 (23.2)
Main reason behind preferring OHI		
	Easy access	231(47.2)
	Improving understanding of a specific health problem	151 (30.8)
	Anonymity	57 (11.6)
	Recommended by family, friend, or physician	51(10.4)
Web pages visited for health-related information		
	Scientific societies	111 (22.6)
	Websites recommended by search engines such as Google, Yahoo, etc.	163 (33.3)
	Websites related to Ministry of Health or universities	81 (16.5)
	No special preference	135 (27.6)
Participant rating of the OHI quality		
	Excellent	172 (35.2)
	Good	235 (47.9)
	Fair	83 (16.9)
	Poor	0
Looking for OHI		
	For themselves	262 (53.5)
	For someone else	150 (30.6)
	Both	78 (15.9)
Search topics		
	New health issue	126 (25.7)
	Longstanding health issue	152 (31.0)
	Both	212 (43.3)
Preferred search location for OHI		
	Home	309 (63.1)
	Work	94 (19. 3)
	Others[c]	87 (17.6)

[a] OHI: online health information.

[b] Multiple responses allowed.

[c] Others include public places and Internet cafe.

Table 3. Cross-tabulations between personal characteristics of the study participants and using the Internet as a main source of health information(N=490).

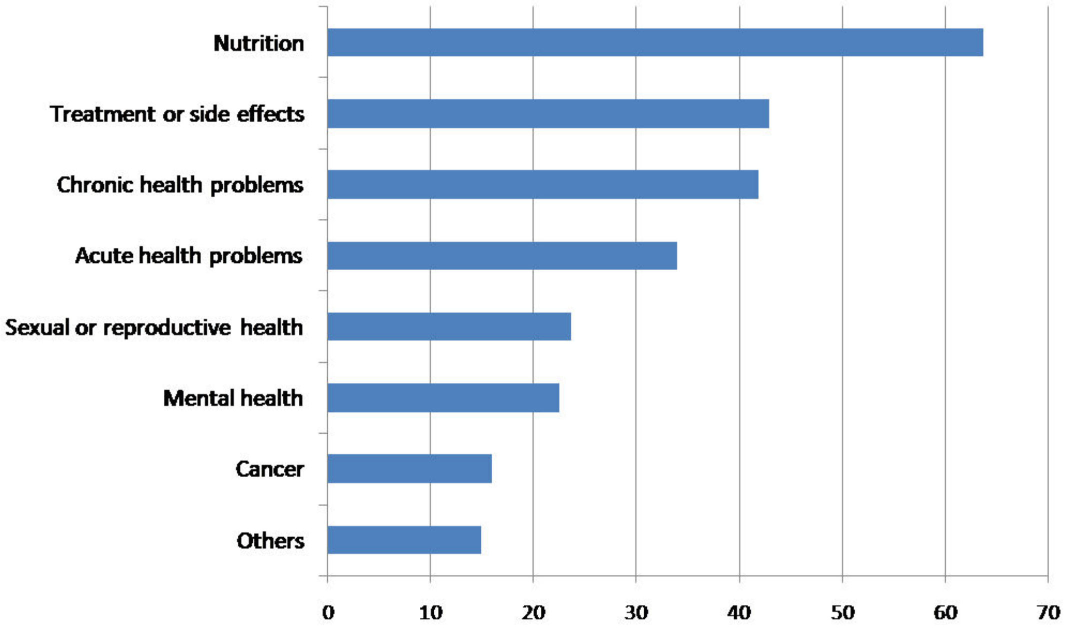
Personal characteristics		n	Using the Internet as a main source of health information n (%)	*P* value[a]
Age (years)				
	<35	278	208 (74.8)	.001
	≥35	212	129 (60.8)	
Gender				
	Male	210	103 (49.1)	.01
	Female	280	168 (60.0)	
Having chronic diseases[b]	Yes	182	121 (66.5)	.001
Reporting behavioral changes after obtaining OHI[c]	Yes	283	173 (61.1)	.002

[a] Result is significant at *P*<.05.

[b] Having one or more chronic health problems.

[c] OHI: online health information.

Figure 1. Common health topics sought on the Internet by the study participants (%).



The majority of participants (83.1%, 407/490) rated the quality of OHI as good or excellent, and 16.9% (83/490) rated it as fair, whereas none rated it as poor. About 53.5% (262/490) of participants reported looking for health information for their own health issues, whereas 30.6% (150/490) were looking for health information for someone else, for example, family member or a friend (Table 2).

Table 3 compares personal characteristics of the participants and their use of the Internet as the main source of information, which was significantly higher among female participants (χ^2_1=5.8, *P*=.01), younger age group participants (χ^2_1=10.4, *P*=.001), and participants who reported having chronic health conditions (χ^2_1=14.6, *P*=.001; Table 3). Using the Internet as the main source of health information was significantly higher among participants who reported behavioral changes from the obtained (χ^2_1=9.1, *P*=.002, Table 3).

Participants reported seeking OHI for one or more of a broad range of health topics. About 63.8% (312/490) of the participants indicated that they sought information about nutrition and 41.9% (205/490) of the participants reported using the Internet to find information about chronic health conditions, whereas 23.7% (116/490) searched for sexual health information and 22.6% (110/490) participants searched for mental health issues (Figure 1).

Table 4. Multiple regression analysis with frequency of using online health information (OHI) as dependent variable and personal characteristics of Internet users as independent variables (*P*<.05. statistically significant).

Personal characteristics	Model 1		Model 2		Model 3		Model 4	
	Beta[a]	SE[b]	Beta	SE	Beta	SE	Beta	SE
Age	−0.19	0.04	−0.19	0.04	−0.16	0.03	−0.16	0.03
Female (vs male)	.22	0.04	.21	0.04	.22	0.03	.20	0.04
Primary educated (vs college graduate)	.62	0.08	.09	0.06	.09	0.10	.09	0.08
Secondary educated (vs college graduates)	.12	0.09	.11	0.08	−0.12	0.09	.13	0.09
Poor general health (vs good)			.07	0.12	.07	0.10	.07	0.12
Fair general health (vs good)			−0.01	0.26	−0.05	0.26	.02	0.26
With chronic disease (vs without).			.06	0.02	−0.02	0.02	.03	0.08
Health insurance (vs noninsured)			.07	0.18	.06	0.18	.07	0.18
Internet as the main source of health information (vs not)					.12	0.19	.07	0.23
Behavior changes (vs no changes)							. 95	0.19
Constant	-	3.13	-	5.71	-	5.80	-	5.49
F value	18.02		12.80		12.71		11.94	
R^2	0.23		0.28		0.29		0.31	

[a] Beta: unstandardized.

[b] SE: standard error.

Table 4 shows the results of the multiple regression analysis with personal characteristics as independent variables and the frequency of using OHI as a dependent variable. In model 1, the regression coefficient of the independent variables age (beta=−0.16), gender (beta=.02), and level of education (beta=.13) were statistically significant. The model showed that younger participants spent more time on searching OHI than older participants. Model 1 can predict 23.2% of the variance in using OHI among the participants **.**

Adding the self-reported general health status in regression model 2, a statistically significant increase in R^2(from 0.23 to 0.28, *P*<.05) occurred. The direction and the significance effect of age, gender, and education remain the same as in model 1. Adding the use of the Internet as a main source of health information in regression model 3, a further statistically significant increase occurred in the variance in frequency of using OHI (R^2=.29, *P*<.05). The regression coefficients were statistically significant for the following independent variables (*P*<.05): age, gender, levels of education, and having a chronic health problem.

Overall, 31.0% of the variance in frequency of using OHI can be predicted by the personal characteristics of Internet users (*F*=9.94, *P*<.05, Model 4). The multiple regression analysis showed that the main predictors are being female, of younger age, having a higher level of education, reporting good general health, having a chronic health problem, and using the Internet as a main source of health information.

Discussion

Principal Findings

In our study, we explored characteristics and behaviors of Egyptian OHI seekers. Our findings showed that the main personal characteristics associated with more frequent OHI searching were younger age, female gender, higher education, and reporting good general health.

The study participants represent a small sample of Egyptian Internet users who seek OHI. We have no data of those who declined to take part in the study or those who had not searched OHI at any given time. Accordingly, the results of the study can be generalized only to Egyptian OHI seekers. Participants' answers to the questionnaire were self-reported and could not be independently verified. None of the participants rated the quality of OHI as poor, which may be because participants who took part in the study were active users of a health website.

In relation to the association between gender and OHI-seeking behavior, our findings show that female participants (57.1%) have looked for OHI more frequently than male participants (42.9%). This may arise from curiosity about their own health, as women seem to be more involved in decisions about their health [26]. Similar results were reported by Rice et al, which shows that the female gender is an important predictor for OHI-seeking behavior [12]. However, this should be seen in the context that according to a 2013 report from the Egyptian MCIT, a minority (44.4%) of Egyptian Internet users were females, whereas 56.6% were males [5].

The report also stated that 72% of Egyptian Internet users ranged between 15-44 years old, whereas only 19% were aged 45 years or older [5]. In this study, we found that younger participants sought OHI more frequently than older participants. The reason for that may be that younger OHI seekers are more familiar with Internet use [25]. Similarly, Nolke et al reported that the probability of seeking OHI in German participants decreased with increasing age [27].

Participants' level of education also had a considerable impact on OHI seeking behavior. In our study, participants with higher educational level spent more time seeking for OHI. This is to be expected given that 73% of Egyptian Internet users were university graduates [5]. Cotten et al found that OHI seekers tended to have higher levels of education than those who did not seek OHI, which confirms a link between educational level and OHI seeking [28].

Our findings also show that OHI seekers with good self-reported general health were likely to search for OHI more frequently than those with poor general health. Maybe because more than 50% of the participants were less than 35 years old, this meant that they were more likely to be in good general health. Similarly, Weaver et al found that healthier people use OHI more frequently in a proactive manner for health promotion or to maintain a healthy lifestyle [8]. However, Cotten et al found that healthier persons seek OHI less frequently than those with poor general health [28].

We found that people who reported one or more chronic health problems were more likely to search for OHI more frequently than people without any chronic health conditions. This link between having a chronic health problem and the frequency of OHI seeking is supported by some researchers, whereas others did not find a significant association [10,29-32]. The reason for this discrepancy is not clear but it might be argued that the two groups used OHI differently: healthy participants focused on improving their lifestyles, whereas people with chronic illness sought to improve their current health condition.

In this study, participants reported seeking OHI for a broad range of health topics including nutrition, sexual health, chronic health problems, and mental health. We found that nutrition-related health information is the most common topic sought on the Internet especially by female participants. This finding can be seen in the context of increased prevalence of nutrition-related health problems among Egyptian adults [33]. As reported in the DHS report in 2015, 30% of Egyptian adults are overweight or obese, and 6.8% of Egyptian children under the age of 5 years were malnourished (eg, micronutrients deficiency or anemic) [34,35].

Despite the ease of finding health information on the Internet, it may be difficult to judge the reliability of this health information. This may explain the variability in the level of trust in OHI [17]. Most OHI seekers in Arabic countries rely on health sites maintained mainly by the private sector rather than by ministries of health or other public services. Arabic language health information websites need a substantial improvement of their quality in order to improve their trustworthiness [6]. However, most of our participants rated the quality of the OHI accessed as “good” or “excellent.” For some, this may be linked to lack of trust in health professionals as much as increased trust in OHI sites.

Conclusions

The Internet has become an important tool with the potential to improve information dissemination and health care delivery to consumers. Continuing efforts to maximize the potential of this tool could have great value for users. The great range of health topics as well as the easy access to OHI has made the Internet a major and growing source of health information for Egyptian Internet users. We found that personal characteristics of OHI seekers such as younger age, females, higher levels of education, and having generally good health make a difference in how frequently they access the Internet for health information. An understanding of personal characteristics may inform education of OHI seekers to help them to take an active role in improving their health and engaging with treatment plans; however, more research is needed to identify whether these characteristics can be confirmed in a representative sample of Egyptian Internet users.

It is also concerning that most providers of Arabic language OHI are privately owned and therefore, their priority may be increasing advertising revenue rather than providing reliable information. There is a need for public funding for quality controlled websites, for example, equivalent to the United Kingdom's National Health Service (NHS) Choices. Cooperation from health care providers is also needed in supporting their patients with reliable resources and assisting them to properly evaluate the quality of information available on the Internet in order to promote efficient, reliable, and appropriate OHI use for Internet users.

Conflicts of Interest

The authors whose names are listed in this manuscript certify that they have no conflict of interest in the subject matter or materials discussed in this manuscript.

References

1. Laugesen J, Hassanein K, Yuan Y. The impact of internet health information on patient compliance: a research model and an empirical study. J Med Internet Res 2015;17(6):e143 [FREE Full text] [doi: 10.2196/jmir.4333] [Medline: 26068214]
2. Cutilli CC. Seeking health information: what sources do your patients use? Orthop Nurs 2010;29(3):214-219. [doi: 10.1097/NOR.0b013e3181db5471] [Medline: 20505493]
3. Gaal I. [Impact of the internet on physician-patient relationship]. Orv Hetil 2016 Apr 24;157(17):680-684. [doi: 10.1556/650.2016.30456] [Medline: 27084443]

4. Internet world status. 2016. Egypt internet usage and telecommunications report URL: http://www.internetworldstats.com/stats1.htm [accessed 2016-09-08] [WebCite Cache ID 6kMtXk4zQ]
5. MCIT. Measuring the digital society in Egypt: internet at a glance URL: http://www.mcit.gov.eg/Upcont/Documents/Publications_1272015000_Measuring_the_Digital_Society_in_Egypt_12_.pdf [accessed 2017-05-30] [WebCite Cache ID 6qqPPeMqg]
6. Huziah M. Hon.ch. 2010. Assessment of online health information for Arabic sites URL: http://www.hon.ch/Global/pdf/Press/Assessment_of_Arabic_Sites_Report_17-09-09.pdf [accessed 2017-02-15] [WebCite Cache ID 6oIJVgVIZ]
7. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
8. Weaver JB, Mays D, Weaver SS, Hopkins GL, Eroglu D, Bernhardt JM. Health information-seeking behaviors, health indicators, and health risks. Am J Public Health 2010 Aug;100(8):1520-1525. [doi: 10.2105/AJPH.2009.180521] [Medline: 20558794]
9. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [Medline: 11929503]
10. 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]
11. Gallagher S, Doherty DT. Searching for health information online: characteristics of online health seekers. J Evid Based Med 2009 May;2(2):99-106. [doi: 10.1111/j.1756-5391.2009.01021.x] [Medline: 21348996]
12. 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]
13. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
14. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
15. Bessière K, Pressman S, Kiesler S, Kraut R. Effects of internet use on health and depression: a longitudinal study. J Med Internet Res 2010;12(1):e6 [FREE Full text] [doi: 10.2196/jmir.1149] [Medline: 20228047]
16. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 4;362(9):859-860. [doi: 10.1056/NEJMc0909595] [Medline: 20200398]
17. Moretti FA, Oliveira VE, Silva EM. Access to health information on the internet: a public health issue? Rev Assoc Med Bras 2012;58(6):650-658 [FREE Full text] [Medline: 23250092]
18. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
19. LaValley SA, Kiviniemi MT, Gage-Bouchard EA. Where people look for online health information. Health Info Libr J 2017 Jun;34(2):146-155. [doi: 10.1111/hir.12143] [Medline: 27207817]
20. 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]
21. Edejer TT. Disseminating health information in developing countries: the role of the internet. BMJ 2000 Sep 30;321(7264):797-800 [FREE Full text] [Medline: 11009519]
22. Al-Shorbaji N. E-health in the Eastern Mediterranean Region: a decade of challenges and achievements. East Mediterr Health J 2008;14 Suppl:S157-S173. [Medline: 19205616]
23. AlSaadi MM. Evaluation of internet use for health information by parents of asthmatic children attending pediatric clinics in Riyadh, Saudi Arabia. Ann Saudi Med 2012;32(6):630-636. [doi: 10.5144/0256-4947.2012.630] [Medline: 23396028]
24. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
25. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
26. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. Health Promot Int 2015 Sep;30(3):736-745 [FREE Full text] [doi: 10.1093/heapro/dat063] [Medline: 23985248]
27. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
28. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
29. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]

30. Wilson V. Patient use of the internet for diabetes information. Nurs Times 2013;109(23):18-20. [Medline: 23888634]
31. Weymann N, Dirmaier J, von WA, Kriston L, Härter M. Effectiveness of a web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
32. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
33. Hegazi R, El-Gamal M, Abdel-Hady N, Hamdy O. Epidemiology of and risk factors for type 2 diabetes in Egypt. Ann Glob Health 2015;81(6):814-820 [FREE Full text] [doi: 10.1016/j.aogh.2015.12.011] [Medline: 27108148]
34. The DHS Program. 2015. Egypt health issues survey URL: https://dhsprogram.com/pubs/pdf/FR313/FR313.pdf [accessed 2017-02-15] [WebCite Cache ID 6oIj8n2yt]
35. Kavle JA, Flax VL, Abdelmegeid A, Salah F, Hafez S, Ramzy M, et al. Factors associated with early growth in Egyptian infants: implications for addressing the dual burden of malnutrition. Matern Child Nutr 2016 Jan;12(1):139-151 [FREE Full text] [doi: 10.1111/mcn.12213] [Medline: 26373408]

Abbreviations

OHI: online health information
MCIT: Ministry of Communication and Information Technology
NHS: National Health Service

Edited by G Eysenbach; submitted 21.10.16; peer-reviewed by C Fincham, K Hassanein; comments to author 05.01.17; revised version received 15.02.17; accepted 06.04.17; published 22.06.17.

Please cite as:
Ghweeba M, Lindenmeyer A, Shishi S, Abbas M, Waheed A, Amer S
What Predicts Online Health Information-Seeking Behavior Among Egyptian Adults? A Cross-Sectional Study
J Med Internet Res 2017;19(6):e216
URL: http://www.jmir.org/2017/6/e216/
doi:10.2196/jmir.6855
PMID:28642216

Review

Researching Mental Health Disorders in the Era of Social Media: Systematic Review

Akkapon Wongkoblap[1], BSc, MSc; Miguel A Vadillo[2,3], PhD; Vasa Curcin[1,2], PhD

[1]Department of Informatics, King's College London, London, United Kingdom
[2]Primary Care and Public Health Sciences, King’s College London, London, United Kingdom
[3]Departamento de Psicología Básica, Universidad Autónoma de Madrid, Madrid, Spain

Corresponding Author:
Akkapon Wongkoblap, BSc, MSc
Department of Informatics
King's College London
Strand
London, WC2R 2LS
United Kingdom
Phone: 44 20 7848 2588
Fax: 44 20 7848 2017
Email: akkapon.wongkoblap@kcl.ac.uk

Abstract

Background: Mental illness is quickly becoming one of the most prevalent public health problems worldwide. Social network platforms, where users can express their emotions, feelings, and thoughts, are a valuable source of data for researching mental health, and techniques based on machine learning are increasingly used for this purpose.

Objective: The objective of this review was to explore the scope and limits of cutting-edge techniques that researchers are using for predictive analytics in mental health and to review associated issues, such as ethical concerns, in this area of research.

Methods: We performed a systematic literature review in March 2017, using keywords to search articles on data mining of social network data in the context of common mental health disorders, published between 2010 and March 8, 2017 in medical and computer science journals.

Results: The initial search returned a total of 5386 articles. Following a careful analysis of the titles, abstracts, and main texts, we selected 48 articles for review. We coded the articles according to key characteristics, techniques used for data collection, data preprocessing, feature extraction, feature selection, model construction, and model verification. The most common analytical method was text analysis, with several studies using different flavors of image analysis and social interaction graph analysis.

Conclusions: Despite an increasing number of studies investigating mental health issues using social network data, some common problems persist. Assembling large, high-quality datasets of social media users with mental disorder is problematic, not only due to biases associated with the collection methods, but also with regard to managing consent and selecting appropriate analytics techniques.

(J Med Internet Res 2017;19(6):e228) doi:10.2196/jmir.7215

KEYWORDS

mental health; mental disorders; social networking; artificial intelligence; machine learning; public health informatics; depression; anxiety; infodemiology

Introduction

Mental illness is quickly becoming one of the most serious and prevalent public health problems worldwide [1]. Around 25% of the population of the United Kingdom have mental disorders every year [2]. According to statistics published by the World Health Organization, more than 350 million people have depression. In terms of economic impact, the global costs of mental health problems were approximately US $2.5 trillion in 2010. By 2030, it is estimated that the costs will increase further to US $6.0 trillion [3]. Mental disorders include many different illnesses, with depression being the most prominent. Additionally, depression and anxiety disorders can lead to suicidal ideation and suicide attempts [1]. These figures show that mental health problems have effects across society, and demand new prevention and intervention strategies. Early detection of mental illness is an essential step in applying these strategies, with the mental illnesses typically being diagnosed using validated questionnaires designed to detect specific patterns of feelings or social interaction [4-6].

Online social media have become increasingly popular over the last few years as a means of sharing different types of user-generated or user-curated content, such as publishing personal status updates, uploading pictures, and sharing current geographical locations. Users can also interact with other users by commenting on their posts and establishing conversations. Through these interactions, users can express their feelings and thoughts, and report on their daily activities [7], creating a wealth of useful information about their social behaviors [8]. To name just 2 particularly popular social networks, Facebook is accessed regularly by more than 1.7 billion monthly active users [9] and Twitter has over 310 million active accounts [10], producing large volumes of data that could be mined, subject to ethical constraints, to find meaningful patterns in users' behaviors.

The field of data science has emerged as a way of addressing the growing scale of data, and the analytics and computational power it requires. Machine learning techniques that allow researchers to extract information from complex datasets have been repurposed to this new environment and used to interpret data and create predictive models in various domains, such as finance [11], economics [12], politics [13], and crime [14]. In medical research, data science approaches have allowed researchers to mine large health care datasets to detect patterns and accrue meaningful knowledge [15-18]. A specific segment of this work has focused on analyzing and detecting symptoms of mental disorders through status updates in social networking websites [19].

Based on the symptoms and indicators of mental disorders, it is possible to use data mining and machine learning techniques to develop automatic detection systems for mental health problems. Unusual actions and uncommon patterns of interaction expressed in social network platforms [19] can be detected through existing tools, based on text mining, social network analysis, and image analysis.

Even though the current performance of predictive models is suboptimal, reliable predictive models will eventually allow early detection and pave the way for health interventions in the forms of promoting relevant health services or delivering useful health information links. By harnessing the capabilities offered to commercial entities on social networks, there is a potential to deliver real health benefits to users.

This systematic review aimed to explore the scope and limits of cutting-edge techniques for predictive analytics in mental health. Specifically, in this review we tried to answer the following questions: (1) What methods are researchers using to collect data from online social network sites such as Facebook and Twitter? (2) What are the state-of-the-art techniques in predictive analytics of social network data in mental health? (3) What are the main ethical concerns in this area of research?

Methods

We conducted a systematic review to examine how social media data have been used to classify and predict the mental health state of users. The procedure followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) to outline and assess relevant articles [20].

Literature Search Strategy

We searched the literature in March 2017, collecting articles published between 2010 and March 8, 2017 in medical and computer science databases. We searched PubMed, Institute of Electrical and Electronics Engineers (IEEE Xplore), Association for Computing Machinery (ACM Digital Library), Web of Science, and Scopus using sets of keywords focused on the prediction of mental health problems based on data from social media. We restricted our searches to common mental health disorders, as defined by the UK National Institute for Health and Care Excellence [21]: depression, generalized anxiety disorder, panic disorder, phobias, social anxiety disorder, obsessive-compulsive disorder (OCD), and posttraumatic stress disorder (PTSD). To ensure that our literature search strategy was as inclusive as possible, we explored Medical Subject Headings (MeSH) for relevant key terms. MeSH terms were used in all databases that made this option available. Search terms are outlined in Textbox 1.

In addition, we manually searched the proceedings of the Computational Linguistics and Clinical Psychology Workshops (CLPsych) and the outputs of the World Well-Being Project [22] to find additional articles that our search terms might have excluded. Furthermore, we examined the reference lists of included articles for additional sources.

Textbox 1. Search strategy to identify articles on the prediction of mental health problems based social media data.

Medical Subject Headings (MeSH)

1. Depression/ or Mental Health/ or Mental Disorders/ or Suicide or Life Satisfaction/ or Well Being/ or Anxiety/ or Panic/ or Phobia/ or OCD/ or PTSD
2. Social Media/ or Social Networks/ or Facebook/ or Twitter/ or Tweet
3. Machine Learning/ or Data Mining/ or Big Data/ or Text Analysis/ or Text Mining/ or Predictive Analytics/ or Prediction/ or Detection/ or Deep Learning
4. (1) and (2)
5. (1) and (3)

Inclusion and Exclusion Criteria

We further filtered the titles and abstracts of articles retrieved using the search terms outlined in Textbox 1. Only articles published in peer-reviewed journals and written in English were included. Further inclusion criteria were that studies had to (1) focus on predicting mental health problems through social media data, and (2) investigate prediction or classification models based on users' text posts, network interactions, or other features of social network platforms. Within this review, we focused on social network platforms—that is, those allowing users to create personal profiles, post content, and establish new or maintain existing relationships.

Studies were excluded if they (1) only analyzed the correlation between social network data and symptoms of mental illness, (2) analyzed textual contents only by human coding or manual annotation, (3) examined data from online communities (eg, LiveJournal), (4) focused on the relationship between social media use and mental health disorders (eg, so-called Internet addiction), (5) examined the influence of cyberbullying on mental health, or (6) did not explain where the datasets came from.

Data Extraction

After screening articles and obtaining a set of studies that met our inclusion criteria, we extracted the most relevant data from the main texts. These are title, author, aims, findings, methods, data collection on machine learning techniques, sampling, questionnaire, platform, and language.

Results

Overview

Figure 1 presents a PRISMA flow diagram of the results of searching and screening articles following the above search methodology. The initial search resulted in a total of 5371 articles plus 11 additional articles obtained through CLPsych, 1 from the World Well-Being Project, and 3 from the reference lists of included articles. We removed 1864 of these articles because of duplication. Each of the remaining articles (n=3522) was screened by reviewing its title and abstract. If an article analyzed data from other sources (such as brain signals, mental health detection from face detection, or mobile sensing), we discarded it. This resulted in a set of 106 articles. By matching these with our inclusion and exclusion criteria, we removed a further 58 articles. To sum up, we excluded 5338 articles and included 48 in the review (see Figure 1).

We extracted data from each of the 48 articles. Table 1 and Multimedia Appendix 1 (whose format is adapted from previous work [11,23]) show the key characteristics of the selected studies [24-71], ordered by year published. Of the studies reviewed, 46 were published from 2013 onward, while only 2 peer-reviewed articles were published between 2011 and 2012. None of the selected articles was published in 2010.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. CLPsych: Computational Linguistics and Clinical Psychology Workshops.

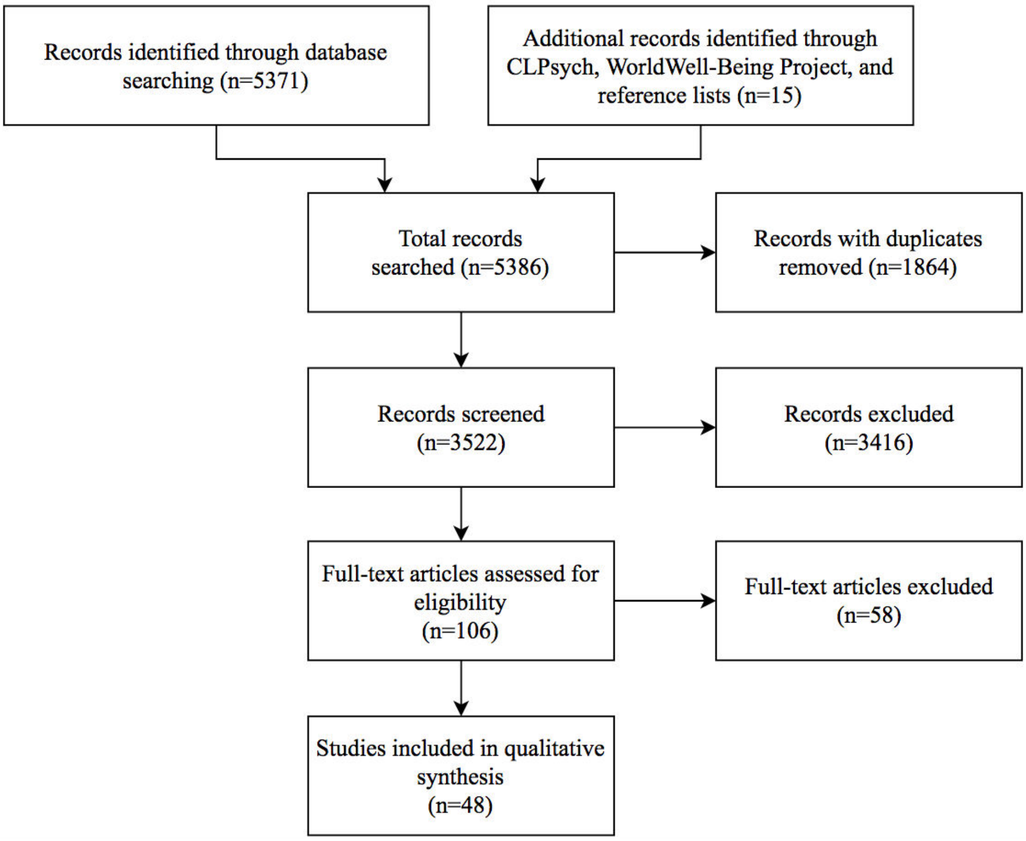


Table 1. Summaries of articles reviewed.

First author, date, reference	Aims	Findings
Wang, 2017 [24]	To explore and characterize the structure of the community of people with eating disorders using Twitter data and then classify users into those with and without the disorder.	There was assortativity among users with eating disorder. The classifier distinguished 2 groups of people.
Volkava, 2016 [25]	To explore academic discourse from tweets and build predictive models to analyze the data.	Tweets from students across 44 universities were related to student surveys on satisfaction and happiness.
Saravia, 2016 [26]	To present a new data collection method and classify individuals with mental illness and nonmental illness.	The proposed method and a classifier were built as an online system, which distinguished 2 groups of individuals and provided mental illness information.
Kang, 2016 [27]	To propose classification models to detect tweets of users with depression for a long period of time. Classifiers were based on the texts, emoticons, and images they posted.	The models detected users with depression.
Schwartz, 2016 [28]	To present predictive models to estimate individual well-being through textual content on social networks.	A combination of message- and user-level aggregation of posts performed well.
Chancellor, 2016 [29]	To explore posts from Instagram to forecast levels of mental illness severity of pro-eating disorder.	Future mental illness severity could be predicted from user-generated messages.
Braithwaite, 2016 [30]	To explore machine learning algorithms to measure suicide risk in the United States.	Machine learning algorithms successfully classified users with suicidal ideation.
Coppersmith, 2016 [31]	To explore linguistics and emotional patterns in Twitter users with and without suicide attempt.	There were quantifiable signals of suicide attempt in tweets.
Lv, 2015 [32]	To build a Chinese suicide dictionary, based on Weibo posts, to detect suicide risk.	The Chinese suicide dictionary detected individuals and tweets at suicide risk.
O'Dea, 2015 [33]	To explore machine learning models to automatically detect the level of concern for each suicide-related tweet.	Machine learning classifiers estimated the level of concern from suicide-related tweets.
Liu, 2015 [34]	To investigate and predict users' subjective well-being based on Facebook posts.	Users' subjective well-being could be predicted from posts and their time frame.
Burnap, 2015 [35]	To explore suicide-related tweets to understand users' communications on social media.	Classification models classified tweets into relevant suicide categories.
Park, 2015 [36]	To analyze the relationships between Facebook activities and the depression state of users.	Participants with depression had fewer interactions, such as receiving likes and comments. Depressed users posted at a higher rate.
Hu, 2015 [37]	To present classifiers with different lengths of observation time to detect depressed users.	Behavioral and linguistic features predicted depression. A 2-month period of observation enabled prediction cues of depression half a month in advance.
Tsugawa, 2015 [38]	To develop a model to recognize individuals with depression from non-English social media posts and activities.	Activities extracted from Twitter were useful to detect depression; 2 months of observation data enabled detection of symptoms of depression. The topics estimated by LDA[a] were useful.
Zhang, 2015 [39]	To explore 2 natural language processing algorithms to identify posts predicting the probability of suicide.	LDA automatically detected suicide probability from textual contents on social media.
Coppersmith, 2015 [40]	To explore tweet content with self-reported health sentences and language differences in 10 mental health conditions.	There were quantifiable signals of 10 mental health conditions in social network messages and relations between them.
Preotiuc-Pietro, 2015 [41]	To implement linear classifiers to detect users with PTSD[c] and depression based on user metadata, and several textual and topic features.	The combination of linear classifiers performed better than average classifiers. All unigram features performed well.
Mitchell, 2015 [42]	To use several natural language processing techniques to explore the language of schizophrenic users on Twitter.	Character ngram features were used to train models to classify users with and without schizophrenia. LDA outperformed linguistic inquiry and word count.
Preotiuc-Pietro, 2015 [43]	To study differences in language use in tweets about mental health depending on the role of personality, age, and sex of users.	Personality and demographic data extracted from tweets detected users with depression or PTSD.
Pedersen, 2015 [44]	To explore and study the accuracy of decision lists of ngrams to classify users with depression and PTSD.	Bigram features underperformed ngram 1-6 features.

First author, date, reference	Aims	Findings
Resnik, 2015 [45]	To build classifiers to categorize depressed and nondepressed users, based on supervised topic models.	LDA mined useful information from tweets. Supervised topic models such as supervised LDA and supervised anchor model improved LDA accuracy.
Resnik, 2015 [46]	To build classifiers with TF-IDF[d] weighting, using support vector machine with a linear kernel or radial basis function kernel.	TF-IDF showed good performance, and TF-IDF with supervised topic model performed even better.
Durahim, 2015 [47]	To explore data from social networks to measure the Gross National Happiness of Turkey.	Sentiment analysis estimated Gross National Happiness levels similar to Turkish statistics.
Guan, 2015 [48]	To explore 2 types of classifiers to detect posts revealing high suicide risk.	Users' profiles and their generated text were used to classify users with high or low suicide risk.
Landeiro Dos Reis, 2015 [49]	To explore exercise-related tweets to measure their association with mental health.	Users who posted workouts regularly tended to express lower levels of depression and anxiety.
De Choudhury, 2014 [50]	To explore several types of Facebook data to detect and predict postpartum depression.	Postpartum depression was predicted from an increase of social isolation and a decrease of social capital.
Huang, 2014 [51]	To present a framework to detect posts related to suicidal ideation.	The best predictive model was based on support vector machine.
Wilson, 2014 [52]	To explore the types of mental health information posted and shared on Twitter.	The study distinguished 8 themes of information about depression in Twitter posts, each having different features.
Coppersmith, 2014 [53]	To present a novel method to collect posts related to PTSD and build a classifier.	The classifier distinguished users with and without self-reported PTSD.
Kuang, 2014 [54]	To create the Chinese version of the extended PERMA[e] corpus and use it to measure happiness scores.	The proposed model measured happiness.
Hao, 2014 [55]	To propose machine learning models to measure subjective well-being of social media users.	The model measured subjective well-being from social media data.
Prieto, 2014 [56]	To develop a machine learning model to detect and measure the prevalence of health conditions.	The proposed methods identified the presence of health conditions on Twitter.
Lin, 2014 [57]	To develop a deep neural network model to classify users with or without stress.	The trained model detected stress from user-generated content.
Schwartz, 2014 [58]	To build predictive models to detect depression based on Facebook text.	Facebook updates enabled distinguishing depressed users. Predictive models offered insights into seasonal affective disorder.
Coppersmith, 2014 [59]	To analyze tweets related to health and propose a new method to quickly collect public tweets containing statements of mental illnesses.	There were differences in quantifiable linguistic signals of bipolar disorder, depression, PTSD, and seasonal affective disorder in tweets.
Homan, 2014 [60]	To examine the potential of tweet content to classify suicidal risk factors.	Annotations from novices and experts were used to train classifiers, although expert annotations outperformed novice annotations.
Park, 2013 [61]	To develop a Web app to detect symptoms of depression from features extracted from Facebook.	Depressed users had fewer Facebook friends, used fewer location tags, and tended to have fewer interactions.
Wang, 2013 [62]	To build a depression detection model based on sentiment analysis of data from social media.	Sentiment analysis with 10 features detected users with depression, with 80% accuracy.
Wang, 2013 [63]	To explore a detection model, based on node and linkage features, to recognize the presence of depression in social media users. This was an extended version of their earlier study [62].	The node and linkage features model performed better than the model based just on node features.
Tsugawa, 2013 [64]	To explore the effectiveness of an analytic model to estimate depressive tendencies from users' activities on a social network.	There was a correlation between the Zung Self-Rating Depression Scale and the model estimations.
De Choudhury, 2013 [65]	To explore predictive models to classify mothers with a tendency to change behavior after giving birth or to experience postpartum depression.	Tweets during prenatal and early postnatal periods predicted future behavior changes, with an accuracy of 71%. Data over 2-3 weeks after giving birth improved prediction results, with an accuracy of 80%-83%.
De Choudhury, 2013 [66]	To explore the potential of a machine learning model to measure levels of depression in populations.	The proposed model estimated levels of depression.
De Choudhury, 2013 [67]	To develop a prediction model to classify individual users with depression.	The predictive model classified users with depression.

First author, date, reference	Aims	Findings
Schwartz, 2013 [68]	To analyze tweets from different US counties to predict well-being of people in those areas.	Topic features provided useful information about life satisfaction.
Hao, 2013 [69]	To explore the mental state of users through their online behavior.	Online behavior enabled prediction of mental health problems.
Jamison-Powell, 2012 [70]	To explore the characteristics of tweets that included the #insomnia hashtag.	Tweets about insomnia contained more negative words. People used Twitter to express their symptoms and ideas for coping strategies.
Bollen, 2011 [71]	To explore an online social network to measure subjective well-being levels of users and calculated assortativity.	There was assortativity among Twitter users.

[a] LDA: latent Dirichlet allocation.

[b] CLPsych: Computational Linguistics and Clinical Psychology Workshops.

[c] PTSD: posttraumatic stress disorder.

[d] TF-IDF: term frequency-inverse document frequency.

[e] PERMA: positive emotions, engagement, relationships, meaning, and accomplishment.

The selected studies can be divided into several distinct categories. Several studies [27,36-38,40-46,52,56-59,61-64,66,67] used datasets from social networks to examine depression. Postpartum depression disorder was explored by De Choudhury et al [50,65], PTSD was investigated by 8 studies [40,41,43-46,53,59]. Anxiety and OCD were investigated by 2 studies [40,69]. Borderline disorder and bipolar disorder were investigated by 3 studies [26,40,59]. Seasonal affective disorder was studied by Coppersmith et al [40,59]. Eating disorder was explored by Chancellor et al [29], Coppersmith et al [40], and Prieto et al [56]. Attention-deficit/ hyperactivity disorder, anxiety, and schizophrenia were examined by Coppersmith et al [40], and sleep disorder was studied by Jamison-Powell et al [70]. None of the included studies explored phobias or panic disorders. Users with suicidal ideation were investigated by 8 studies [30-33,35,39,51,60]. Happiness, satisfaction with life, and well-being were investigated by 7 studies [28,34,47,54,55,68,71].

Of the studies included in this review, 31 analyzed social network contents written in English [24-31,33-35,40-46,49,50,52,53,58-60,65-68,70,71]; 11 studies investigated Chinese text [32,37,39,48,51,54,55,57,62,63,69]; 2 focused on Korean [36,61] and 2 on Japanese text [38,64], 1 looked at Turkish content [47], and 1 jointly at Spanish and Portuguese [56].

Data Collection Techniques

Each of the selected articles was based on a dataset directly or indirectly obtained from social networks. We identified 2 broad approaches to data collection: (1) collecting data directly from the participants with their consent using surveys and electronic data collection instruments (eg, Facebook apps), and (2) aggregating data extracted from public posts.

The methods for collecting data directly from participants varied with the purpose of the studies and the target platform. These methods included posting project information on relevant websites inviting participants to take part in the project [32,38,50] and posting tasks on crowdsourcing platforms asking for project volunteers [28,30,66,67]. For crowdsourcing, researchers posted detailed information about their studies on platforms such as Amazon Mechanical Turk [74] to attract participants. As part of a questionnaire, the participants would typically be asked to provide informed consent allowing collection of their social network data.

A range of questionnaires were used to measure participants' levels of depression and life satisfaction, including the Center for Epidemiologic Studies Depression Scale [36,38,61,66,67], Patient Health Questionnaire-9 [50], Beck Depression Inventory [36,38,61,67], Zung Self-Rating Depression Scale [64], Depressive Symptom Inventory-Suicidality Subscale [30], and Symptom Checklist-90-Revised [69]. The instruments used to detect suicidal ideation and the possibility of an individual committing suicide were the Suicide Probability Scale [32,39,48], the Acquired Capability for Suicide Scale [30], and the Interpersonal Needs Questionnaire [30]. Satisfaction with life and well-being were measured with the Satisfaction with Life Scale [28,34], the Positive and Negative Affect Schedule [55], and the Psychological Well-Being Scale [55]. One study used the Revised NEO Personality Inventory-Revised to assess personality [58].

The second approach was to pool only public posts from social network platforms, by using regular expressions to search for relevant posts, such as “I was diagnosed with [condition name]” [40,42,43,59].

To collect social network data, each data source required a custom capture mechanism, due to a lack of standards for data collection. Facebook-based experiments gathered user datasets by developing custom tools or Web apps connecting to the Facebook application programming interfaces (APIs) [36,50,61]. Another group of studies used Twitter APIs to explore cues for mental disorders [24-27,30,31,33,35,38,47,52,53,56-60,64-68,70,71]. A similar approach was used for Instagram APIs [29] and Sina Weibo APIs [32,37,39,51,54,55,57,62,63].

Another way of obtaining data was promoted by the myPersonality project, which provides both social network data and a variety of psychometric test scores for academic researchers [75], and was used by 3 studies [28,34,58]. Some studies [41,44-46] originated from workshops where the organizers shared data already approved by an institutional review board (IRB) for analysis.

Translating Collected Data Into Knowledge and Results

In all of the selected studies, several standard steps had to be taken before machine learning algorithms could be applied to data. First, data were cleaned and preprocessed to ensure that they were in the form required by the analytical algorithms. Second, the key features (the term “feature” in machine learning denotes a set of observations that is relevant to the modelling problem, typically represented numerically [76]) related to the research domain were prepared for model construction. Overall, this involves feature extraction and feature selection, producing sets of features to be used in learning and validating predictive models.

Data Preprocessing

The corpus of data is typically preprocessed by (1) removing unsuitable samples and (2) cleaning and preparing the data for analysis. Information and questionnaires from participants might contain useless data and incomplete details, which are usually removed from studies in order to improve the accuracy of prediction and classification results. Participants who take an abnormally short or long time to complete the questionnaires were excluded from 4 studies [38,39,66,67]. Low-activity participants who had published less than a defined threshold of posts were removed from 8 studies [26,32,34,37,39,55,59,71]. Participants with poor correlations between 2 different questionnaires were excluded from the final dataset in 2 studies [38,67].

As part of the data cleaning process, each post was checked for the majority written language (eg, contained at least 70% English words [28,40,42,53,59,70]). This ensured that the available tools were suitable to analyze the posts. Each post was preprocessed by eliminating stop words and irrelevant data (eg, retweets, hashtags, URLs), lowercasing characters, and segmenting sentences [31,44,46,53,56,60,66]. Emoticons were converted to other forms such as ASCII codes [45] to ensure data were machine readable. Anonymization was also performed to remove any potentially identifiable usernames [31,33,35,52,53,70].

Feature Extraction

There are many potential techniques to extract features that could be used for predicting mental health problems in social network users. Several studies have attempted to investigate the textual contents of social networks to understand what factors contain cues for mental disorders. However, some research projects have used alternative techniques. In this review, we identified three broad approaches to feature extraction: text analysis, image analysis, and social interaction.

In text mining, sentiment analysis is a popular tool for understanding emotion expression. It is employed to classify the polarity of a given text into categories such as positive, negative, and neutral [77]. Several studies [24,28,30,32,34,39,49,50,52-55,57,60,65-68,70] used the well-known linguistic inquiry and word count (LIWC) [78] to extract potential signals of mental problems from textual content (eg, the word frequency of the first personal pronoun “I” or “me” or of the second personal pronoun, positive and negative emotions being used by a user or in a post). OpinionFinder [79] was used by Bollen et al [71] and SentiStrength [80] was used by Kang et al [27] and by Durahim and Coşkun [47] to carry out sentiment analysis. Custom tools were also developed for performing sentiment analysis. Affective Norms for English Words [81] was used to qualify the emotional intensity of English words in 2 studies [65,66], while topic modelling was employed in 4 studies [28,29,38,39] to extract topics from user-generated posts.

Social media posts tend to be rich in various emoticons. As a consequence, several studies [27,62] looked into the meaning and mood states associated with their use.

Apart from posting text messages, social network platforms allow users to post images. Some studies investigated these images for cues of mental disorders [27,57]. Color compositions and scale-variant feature transform descriptor techniques were used to extract emotional meanings of each individual image [27]. Image properties, comprising color theme, saturation, brightness, color temperature, and color clarity, were analyzed be Lin et al [57].

Finally, social network platforms contain millions of interactions and relationships among users. Social network users not only can connect and add online friends, but also can post, comment, and reply to their friends. The resulting graph structure, comprising information about interactions, relationships, and friendships, was mined to understand the cues that can be connected to symptoms of mental disorders (eg, interactions among depressed users and assortative mixing patterns) [24,63,71].

Feature Selection

Feature selection isolates a relevant subset of features that are able to predict symptoms of mental disorders or correctly label participants, while avoiding overfitting. Statistical analysis is typically performed to discover a set of parameters that can differentiate between users with mental disorders and users without mental disorders. The techniques used in the selected studies were Pearson correlation coefficient [36,55,56], correlation-based feature selection [56], Spearman rank correlation coefficient [61], and Mann-Whitney *U* test [61]. The dimensionality of features was reduced by principal component analysis [35,58,65-67], randomized principal component analysis [28], convolutional neural network with cross-autoencoder technique [57], forward greedy stepwise [37], binary logistic regression [62], gain ratio technique [56], and relief technique [56].

Predictive Model Construction

In the selected studies, prediction models were used to detect and classify users according to mental disorders and satisfaction with life. To build a predictive model, a selected set of features is used as training data for machine learning algorithms to learn patterns from those data.

All the articles included in this review used supervised learning techniques, where the sample data contain both the inputs and the labeled outputs. The model learns from these to predict unlabeled inputs from other sources and provide prediction outputs [82]. The techniques used in these studies included support vector machine (SVM) [32,33,35,38,42,56,69], linear SVM [24,27,41,46,60], and SVM with a radial basis function kernel [24,27,46,51,65-67]. Regression techniques included ridge regression [28], linear regression [37,58], log-linear regression [53,59], logistic regression [25,31,33,37,48,49,51], binary logistic regression with elastic net regularization [41,43], linear regression with stepwise selection [39,55,64], stepwise logistic regression with forward selection [50], regularized multinomial logistic regression [29], linear support vector regression [45,55], least absolute shrinkage and selection operator [55,68], and multivariate adaptive regression splines [55]. Other algorithms used for binary classification were decision trees [35,51,56,62,63], random forest [26,48,51], rules decision [62], naive Bayes [24,35,51,56,62,69], k-nearest neighbor [24,56], maximum entropy [42], neural network [69], and deep learning neural network [57].

Model Verification

Following model construction, its accuracy was measured using a test dataset. The most common model validation technique was *n*-fold cross-validation, which randomly partitions a dataset into *n* equal subsets and proceeds to iterate *n* times, with each subset used for validation exactly once, while the remaining *n*–1 subsets are used as training data [82]. Several studies [26,31,35,37-43,49,51,56,67,69] employed 10-fold cross-validation to verify their prediction models and classifiers, while 5-fold cross-validation was used by 4 studies [24,48,55,57]. Leave-one-out cross-validation was used in 2 studies [30,59].

The performance of predictive models can also be evaluated in other datasets. Several studies [27-29,33,45,46,49,58,60,64,68] divided the collected dataset into training and test subsets to measure the accuracy of their models. Some [47,48,53,54,66] collected a new dataset to evaluate the accuracy of the predicted results and compare the predicted results with a set of known statistics (eg, depression rates in US cities, student satisfaction survey, and Gross National Happiness percentages of provinces of Turkey).

Ethics

The ethical aspects of using social network data for research are still not clearly defined, particularly when working with information that is publicly available. Thus, the studies that we surveyed adopted a wide range of approaches to handle ethical constraints.

Among the articles included in this review, 9 [30,32,33,36,38,40,42,48,61] were approved by their authors' IRBs, and 8 [34,36-38,48,50,55,61] reported receiving informed consent from participants prior to data analysis. For public data collected from crowdsourcing platforms, participants who opted in provided their consent to data sharing [67]. For myPersonality data, Liu et al [34] stated that the dataset itself had IRB approval, so the authors did not report obtaining any further approval from their institution. Youyou et al [83] also concluded that no IRB approval was needed for using myPersonality data. Chancellor et al [29] did not seek IRB approval, because their study used Instagram data without personally identifiable information.

Researchers in 6 studies [31,33,35,52,53,70] reported that the social network datasets collected from participants were anonymized. O'Dea et al [33] removed names, user identifiers, and user identities, and the data collected had to be analyzed after 3 months. Names and usernames in tweets were removed or replaced with other text in 3 studies [31,52,53]. Jamison-Powell et al [70] reported that they removed user identifiers from tweets illustrated in their published article.

The performance of these models is still fuzzy and unstable. As a consequence, none of these studies presented the model's predicted output to participants themselves. Schwartz et al [28] also noted that mental health predictive models are still under development and not sufficiently accurate to be used in practice, and little research has been done on user acceptability of such tools.

Discussion

Principal Findings

The purpose of this review was to investigate the state of development of research on machine learning techniques predicting mental health from social network data. This review also focused on identifying gaps in research and potential applications to detect users with mental health problems. Despite the thousands of articles collected through our search terms, the results of our review suggest that there is a relatively small but growing number of studies using machine learning models to predict mental health problems from social network data. From the initial set of matched articles, only 48 met our inclusion criteria and were selected for review. Some of the excluded studies focused on analysis of the effects of social media use on mental health and well-being states of individual users, and the influence of cyberbullying in social networks on other users.

What Were the Most Surprising Findings?

From the above results, we observed that the same methods could be adapted to analyze posts in different languages. For example, Tsugawa et al [38] adapted De Choudhury's methods [67], originally designed for the analysis of English textual content, to Japanese textual content. Both of them achieved similar results, although some outcomes were dissimilar due to differences in contexts and cultures. This example illustrates that same methods can be used to facilitate studies in different languages.

Several sites were used as sources of data. Facebook is possibly the most popular social network platform. However, only a few studies relied on Facebook datasets to predict mental disorders. One reason for this might be that, by default, users on this site do not make their profiles publicly accessible. Another reason is that getting data from Facebook requires consent from users.

From the selected studies, we can acknowledge several benefits and drawbacks of the methods used in the experiments.

Data Collection

Twitter was a popular source of social network data in the surveyed articles. It provides two different ways of accessing the data: retrospective (using their search APIs) and prospective (via their streaming APIs). Retrospective access allows a regular expression search on the full set of historical tweets, while prospective access allows a search to be set to capture all matching tweets going forward. However, the prospective search grants access to a sample of only 1% of all real-time public tweets based on specific filters. Twitter provides an alternative resource, Firehose, which can provide a standing search over all public tweets, as used in some studies [65-68], but it is only accessible through paid subscription [84].

There are some important differences between studies conducted on Facebook and those using microblogging platforms like Twitter or Sina Weibo. Facebook does not allow developers to access interactions and friendships between users. In addition, users must provide explicit consent to allow an app to pool their data. As a consequence of this, no previous research has used social network analysis to measure and predict mental health problems from Facebook data. On the other hand, microblogging sites grant access to such data. These sites provide APIs that allow developers to get information about followers and followees, and to construct social network graphs of interacting users.

In terms of data collection from users, there are some differences between obtaining data through participants' consent and using regular expression to search for relevant posts. The former option can provide us the real results of the prevalence of mental disorders from participants. The latter approach reduces the time and cost of identifying users with mental illness [59].

Feature Extraction Techniques

The LIWC tool is mostly used for text analysis in psychological research. It extracts many category features, such as style words, emotional words, and parts of speech, from textual contents. It is relatively easy to use and does not require programming skills. Users can just select and open a file or a set of files and LIWC will extract the relevant features and values of each feature. However, there are some disadvantages too. First, LIWC is a proprietary software and users have to purchase a license to use it. Second, the feature database of the tool is not easy to modify. To do this, researchers might need programming skills.

To overcome these shortcomings, there are alternative tools to extract features. However, these tools are rather limited in that they can extract only some features. WordNet is a large English lexicon that can be used to extract parts of speech from text and find semantic meanings of words [85]. SentiStrength assesses the polarity between positive and negative words and the levels of strength of positive and negative words in a textual message [80]. OpinionFinder performs subjectivity, objectivity, and sentiment analysis [79]. Mallet is a useful natural language processing tool to classify or cluster documents, create topics, and perform sequence labelling [86]. Latent Dirichlet allocation is a useful and powerful technique to create topic models. Latent Dirichlet allocation analyzes latent topics, based on word distribution, and then assigns a topic to each document [87]. Each word from an assigned text can be tagged with parts of speech by Part-Of-Speech Tagger [88].

What Can Be Done to Improve the Area?

The selected articles were largely focused on depression, at around 46% (22/48), while 17% (8/48) focused on suicide. Nearly 15% (7/48) of the articles reported a study of well-being and happiness. The rest of the articles investigated postpartum depression, eating disorder, and PTSD. Worthy of note, there is a lack of models for detection of chronic stress and anxiety disorders. Only 1 study in our sample built a stress state detection model [57]. Therefore, this is likely to become an interesting avenue for future research. If a user has a long period of chronic stress, he or she might be becoming depressed. For instance, Hammen [89] reported that chronic stress is a symptomatic source of depression and can develop into other disorders.

Furthermore, based on the selected articles, no study used social network data from actual patients with mental illnesses clinically identified by a doctor or psychologist. Most of the studies included in our review assessed mental disorders with surveys, which are open to self-identifying biases. It would be interesting to promote a closer collaboration between computer scientists and doctors or psychologists, who could provide access to patients with a diagnosis of mental disorders. This might improve the accuracy and reliability of data, making it possible to build predictive models based on features extracted from real patients' social networks. However, mental health conditions might only be formally diagnosed in a specific subset of patients with those conditions, which may lead to a different type of bias.

Importantly, this area of research can benefit from the adoption of open science standards [90]. Many of the studies we reviewed were based on an analysis of openly available data from social networking sites or from the myPersonality project. Parts of the materials used in some of the studies are posted online (eg, [28,32]) or available upon request (eg, [61]). However, most of the studies did not share their entire computational workflow, including not only the datasets, but also the specific code used to preprocess and analyze them. Therefore, future studies should comply with the Transparency and Openness Promotion guidelines [91] at level 2 (which requires authors to deposit data and code in trusted repositories) or 3 (which also requires the reported analyses to be reproduced independently before publication). Of course, to avoid the dissemination of sensitive personal information, the datasets should be properly deidentified when necessary.

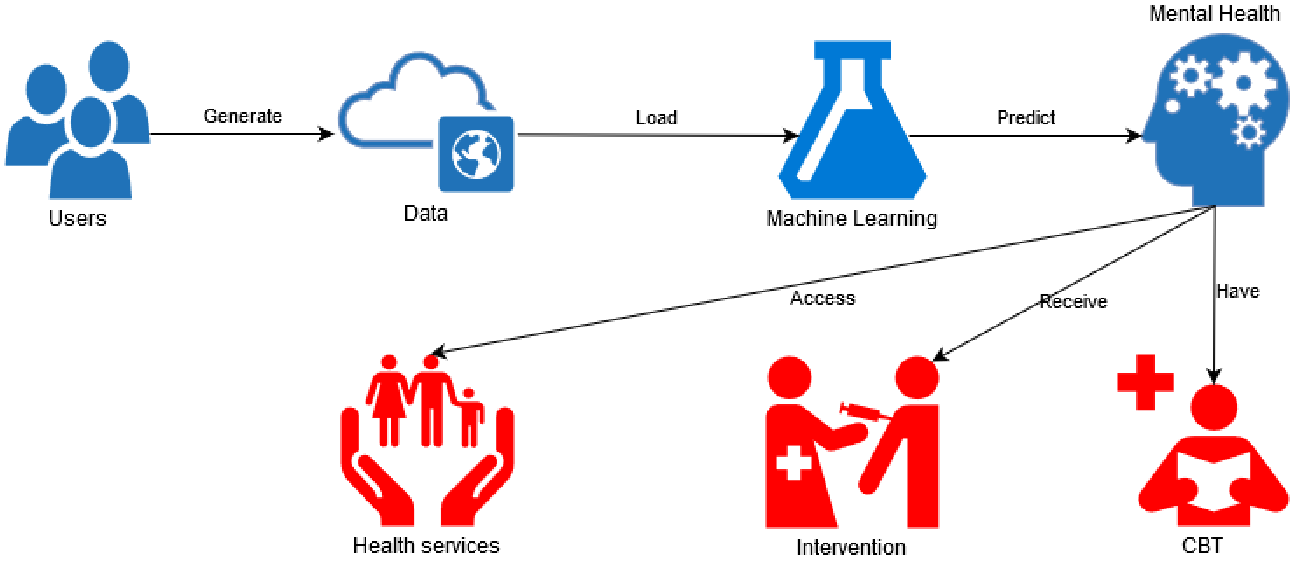
What Are the Novel Trends?

The next generation of predictive models will include more technical analyses. Most of the selected studies relied on textual analysis. But apart from text mining techniques, other methods can be used to gain insights into mental disorders in collected datasets. For instance, image analysis can be used to extract meaningful features from images posted by users. Users facing mental disorders may post images with specific color filters or contents. Among our reviewed studies, 2 found a significant relationship between emotions and color use [92,93]. Another interesting technique is social network analysis. In this review, we selected 3 studies that used social network analysis to examine mental health. However, only 2 studies analyzed symptoms of mental disorders through social network analysis [24,63], while 1 study explored well-being [71]. One study reported that symptoms of depression can be observed through social networks. In other words, depression can be detected through each person’s friends [94]. These examples show that social network analysis is a promising tool to investigate the prevalence of mental illness among online users.

A wide range of machine learning algorithms were used in the reviewed studies. Only 1 study used deep learning algorithms to build a classifier [57], with the rest relying on SVMs, regression models, and decision trees to build classification models. It is expected that, with the rise in popularity of deep learning techniques, this will be changing soon. However, deep learning models are a black box, as opposed to human-interpretable models, such as regression and decision trees, raising the issue of whether it is possible, or indeed necessary, to have these algorithms validated by clinical experts [95].

As this review showed, it is now possible to detect social network users with mental health problems. However, a supporting methodology needs to be developed to translate this innovation into practice and provide help to individuals. Thus, mechanisms are needed to integrate the data science efforts with digital interventions on social network platforms, such as promoting access to health services, offering real-time interventions [96], delivering useful health information links, and conducting cognitive behavioral therapy [97] (see Figure 2).

Figure 2. Conceptual view of social network-based mental health research. CBT: cognitive behavioral therapy.



Ethical Concerns

Several studies outside the scope of this review are particularly useful in highlighting the importance of ethical issues in this area of research. For instance, researchers from Facebook and Cornell University [98] collected and used datasets from Facebook, without offering the possibility to opt out. According to the US Federal Policy for the Protection of Human Subjects (‘Common Rule’), all studies conducted in the United States are required to offer an opt-out for participants. However, private companies do not fall under this rule [99]. This study was not approved by the Cornell University IRB either, “[b]ecause this experiment was conducted by Facebook, Inc. for internal purposes, the Cornell University IRB determined that the project did not fall under Cornell’s Human Research Protection Program” [99].

Another study collected public Facebook posts and made the dataset publicly available to other researchers on the Internet [100]. The posts were manually collected by accessing authors’ friends’ profiles, and anonymizing them. But even so, the posts could still be easily identified [101].

As a result of privacy issues in research with human subjects, the Association of Internet Researchers and other authors have proposed not only ethical questions to evaluate the ethical implications of a research project before starting, but also specific guidelines to eliminate and deal with these issues [102,103].

Surprisingly, few of the studies focused on ethical issues. Conway [104] provided a taxonomy of ethical concepts to bear in mind when using Twitter data for public health studies. Conway [104] and McKee [105] reviewed and presented normative rules for using public Twitter data, including paraphrasing collected posts, receiving informed consent from participants, hiding a participant’s identity, and protecting collected data. Some ethical issues, including context sensitivity, complication of ethics and methodology, and legitimacy requirements, were explicitly addressed by Vayena et al [106].

Mikal et al [107] focused on the perspectives of participants in using social media for population health monitoring. The authors reported that most research participants agreed to have their public posts used for health monitoring, with anonymized data, although they also thought that informed consent would be necessary in some cases.

One approach to reducing the ethical issues of accessing to and using personal information in this area of research is to anonymize the collected datasets to prevent the identification of participants. Wilkinson et al [103] suggested that researchers should not directly quote messages or the public URLs of messages in publication, because these can be used to identify content creators. Sula [108] provided strategies to deal with research in social media including involving participants in studies (not just collect public contents), not collect personally identifiable information (eg, social network profile names), provide participants with a chance to opt out, and make resulting research findings easily accessible and understandable to participants. In most localities, doing any research that collects private information (including social networking posts) from human participants is required to provide project information to IRBs or ethics committees to obtain approval prior to data collection [102,109].

Related Work

This review focused on studies building predictive machine learning models to automatically detect mental health conditions from social network data. Some studies linking mental health and other sources of data did not meet our selection criteria but provide interesting insights about research trends in this area. For instance, previous research has tried to predict mental health conditions or suicidal risk from alternative sources of data such as clinical notes [110], voice analysis [111,112], face analysis [113], and multimodal analysis [114]. We excluded other studies from this review because they used social media data to predict different outcomes; for example, Hanson et al [115] used Twitter data to predict drug abuse. Additionally, recent work has investigated reasons behind Twitter users posting about their mental health [116].

Conclusion

The purpose of this review was to provide an overview of the state-of-the-art in research on machine learning techniques predicting mental health from social network data. Most of the selected studies approached this problem using text analysis. However, some studies also relied on image analysis and social network analysis to gain insights into mental health problems from social network datasets. Predictive models and binary classifiers can be trained based on features obtained from all these techniques. Based on our selected articles, there were relatively few studies applying predictive machine learning models to detect users with mental disorders in real social networks. Moving forward, this research can help in designing and validating new classification models for detecting social network users with mental illnesses and recommend a suitable individually tailored intervention. These interventions might be delivered in the form of advertisements, information links, online advice, or cognitive behavioral therapy; for example, Facebook is considering offering users deemed at risk of suicide online help in real time [117]. However, the reliability of the provided social network data and the general desirability of such interventions should be carefully studied with the users.

With advances in smart data capture devices, such as mobile phones, smart watches, and fitness accessories, future research could combine physical symptoms, such as movements, heart signs, or sleep patterns, with online social network activity to improve the accuracy and reliability of predictions. Finally, scholars interested in conducting research in this area should pay particular attention to the ethical issues of research with human subjects and data privacy in social media, as these are still not fully understood by ethics boards and the wider public.

Acknowledgments

AW is fully funded by scholarships from the Royal Thai Government to study for a PhD, and MAV was supported by grant 2016-T1/SOC-1395 from Madrid Science Foundation. This research was supported by the UK National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the National Health Service, the NIHR, or the Department of Health. Open access for this article was funded by King's College London. We would like to thank Dr Elizabeth Ford for useful comments while finalizing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summaries of methodologies of the selected articles.

[PDF File (Adobe PDF File), 179KB - jmir_v19i6e228_app1.pdf]

References

1. Marcus M, Yasamy M, van Ommeren OM, Chisholm D. Depression: a global public health concern. Geneva, Switzerland: World Health Organization, Department of Mental Health and Substance Abuse; 2012. URL: http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf [accessed 2016-08-24] [WebCite Cache ID 6j4f0tPeN]
2. McManus S, Meltzer H, Brugha T, Bebbington P, Jenkins R. Adult Psychiatric Morbidity in England: Results of a Household Survey. Leeds, UK: NHS Information Centre for Health and Social Care; 2007.
3. Bloom D, Cafiero E, Jané-Llopis E, Abrahams-Gessel S, Bloom R, Fathima S, et al. The global economic burden of noncommunicable diseases.: World Economic Forum; 2011. URL: http://www3.weforum.org/docs/

WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf [accessed 2012-11-26] [WebCite Cache ID 6CSThUnbF]

4. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967 Dec;6(4):278-296. [Medline: 6080235]
5. Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965 Jan 01;12(1):63. [doi: 10.1001/archpsyc.1965.01720310065008]
6. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
7. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
8. Barbier G, Liu H. Data mining in social media. In: Aggarwal CC, editor. Social Network Data Analytics. Volume 6, number 7. New York, NY: Springer; 2011:327-352.
9. Facebook. Facebook reports second quarter 2016 results. 2016. URL: https://investor.fb.com/investor-news/press-release-details/2016/Facebook-Reports-Second-Quarter-2016-Results/default.aspx [accessed 2016-12-21] [WebCite Cache ID 6mufUE0bW]
10. Twitter. Twitter usage and company facts. 2016. URL: https://about.twitter.com/company [accessed 2016-12-21] [WebCite Cache ID 6mugoNvWf]
11. Khadjeh Nassirtoussi A, Aghabozorgi S, Ying Wah T, Ngo DCL. Text mining for market prediction: a systematic review. Expert Syst Appl 2014 Nov;41(16):7653-7670. [doi: 10.1016/j.eswa.2014.06.009]
12. Yu Q, Miche Y, Séverin E, Lendasse A. Bankruptcy prediction using extreme learning machine and financial expertise. Neurocomputing 2014 Mar;128:296-302. [doi: 10.1016/j.neucom.2013.01.063]
13. Jungherr A. Twitter use in election campaigns: a systematic literature review. J Inf Technol Polit 2015 Dec 21;13(1):72-91. [doi: 10.1080/19331681.2015.1132401]
14. Wang T, Rudin C, Wagner D, Sevieri R. Learning to detect patterns of crime. In: Blockeel H, Kersting K, Nijssen S, Zelezny F, editors. Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2013. Lecture Notes in Computer Science. Volume 8190. Berlin, Germany: Springer; 2013:515-530.
15. Kaur H, Wasan SK. Empirical study on applications of data mining techniques in healthcare. J Comput Sci 2006 Feb 1;2(2):194-200. [doi: 10.3844/jcssp.2006.194.200]
16. Cortés R, Bonnaire X, Marin O, Sens P. Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective. Procedia Comput Sci 2015;52:1004-1009. [doi: 10.1016/j.procs.2015.05.093]
17. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 Jun;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
18. Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. J Big Data 2014;1(1):2. [doi: 10.1186/2196-1115-1-2]
19. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011 Jun;28(6):447-455 [FREE Full text] [doi: 10.1002/da.20805] [Medline: 21400639]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
21. National Institute for Health and Care Excellence. Common mental health problems: identification and pathways to care. 2011 May. URL: https://www.nice.org.uk/guidance/cg123 [accessed 2017-03-13] [WebCite Cache ID 6ovoAde2H]
22. Penn Positive Psychology Center. World Well-Being Project. Philadelphia, PA: Penn Positive Psychology Center; 2017. URL: http://www.wwbp.org/ [accessed 2017-03-05] [WebCite Cache ID 6okLXu9c1]
23. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2015 Feb 19:103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
24. Wang T, Brede M, Ianni A, Mentzakis E. Detecting and characterizing eating-disorder communities on social media. 2017 Presented at: Tenth ACM International Conference on Web Search and Data Mining - WSDM ’17; Feb 6-10, 2017; Cambridge, UK p. 91-100. [doi: 10.1145/3018661.3018706]
25. Volkova S, Han K, Corley C. Using social media to measure student wellbeing: a large-scale study of emotional response in academic discourse. In: Spiro E, Ahn YY, editors. Social Informatics. SocInfo 2016. Lecture Notes in Computer Science. Volume 10046. Cham, Switzerland: Springer; 2016:510-526.
26. Saravia E, Chang C, De Lorenzo RJ, Chen Y. MIDAS: mental illness detection and analysis via social media. 2016 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); Aug 18-21, 2016; San Francisco, CA, USA p. 1418-1421. [doi: 10.1109/ASONAM.2016.7752434]
27. Kang K, Yoon C, Kim E. Identifying depressive users in Twitter using multimodal analysis. 2016 Presented at: International Conference on Big Data and Smart Computing (BigComp); Jan 18-20, 2016; Hong Kong, China p. 18-20. [doi: 10.1109/BIGCOMP.2016.7425918]

28. Schwartz HA, Sap M, Kern ML, Eichstaedt JC, Kapelner A, Agrawal M, et al. Predicting individual well-being through the language of social media. Pac Symp Biocomput 2016;21:516-527 [FREE Full text] [Medline: 26776214]
29. Chancellor S, Lin Z, Goodman E, Zerwas S, De Choudhury CM. Quantifying and predicting mental illness severity in online pro-eating disorder communities. 2016 Presented at: 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; Feb 27-Mar 2, 2016; San Francisco, CA, US p. 1171-1184. [doi: 10.1145/2818048.2819973]
30. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for Twitter data against established measures of suicidality. JMIR Ment Health 2016 May 16;3(2):e21 [FREE Full text] [doi: 10.2196/mental.4822] [Medline: 27185366]
31. Coppersmith G, Ngo K, Leary R, Wood A. Exploratory analysis of social media prior to a suicide attempt. 2016 Presented at: 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 16, 2016; San Diego, CA, USA p. 106-117.
32. Lv M, Li A, Liu T, Zhu T. Creating a Chinese suicide dictionary for identifying suicide risk on social media. PeerJ 2015;3:e1455 [FREE Full text] [doi: 10.7717/peerj.1455] [Medline: 26713232]
33. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]
34. Liu P, Tov W, Kosinski M, Stillwell DJ, Qiu L. Do Facebook status updates reflect subjective well-being? Cyberpsychol Behav Soc Netw 2015 Jul;18(7):373-379. [doi: 10.1089/cyber.2015.0022] [Medline: 26167835]
35. Burnap P, Colombo W, Scourfield J. Machine classification and analysis of suicide-related communication on Twitter. 2015 Presented at: 26th ACM Conference on Hypertext & Social Media; Sep 1-4, 2015; Guzelyurt, Northern Cyprus p. 75-84. [doi: 10.1145/2700171.2791023]
36. Park S, Kim I, Lee S, Yoo J, Jeong B, Cha M. Manifestation of depression and loneliness on social networks. 2015 Presented at: 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW ’15; Mar 14-18, 2015; Vancouver, BC, Canada p. 14-18. [doi: 10.1145/2675133.2675139]
37. Hu Q, Li A, Heng F, Li J, Zhu T. Predicting depression of social media user on different observation windows. 2015 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology; Dec 6-9, 2015; Singapore p. 361-364. [doi: 10.1109/WI-IAT.2015.166]
38. Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H. Recognizing depression from Twitter activity. 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; Apr 18-23, 2015; Seoul, Republic of Korea p. 3187-3196. [doi: 10.1145/2702123.2702280]
39. Zhang L, Huang X, Liu T, Li A, Chen Z, Zhu T. Using linguistic features to estimate suicide probability of Chinese microblog users. In: Zu Q, Hu B, Gu N, Seng S, editors. Human Centered Computing. HCC 2014. Lecture Notes in Computer Science. Volume 8944. Cham, Switzerland: Springer; 2015:549-559.
40. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: analyzing the language of mental health on Twitter through self-reported diagnoses. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 1-10.
41. Preotiuc-Pietro D, Sap M, Schwartz H, Ungar L. Mental illness detection at the World Well-Being Project for the CLPsych 2015 shared task. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 40-45.
42. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 11-20.
43. Preotiuc-Pietro D, Eichstaedt J, Park G, Sap M, Smith L, Tobolsky V, et al. The role of personality, age, and gender in tweeting about mental illness. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 21-30. [doi: 10.3115/v1/W15-1203]
44. Pedersen T. Screening Twitter users for depression and PTSD with lexical decision lists. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 46-53.
45. Resnik P, Armstrong W, Claudino L, Nguyen T, Nguyen VA, Boyd-Graber J. Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 99-107.
46. Resnik P, Armstrong W, Claudino L, Nguyen T, Nguyen V, Boyd-Graber J. The University of Maryland CLPsych 2015 shared task system. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 54-60.
47. Durahim AO, Coşkun M. #iamhappybecause: Gross National Happiness through Twitter analysis and big data. Technol Forecast Soc Change 2015 Oct;99:92-105. [doi: 10.1016/j.techfore.2015.06.035]
48. Guan L, Hao B, Cheng Q, Yip PS, Zhu T. Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. JMIR Ment Health 2015;2(2):e17 [FREE Full text] [doi: 10.2196/mental.4227] [Medline: 26543921]

49. Landeiro Dos Reis V, Culotta A. Using matched samples to estimate the effects of exercise on mental health from Twitter. 2015 Presented at: Twenty-Ninth AAAI Conference on Artificial Intelligence; Jan 25-30, 2015; Austin, TX, USA p. 182-188.
50. De Choudhury M, Counts S, Horvitz E, Hoff A. Characterizing and predicting postpartum depression from shared Facebook data. 2014 Presented at: 17th ACM conference on Computer supported cooperative work & social computing - CSCW ’14; Feb 15-19, 2014; Baltimore, MD, USA p. 626-638. [doi: 10.1145/2531602.2531675]
51. Huang X, Zhang L, Chiu D, Liu T, Li X, Zhu T. Detecting suicidal ideation in Chinese microblogs with psychological lexicons. 2014 Presented at: 2014 IEEE International Conference on Scalable Computing and Communications and its Associated Workshops; Dec 9-12, 2014; Bali, Indonesia p. 844-849. [doi: 10.1109/UIC-ATC-ScalCom.2014.48]
52. Wilson M, Ali S, Valstar M. Finding information about mental health in microblogging platforms. 2014 Presented at: 5th Information Interaction in Context Symposium; Aug 26-30, 2014; Regensburg, Germany p. 8-17. [doi: 10.1145/2637002.2637006]
53. Coppersmith G, Harman C, Dredze M. Measuring post traumatic stress disorder in Twitter. 2014 Presented at: 8th International AAAI Conference on Weblogs and Social Media; Jun 2-4, 2014; Ann Arbor, MI, USA p. 23-45.
54. Kuang C, Liu Z, Sun M, Yu F, Ma P. Quantifying Chinese happiness via large-scale microblogging data. 2014 Presented at: Web Information System and Application Conference; Sep 12-14, 2014; Tianjin, China p. 227-230. [doi: 10.1109/WISA.2014.48]
55. Hao B, Li L, Gao R, Li A, Zhu T. Sensing subjective well-being from social media. In: Slezak D, Schaefer G, Vuong ST, Kim YS, editors. Active Media Technology. AMT 2014. Lecture Notes in Computer Science. Volume 8610. Cham, Switzerland: Springer; 2014:324-335.
56. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: a good place to detect health conditions. PLoS One 2014;9(1):e86191 [FREE Full text] [doi: 10.1371/journal.pone.0086191] [Medline: 24489699]
57. Lin H, Jia J, Guo Q, Xue Y, Li Q, Huang J, et al. User-level psychological stress detection from social media using deep neural network. 2014 Presented at: 22nd ACM international conference on Multimedia; Nov 3-7, 2014; Orlando, FL, USA p. 507-516. [doi: 10.1145/2647868.2654945]
58. Schwartz HA, Eichstaedt J, Kern ML, Park G, Sap M, Stillwell D, et al. Towards assessing changes in degree of depression through Facebook. 2014 Presented at: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 27, 2014; Baltimore, MD, USA p. 118-125.
59. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. 2014 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 27, 2014; Baltimore, MD, USA p. 51-60. [doi: 10.3115/v1/W14-3207]
60. Homan C, Johar R, Liu T, Lytle M, Silenzio V, Ovesdotter AC. Toward macro-insights for suicide prevention: analyzing fine-grained distress at scale. 2014 Presented at: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 27, 2014; Baltimore, MD, USA p. 107-117.
61. Park S, Lee SW, Kwak J, Cha M, Jeong B. Activities on Facebook reveal the depressive state of users. J Med Internet Res 2013;15(10):e217 [FREE Full text] [doi: 10.2196/jmir.2718] [Medline: 24084314]
62. Wang X, Zhang C, Ji Y, Sun L, Wu L, Bao Z. A Depression detection model based on sentiment analysis in micro-blog social network. In: Li J, Cao L, Wang C, Chen Tan K, Liu B, Pei J, et al, editors. Trends and Applications in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science. Volume 7867. Berlin, Germany: Springer; 2013:201-213.
63. Wang X, Zhang C, Sun L. An improved model for depression detection in micro-blog social network. 2013 Presented at: IEEE 13th International Conference on Data Mining Workshops; Dec 7-10, 2013; Dallas, TX, USA p. 80-87. [doi: 10.1109/ICDMW.2013.132]
64. Tsugawa S, Mogi Y, Kikuchi Y, Kishino F, Fujita K, Itoh Y, et al. On estimating depressive tendencies of Twitter users utilizing their tweet data. 2013 Presented at: IEEE Virtual Reality; Mar 18-20, 2013; Lake Buena Vista, FL, USA. [doi: 10.1109/VR.2013.6549431]
65. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 27-May 2, 2013; Paris, France p. 3267-3276. [doi: 10.1145/2470654.2466447]
66. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. 2013 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
67. De Choudhury M, Gamon M. Predicting depression via social media. 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media; Jul 8-11, 2013; Cambridge, MA, USA p. 128-137.
68. Schwartz H, Eichstaedt J, Kern M, Dziurzynski L, Lucas R, Agrawal M, et al. Characterizing geographic variation in well-being using tweets. 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media; Jul 8-11, 2013; Cambridge, MA, USA.
69. Hao B, Li L, Li A, Zhu T. Predicting mental health status on social media. In: Rau PLP, editor. Cross-Cultural Design. Cultural Differences in Everyday Life. CCD 2013. Lecture Notes in Computer Science. Volume 8024. Berlin, Germany: Springer; 2013:101-110.

70. Jamison-Powell S, Linehan C, Daley L, Garbett A, Lawson S. “I can’t get no sleep”: discussing #insomnia on Twitter. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX, USA p. 1501-1510. [doi: 10.1145/2207676.2208612]
71. Bollen J, Gonçalves B, Ruan G, Mao H. Happiness is assortative in online social networks. Artif Life 2011;17(3):237-251. [doi: 10.1162/artl_a_00034] [Medline: 21554117]
72. Seligman MEP. Flourish: A Visionary New Understanding of Happiness and Well-being. London, UK: Free Press; 2011.
73. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 shared task depression and PTSD on Twitter. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 5, 2015; Denver, CO, USA p. 31-39.
74. Kittur A, Chi E, Suh B. Crowdsourcing user studies with Mechanical Turk. 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems - CHI ’08; Apr 5-10, 2008; Florence, Italy p. 453-456. [doi: 10.1145/1357054.1357127]
75. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]
76. Chaoji V, Hoonlor A, Szymanski B. Recursive data mining for role identification. 2008 Presented at: 5th international conference on Soft computing as transdisciplinary science and technology - CSTST ’08; Oct 28-31, 2008; Cergy-Pontoise, France p. 217-225. [doi: 10.1145/1456223.1456270]
77. Kiritchenko S, Zhu X, Mohammad S. Sentiment analysis of short informal texts. J Artif Intell Res 2014;50:723-762. [doi: 10.1613/jair.4272]
78. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
79. Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, et al. OpinionFinder: a system for subjectivity analysis. 2005 Presented at: HLT/EMNLP on Interactive Demonstrations; Oct 7, 2005; Vancouver, BC, Canada p. 34-35. [doi: 10.3115/1225733.1225751]
80. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. J Assoc Inform Sci Technol 2010 Dec 15;61(12):2544-2558. [doi: 10.1002/asi.21416]
81. Bradley M, Lang P. Affective Norms for English Words (ANEW): instruction manual and Affective Ratings. Technical Report C-. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida; 1999. URL: http://www.uvm.edu/pdodds/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf [accessed 2016-08-29] [WebCite Cache ID 6r1Z3AUtu]
82. Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning. Cambridge, Massachusetts: MIT Press; 2012.
83. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. Proc Natl Acad Sci U S A 2015 Jan 27;112(4):1036-1040 [FREE Full text] [doi: 10.1073/pnas.1418680112] [Medline: 25583507]
84. Morstatter F, Pfeffer J, Liu H, Carley K. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. Palo Alto, CA: AAAI Press; 2013 Presented at: 7th International AAAI Conference on Weblogs and Social Media. ICWSM 2013; July 8-11, 2013; Boston, MA, USA p. 400-408 URL: https://arxiv.org/pdf/1306.5204.pdf
85. Miller GA. WordNet: a lexical database for English. Commun ACM 1995 Nov;38(11):39-41. [doi: 10.1145/219717.219748]
86. McCallum AK. MALLET: a machine learning for language toolkit. 2002. URL: https://people.cs.umass.edu/~mccallum/mallet/ [accessed 2017-06-06] [WebCite Cache ID 6r1ZjX4aR]
87. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Machine Learning Res 2003;3(Jan):993-1022.
88. Toutanova K, Klein D, Manning C, Singer Y. Feature-rich part-of-speech tagging with a cyclic dependency network. 2003 Presented at: 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL ’03; May 27-Jun 1, 2003; Edmonton, AB, Canada p. 173-180. [doi: 10.3115/1073445.1073478]
89. Hammen C. Stress and depression. Annu Rev Clin Psychol 2005;1:293-319. [doi: 10.1146/annurev.clinpsy.1.102803.143938] [Medline: 17716090]
90. Stodden V, McNutt M, Bailey DH, Deelman E, Gil Y, Hanson B, et al. Enhancing reproducibility for computational methods. Science 2016 Dec 09;354(6317):1240-1241. [doi: 10.1126/science.aah6168] [Medline: 27940837]
91. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture. Science 2015 Jun 26;348(6242):1422-1425 [FREE Full text] [doi: 10.1126/science.aab2374] [Medline: 26113702]
92. Valdez P, Mehrabian A. Effects of color on emotions. J Exp Psychol Gen 1994 Dec;123(4):394-409. [Medline: 7996122]
93. Nolan RF, Dai Y, Stanley PD. An investigation of the relationship between color choice and depression measured by the Beck Depression Inventory. Percept Mot Skills 1995 Dec;81(3 Pt 2):1195-1200. [doi: 10.2466/pms.1995.81.3f.1195] [Medline: 8684914]
94. Rosenquist JN, Fowler JH, Christakis NA. Social network determinants of depression. Mol Psychiatry 2011 Mar;16(3):273-281 [FREE Full text] [doi: 10.1038/mp.2010.13] [Medline: 20231839]
95. Castelvecchi D. Can we open the black box of AI? Nature 2016 Dec 06;538(7623):20-23. [doi: 10.1038/538020a] [Medline: 27708329]

96. Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. J Med Internet Res 2015;17(6):e141 [FREE Full text] [doi: 10.2196/jmir.3662] [Medline: 26068087]
97. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
98. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
99. Verma IM. Editorial expression of concern: experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jul 22;111(29):10779 [FREE Full text] [doi: 10.1073/pnas.1412469111] [Medline: 24994898]
100. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N. Tastes, ties, and time: a new social network dataset using Facebook.com. Social Networks 2008 Oct;30(4):330-342. [doi: 10.1016/j.socnet.2008.07.002]
101. Zimmer M. “But the data is already public”: on the ethics of research in Facebook. Ethics Inf Technol 2010 Jun 4;12(4):313-325. [doi: 10.1007/s10676-010-9227-5]
102. Markham A, Buchanan E. Ethical decision-making and internet research: recommendations from the AoIR Ethics Working Committee (version 2.0).: The Association of Internet Researchers; 2012. URL: http://aoir.org/reports/ethics2.pdf [accessed 2016-09-23] [WebCite Cache ID 6hhZbLexD]
103. Wilkinson D, Thelwall M. Researching personal information on the public web: methods and ethics. Soc Sci Comput Rev 2010 Aug 17;29(4):387-401. [doi: 10.1177/0894439310378979]
104. Conway M. Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. J Med Internet Res 2014;16(12):e290 [FREE Full text] [doi: 10.2196/jmir.3617] [Medline: 25533619]
105. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
106. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
107. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
108. Sula CA. Research ethics in an age of big data. Bull Assoc Inf Sci Technol 2016 Feb 01;42(2):17-21. [doi: 10.1002/bul2.2016.1720420207]
109. Working Party on Ethical Guidelines for Psychological Research. Code of human research ethics. Leicester, UK: The British Psychological Society; 2010. URL: http://www.bps.org.uk/sites/default/files/documents/code_of_human_research_ethics.pdf [accessed 2016-11-23] [WebCite Cache ID 6r1c032Xg]
110. Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B, et al. Predicting the risk of suicide by analyzing the text of clinical notes. PLoS ONE 2014 Jan 28;9(3):e85733. [doi: 10.1371/journal.pone.0085733] [Medline: 24489669]
111. Roark B, Mitchell M, Hosom J, Hollingshead K, Kaye J. Spoken language derived measures for detecting mild cognitive impairment. IEEE Trans Audio Speech Lang Process 2011 Sep 01;19(7):2081-2090 [FREE Full text] [doi: 10.1109/TASL.2011.2112351] [Medline: 22199464]
112. Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 2014 Jun;55:43-60. [doi: 10.1016/j.cortex.2012.12.006] [Medline: 23332818]
113. Ooi K, Low L, Lech M, Allen N. Prediction of clinical depression in adolescents using facial image analysis. 2011 Presented at: WIAMIS 2011: 12th International Workshop on Image Analysis for Multimedia Interactive Services; Apr 13-15, 2011; Delft, the Netherlands.
114. Gupta R, Malandrakis N, Xiao B, Guha T, Van SM, Black M, et al. Multimodal prediction of affective dimensionsand depression in human-computer interactions. 2014 Presented at: 4th International Workshop on Audio/Visual Emotion Challenge - AVEC ’14; Nov 7, 2014; Orlando, FL, USA p. 33-40. [doi: 10.1145/2661806.2661810]
115. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
116. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Apr 05;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
117. Callison-Burch V, Guadagno J, Davis A. Facebook. 2017. Building a safer community with new suicide prevention tools URL: http://newsroom.fb.com/news/2017/03/building-a-safer-community-with-new-suicide-prevention-tools/ [accessed 2017-03-14] [WebCite Cache ID 6oxEPez41]

Abbreviations

API: application programming interface
CLPsych: Computational Linguistics and Clinical Psychology Workshops
IRB: institutional review board
LIWC: linguistic inquiry and word count.
MeSH: Medical Subject Headings
OCD: obsessive-compulsive disorder
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
SVM: support vector machine

Edited by I Weber; submitted 22.12.16; peer-reviewed by M De Choudhury, G Coppersmith, C Giraud-Carrier; comments to author 06.02.17; revised version received 14.03.17; accepted 27.04.17; published 29.06.17.

Please cite as:
Wongkoblap A, Vadillo MA, Curcin V
Researching Mental Health Disorders in the Era of Social Media: Systematic Review
J Med Internet Res 2017;19(6):e228
URL: http://www.jmir.org/2017/6/e228/
doi:10.2196/jmir.7215
PMID:28663166

Original Paper

Access to Electronic Personal Health Records Among Patients With Multiple Chronic Conditions: A Secondary Data Analysis

Alexandra J Greenberg[1], MPH, PhD; Angela L Falisi[2], MPH; Lila J Finney Rutten[1], MPH, PhD; Wen-Ying Sylvia Chou[2], MPH, PhD; Vaishali Patel[3], MPH, PhD; Richard P Moser[2], PhD; Bradford W Hesse[2], PhD

[1]Mayo Clinic, Rochester, MN, United States
[2]Behavioral Research Program, National Cancer Institute, Bethesda, MD, United States
[3]Office of the National Coordinator, US Department of Health and Human Services, Washington, DC, United States

Corresponding Author:
Alexandra J Greenberg, MPH, PhD
Mayo Clinic
200 First St. SW
Rochester, MN
United States
Phone: 1 507 538 7388
Fax: 1 507 555 1212
Email: worisek.alexandra@gmail.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/6/e39719

Abstract

Background: In the United States, national incentives for offering access to electronic personal health records (ePHRs) through electronic means are geared toward creating a culture of patient engagement. One group of patients who stand to benefit from online access to ePHRs is the growing population with multiple chronic conditions (MCC). However, little is known about the current availability and use of ePHRs and patient portals among those managing MCC.

Objective: The aim was to determine the associations between number of chronic conditions and sociodemographic characteristics and usage of ePHRs, and to assess how the public's use of ePHRs varies across subpopulations, including those with MCC.

Methods: This study used data collected from the 2014 Health Information National Trends Survey (HINTS), and assessed differences in use of ePHRs between those with and without MCC (N=3497) using multiple logistic regression techniques. Variables associated with health care systems (insurance status, having a regular provider) and patient-reported self-efficacy were included in the statistical models.

Results: Those with MCC (n=1555) had significantly higher odds of accessing their records three or more times in the past year compared to those reporting no chronic conditions (n=1050; OR 2.46, 95% CI 1.37-4.45), but the overall percentage of those with MCC using ePHRs remained low (371 of 1529 item respondents, 25.63% weighted). No difference in odds of accessing their records was found between those reporting one chronic condition (n=892) and those reporting none (n=1050; OR 1.02, 95% CI 0.66-1.58). Significant differences in odds of accessing ePHRs were seen between income and age groups ($P<.001$ and $P=.05$, respectively), and by whether respondents had a regular provider ($P=.03$).

Conclusions: We conclude that ePHRs provide a unique opportunity to enhance MCC patient self-management, but additional effort is needed to ensure that these patients are able to access their ePHRs. An increase in availability of patient access to their ePHRs may provide an opportunity to increase patient engagement and support self-management for all patients and especially those with MCC.

(J Med Internet Res 2017;19(6):e188) doi:10.2196/jmir.7417

KEYWORDS

electronic personal health information; electronic health records; patient engagement; multiple chronic conditions

Introduction

Over the next 40 years, the United States is expected to face a significant and rapid growth in its population aged 65 and older due in large part to the aging of the baby boomer generation [1]. With this aging population comes an increase not only in the number of individuals living with chronic disease, but also in the number of individuals living with multiple chronic conditions (MCC) [2]. Currently, 25% of Americans have two or more concurrent chronic conditions; these conditions include both physical conditions (eg, cardiovascular disease, diabetes, and cancer) and psychological conditions (eg, depression and anxiety) [3]. Individuals with two or more chronic conditions account for approximately 93% of Medicare spending; these costs are projected to increase in the coming years [4-7].

The US Department of Health and Human Services emphasizes the importance of patient-facing informatics tools for improving patient self-management and encouraging patient engagement, as evidenced by the creation of the Office of the National Coordinator for Health Information Technology in 2009. Widespread adoption of health information technology tools by the general population has been slow, but it is likely that the demands associated with MCC may encourage adoption in the near future [8]. Among older patients with diabetes, for example, the GroupHealth Cooperative found that the highest rate of use of secure messaging with providers was among those with multiple chronic conditions [9].

Health literacy is also key to using these tools because those with MCC need to have higher levels of health literacy when using electronic personal health records (ePHRs) [10,11]. Patients presenting with MCC have complex treatment regimens, some of which may be in conflict with one another [2]. Care is often fragmented and spread across several different providers and clinics, making coordination of care an additional concern for patients with MCC [2]. Furthermore, the complexity of care for these patients also results in more opportunities for medical errors, often due to missing data and note-taking differences in clinical records [12-15]. Therefore, there is a need for a technologically enabled health care system that can remove the burden from clinical personnel by enabling patients to be more engaged in their own care, simultaneously removing obstacles for timely patient-provider communication and preemptively decreasing the probability of medical error away from the clinic. At the same time, however, care must be taken to ensure that systems are not increasing the burden on the provider's time [16]. From the perspective of the chronic care model, this implies that utilizing these new technologies and their accompanying workflow modifications creates a "prepared health system" [17-19].

At the same time, those diagnosed with MCC face challenges with the self-management of their ongoing care. This may be a result of impairment of physical or cognitive function, complexity of treatment regimens, and poor patient-provider communication [20,21]. In addition, fragmented care delivery poses myriad problems from the patient perspective with regard to managing personal health information (PHI) because it often falls on the patients to bring their records to each of their providers and to ensure that all their medications and therapeutic decisions are communicated accurately across points of care [22-24]. These tasks require a great deal of self-efficacy from patients, which has been linked to clinical outcomes [25-27]. For care of patients with several comorbidities to be successful and to result in improved quality of life, there is a need to have "prepared patients"—that is, patients who are equipped with tools and knowledge that empower them to communicate effectively with their providers, understand and manage their various treatment regimens, feel self-efficacious to take on these tasks, and consolidate their complex health care histories into one place [28,29].

To effectively support care of patients with MCC over time, there is a need to have both a "prepared health care system" and a population of "prepared patients." Two complementary tools for addressing both the concepts of the "prepared system" and of the "prepared patient" are (1) electronic health records (EHRs) on the clinical side, and (2) connected access to that information through electronic personal health records (ePHRs) or patient portals on the patient side [19,30,31]. These ePHRs have the potential to serve as powerful tools for patient self-management, leading to the hypothesis that use of these tools will increase patient activation and self-efficacy and, in turn, improved clinical outcomes [32-34]. Studies have shown that "activated" and engaged patients tend to have lower health care costs and better care experiences than their less "activated" peers [35,36]. ePHRs hold particular promise for those with MCC because they have the potential to allow for care coordination between providers, allow for secure patient-provider communication, appointment management, information consolidation, and prescription refilling [22,37]. However, many individuals with MCCs have multiple providers and may consequently have multiple EHRs, potentially resulting in further confusion [2,25].

Although adoption of ePHRs sets the stage for patient engagement, there are no current data regarding the actual utilization of these tools from the public at large, especially for populations of patients struggling with MCC. To address this evidentiary gap, we analyzed data from the 2014 administration of the Health Information National Trends Survey (HINTS). The HINTS program was initiated in 2001 by the National Cancer Institute to provide surveillance on the public's use of health information in a rapidly changing communication environment [38]. It has become the de facto source of data on many of the health communication and informatics objectives included in the Department of Health and Human Services' Healthy People 2010 and 2020 initiatives [39], and has provided a complementary source of data for clinical audiences on the ways in which the public's use of communication sources may influence practice [40]. In this study, we sought to answer the following questions:

1. What are the associations between number of chronic conditions, sociodemographic characteristics, and usage of ePHRs?
2. How does the public's reported use ePHRs vary across different subpopulations, including populations who report MCC?

3. How likely is it for different groups of adults to have reported that they accessed their ePHRs and, more specifically, does the likelihood of accessing ePHRs increase as the number of chronic conditions increases?

Methods

Study Population and Data Collection

For this study, we analyzed data from the 2014 iteration of the HINTS (HINTS 4, Cycle 4) collected via a self-administered mailed survey between July and November 2014. The probability-based sampling frame used a two-stage design to achieve a nationally representative sample of US adults aged 18 years and older. Briefly, this design first used a stratified sample of residential addresses from the United States Postal Service, from which one adult from each sampled household was randomly selected. Efforts were made to oversample minority populations and those living in central Appalachia. The final number of respondents for this survey cycle was 3677 (response rate=34.4%). All items included in the HINTS administration underwent at least two rounds of cognitive testing through an external testing service (Westat, Rockville, MD, USA) for validation []. Additional information about data collection for HINTS 4, Cycle 4 can be found in the corresponding methodology report [41].

Outcome Measure

The main outcome of interest was use of ePHRs. Respondents were asked: “How many times did you access your PHI online through a secure website or app in the last 12 months?”, with responses categorized as “none,” “1 to 2 times,” “3 to 5 times,” “6 to 9 times,” and “10 or more times.” A dichotomous variable for accessing ePHR was created by dividing respondents into “no/low” use and “medium/high” use. This categorization was based on the fact that no significant difference was found between those who reported never having accessed their record and those who reported accessing it 1 to 2 times (data not shown), and that the median response of those who had accessed their ePHRs at least once was the “3 to 5 times” category. Thus, no/low use consisted of individuals who responded “none” and “1 to 2 times,” and the medium/high use group consisted of individuals who responded “3 to 5 times,” “6 to 9 times,” and “10 or more times.”

Independent Variables

Sociodemographic variables included in the analyses were age, sex, race/ethnicity, education, employment status, and income level. Health care variables included health insurance status and having a regular health care provider. Other independent variables from the survey included in our models were participant access to the Internet or email and self-reported ratings of general health on a five-point scale (ranging from “poor” to “excellent”). Self-efficacy was assessed using an item asking respondents to rate their self-reported ability to take care of their own health on a five-point scale (ranging from “not at all confident” to “very confident”). Response options for these variables were collapsed from five to three [42-45].

Additional items that dealt with use of technology were examined as well. These included items inquiring about whether respondents used mobile phones/tablets and whether they used health-related apps on these devices, whether they had emailed their doctor in the past 12 months, level of confidence that PHI was safe, feelings about control over privacy of ePHRs, whether they had ever withheld information from a provider due to concerns about privacy, and whether they were concerned about security of information when sent electronically between providers.

Chronic conditions were self-reported as part of the survey administration. Two questions were used to assess chronic conditions. The first asked: “Has a doctor or other health professional ever told you that you had any of the following conditions: (1) diabetes or high blood sugar; (2) high blood pressure or hypertension; (3) a heart condition such as heart attack, angina, or congestive heart failure; (4) chronic lung disease, asthma, emphysema, or chronic bronchitis; (5) arthritis or rheumatism; and (6) depression or anxiety disorder?” The second asked: “Have you ever been diagnosed as having cancer?” For each respondent, the number of chronic conditions were totaled; individuals were then categorized as having zero, one, or two or more chronic conditions for analyses. For this study, MCC was defined as having two or more of these conditions.

Data Analyses

Analyses were conducted using SAS-callable SUDAAN 11.0.0 (RTI International, Research Triangle Park, NC, USA) and SAS 9.3 (SAS Institute Inc, Cary, NC, USA), which allowed analyses to account for the complex sampling procedure and to incorporate the jackknife replicate weights used for variance estimation. Descriptive statistics and bivariate analyses using chi-square tests of association were conducted.

Modeling included multivariable logistic regression using the dichotomous ePHR access variable as an outcome and adjusting for both theoretically important and statistically significant sociodemographic and health care characteristics. We regressed ePHR use onto levels of chronic condition, relevant demographics, health insurance / provider status, general health, and self-reported ability to take care of one’s health. Analyses were restricted to those with Internet access or who owned a mobile phone. The final sample weight was used in all analyses to obtain population-level point estimates and model parameters. For each analysis, listwise deletion of subjects was used.

Results

Analyses were restricted to those who responded to the set of questions for self-reporting of chronic conditions. Of the 3497 individuals who responded to items about chronic conditions, 1050 (43.14% weighted) reported having no chronic conditions, 892 (24.85% weighted) reported having one chronic condition, and 1555 (32.01% weighted) reported having two or more chronic conditions.

Associations Between Patient Factors and Number of Chronic Conditions

The first set of analyses explored bivariate relationships between a series of relevant patient characteristics and number of reported chronic conditions. Results from these analyses are presented in Table 1. As expected, there was a significant positive relationship between age and number of chronic conditions reported such that only 60 of 1555 (9.76% weighted) in the 18 to 34 years age range reported having two or more chronic conditions, whereas 274 of 1555 (71.49% weighted) in the 75 years and older range reported two or more conditions (Table 1). Sex, age, race/ethnicity, education, income, having a regular provider, and having health insurance were all significantly associated with number of chronic conditions. In general, having two or more chronic conditions was associated with being older, having health insurance, having a regular provider, being less confident in taking care of themselves, reporting fair to poor health, and being less inclined to use the Internet or to use a mobile phone/tablet (Table 1). No significant difference was found between those who responded to the chronic conditions items and those who did not, save for a difference in self-reported general health (*P*=.02); however, these results are not clinically or contextually significant because of small numbers in individual cells.

Table 1. Associations between patient characteristics, online characteristics, and attitudes with number of chronic conditions (N=3497).

Respondent characteristics	Number of chronic conditions, n (weighted %)[a]			χ^2 (*df*)	*P* value
	0	1	≥2		
Overall	1050 (39.8)	892 (26.0)	1555 (34.1)		
Sex				5.5 (2)	.007
Female	624 (36.2)	543 (27.3)	920 (36.5)		
Male	420 (43.5)	344 (25.2)	605 (31.3)		
Age (years)				59.5 (8)	<.001
18-34	283 (64.6)	117 (25.4)	60 (10.0)		
35-49	322 (48.1)	200 (24.2)	188 (27.8)		
50-64	272 (30.9)	299 (28.9)	556 (40.3)		
65-74	72 (14.9)	149 (27.0)	349 (58.1)		
≥75	28 (6.3)	74 (21.4)	274 (72.2)		
Race/ethnicity				4.1 (8)	.001
Hispanic	185 (41.9)	132 (28.5)	194 (29.6)		
Non-Hispanic White	555 (38.7)	496 (26.1)	844 (35.2)		
Non-Hispanic Black	142 (39.9)	123 (24.0)	252 (36.1)		
Non-Hispanic other	94 (57.4)	56 (22.5)	85 (20.0)		
Missing	74 (28.9)	85 (26.9)	180 (44.2)		
Education				12.1 (6)	<.001
Less than high school	59 (27.5)	65 (27.0)	163 (45.6)		
High school graduate	171 (38.1)	140 (23.0)	323 (38.9)		
Some college	257 (34.9)	282 (27.0)	511 (38.1)		
College graduate	534 (49.6)	377 (26.7)	500 (23.7)		
Income (US$)				8.3 (8)	<.001
<$20,000	163 (31.5)	182 (27.0)	442 (41.5)		
$20,000 to <$35,000	127 (29.1)	116 (22.1)	265 (48.8)		
$35,000 to <$50,000	145 (40.6)	139 (26.7)	220 (32.7)		
$50,000 to <$75,000	180 (38.2)	153 (26.7)	252 (35.0)		
≥$75,000	421 (48.7)	295 (26.5)	349 (24.9)		
Health insurance				9.3 (2)	<.001
Yes	872 (38.3)	768 (25.8)	1397 (35.9)		
No	168 (51.6)	110 (27.2)	130 (21.2)		
Regular provider				50.1 (2)	<.001
Yes	548 (32.4)	612 (25.1)	1256 (42.5)		
No	494 (54.3)	266 (27.7)	268 (18.0)		
Self-reported ability to take care of own health				9.7 (4)	<.001
Completely confident/very confident	787 (42.9)	629 (27.2)	890 (29.8)		
Somewhat confident	224 (34.8)	231 (24.8)	518 (40.4)		
A little confident/not at all confident	36 (25.9)	29 (19.2)	137 (54.9)		
Self-reported general health				52.0 (4)	<.001
Excellent/very good	675 (53.3)	443 (26.4)	427 (20.3)		
Good	301 (31.5)	345 (28.8)	672 (39.7)		
Fair/Poor	69 (17.9)	97 (17.5)	443 (64.7)		

Respondent characteristics	Number of chronic conditions, n (weighted %)[a]			χ^2 (*df*)	*P* value
	0	1	≥2		
Regular Internet use				35.9 (2)	<.001
Yes	923 (42.9)	709 (26.1)	1077 (31.0)		
No	123 (25.3)	173 (25.4)	455 (49.2)		
Accessed EHRs at least once				0.5 (2)	.61
Yes	284 (28.0)	250 (26.7)	371 (25.5)		
No	757 (72.0)	630 (73.3)	1158 (74.5)		
Frequency of EHR access				5.6 (8)	<.001
Never	757 (72.0)	630 (73.3)	1158 (74.5)		
1-2 times	158 (15.4)	124 (14.2)	153 (9.1)		
3-5 times	74 (7.5)	78 (7.4)	101 (7.6)		
6-9 times	24 (2.2)	30 (3.4)	57 (3.7)		
≥10 times	28 (2.9)	18 (1.6)	60 (5.1)		
Use a mobile phone or tablet				36.7 (2)	<.001
Yes	848 (44.9)	610 (26.0)	854 (29.1)		
No	185 (25.7)	256 (26.2)	638 (48.1)		
Use health-related mobile phone/tablet apps				0.7 (2)	.49
Yes	297 (46.0)	204 (24.4)	295 (29.6)		
No	522 (44.2)	388 (27.9)	516 (27.9)		
Exchanged emails with provider(s)				0.3 (2)	.76
Yes	246 (42.0)	206 (25.7)	331 (32.3)		
No	791 (39.6)	662 (26.1)	1179 (34.3)		
Confidence that PHI is safe				2.0 (2)	.11
Very confident	207 (38.9)	178 (23.7)	389 (37.5)		
Somewhat confident	534 (38.3)	473 (27.4)	809 (34.3)		
Not confident	295 (44.4)	221 (24.9)	324 (30.8)		
Control privacy of records				3.3 (4)	.02
Very confident	255 (34.5)	246 (26.8)	487 (38.7)		
Somewhat confident	479 (39.2)	420 (26.0)	733 (34.7)		
Not confident	307 (47.6)	215 (25.2)	302 (27.1)		
Ever withheld information due to privacy concern				0.4 (2)	.66
Yes	160 (43.0)	128 (24.4)	222 (32.6)		
No	882 (39.4)	754 (26.3)	1306 (34.2)		
Concerned about security of information when sent between providers				1.2 (4)	.32
Very concerned	226 (41.9)	191 (25.9)	338 (32.3)		
Somewhat concerned	510 (40.4)	431 (24.4)	756 (35.3)		
Not concerned	305 (37.9)	259 (28.9)	433 (33.2)		

[a] Percentages are weighted.

Assessing Likelihood of Utilizing Personal Health Information Online

In the second set of analyses, we formulated a multivariable binomial logistic regression model based on a combination of model selection techniques (Hosmer-Lemeshow and Akaike information criterion). Results are presented in Table 2.

The first notable finding was that, even when controlling for the influence of other variables in the model, there was a strong and unique contribution from number of chronic conditions;

those reporting two or more chronic conditions had significantly higher odds of reporting medium/high use of ePHRs (OR 2.55, 95% CI 1.36-3.71; Table 2). Once we controlled for numbers of chronic conditions, the relationship between age and ePHR use also persisted with those in the lowest age bracket having the greatest odds of accessing ePHRs (OR 3.81, 95% CI 1.53-9.52; Table 2). The independent relationship between income and ePHR use also persisted, with those earning more than US $75,000 having the highest odds for accessing ePHRs (OR 3.74, 95% CI 1.74-8.07; Table 2). Of the health care system-related variables included in the model, only the regular provider variable showed a significant relationship to ePHR access (OR 1.72, 95% CI 1.07-2.77; Table 2).

Table 2. Weighted multivariate logistic regression model of predictors of using electronic personal health records among those reporting having Internet access or who own a mobile phone (n=2941).

Predictors of use of electronic personal health records	OR (95% CI)	Beta (SE)	Adj Wald *F* (*df*)	*P* value
Number of chronic conditions			4.51 (2)	.02
0	Ref	Ref		
1	0.98 (0.60-1.59)	-0.02 (0.24)		
≥2	1.88 (1.09-3.24)	0.63 (0.27)		
Sex			0.13 (1)	.72
Male	Ref	Ref		
Female	1.06 (0.77-1.45)	0.16 (0.16)		
Age (years)			2.05 (4)	.10
≥75	Ref	Ref		
65-74	1.80 (0.69-4.66)	0.59 (0.48)		
50-64	2.39 (1.01-5.67)	0.87 (0.43)		
35-49	2.68 (1.13-6.36)	0.98 (0.43)		
18-34	3.23 (1.24-8.41)	1.17 (0.47)		
Race/ethnicity			0.98 (4)	.43
Non-Hispanic White	Ref	Ref		
Hispanic	0.62 (0.31-1.26)	-0.47 (0.35)		
Non-Hispanic Black	0.90 (0.57-1.42)	-0.11 (0.23)		
Non-Hispanic other	1.34 (0.70-2.55)	0.29 (0.32)		
Missing	0.47 (0.14-1.54)	-0.76 (0.59)		
Education			1.35 (3)	.27
Less than high school	Ref	Ref		
High school graduate	1.22 (0.25-5.88)	0.20 (0.78)		
Some college	1.51 (0.35-6.52)	0.41 (0.73)		
College graduate	1.85 (0.41-8.31)	0.61 (0.75)		
Income (US$)			3.04 (4)	.03
<$20,000	Ref	Ref		
$20,000 to <$35,000	1.90 (0.81-4.47)	0.42 (-0.21)		
$35,000 to <$50,000	2.75 (1.25-6.08)	0.39 (0.22)		
$50,000 to <$75,000	1.89 (0.85-4.23)	0.40 (-0.16)		
≥$75,000	3.17 (1.50-6.71)	0.37 (0.41)		
Health insurance			1.71 (1)	.20
No	Ref	Ref		
Yes	1.48 (0.81-2.71)	0.30 (-0.21)		
Regular provider			7.43 (1)	.01
No	Ref	Ref		
Yes	1.84 (1.17-2.88)	0.61 (0.22)		
Self-reported ability to take care of own health			0.21 (2)	.81
A little confident/not at all confident	Ref	Ref		
Somewhat confident	0.97 (0.40-2.34)	-0.03 (0.44)		
Completely confident/very confident	1.14 (0.54-2.39)	0.13 (0.37)		
Self-reported general health			1.71 (2)	.19

Predictors of use of electronic personal health records	OR (95% CI)	Beta (SE)	Adj Wald *F* (*df*)	*P* value
Excellent/very good	Ref	Ref		
Good	1.40 (0.94-2.09)	0.34 (0.20)		
Fair/Poor	1.04 (0.52-2.10)	0.04 (0.35)		
Confidence that PHI is safe			5.24 (2)	.01
Not confident	Ref	Ref		
Somewhat confident	1.99 (1.25-3.17)	0.69 (0.23)		
Very confident	2.00 (1.21-3.31)	0.69 (0.25)		

Discussion

This study analyzed data from a nationally representative sample of noninstitutionalized US adults to gain a better understanding of who might be reporting use of ePHRs. Our analysis focused on usage patterns from those with MCC, a specific patient population that stands to benefit greatly from effective implementation and usage of ePHRs. When offered access to their ePHRs, patients with MCC had significantly higher odds of accessing their record more frequently than those without MCC; however, overall usage among those with MCC was lower than their healthy counterparts not reporting a chronic condition (25.6% and 29.1%, respectively).

Patients with MCC require complex care that demands a great deal from both the health care system and the patient. ePHRs and associated patient portals hold great promise for improving care coordination, patient-provider communication, shared decision making, appointment management, information consolidation, and management of medication for those with multiple comorbidities if incorporated into the patient's self-management routines on a regular basis [22,46]. However, there are barriers to accessing and adopting patient portals. Several recent studies have documented barriers, namely a lack of information about the availability of the portals and/or motivation to use them [47,48]. Furthermore, older adults, including those with chronic conditions, had more difficulty using ePHRs and patient portals than their middle-aged counterparts [49]. All these studies on access barriers, however, involved deployment of the ePHR and portals specifically for the study. In the general population, the update rates may be higher, as was shown in one investigation of proactive engagement versus passive delivery of ePHR access in primary care clinics [50].

In order to increase adoption of ePHRs among patients with MCC, there is a need to understand the current state of use especially among these individuals. Thus, we first examined the sociodemographic and health care-related characteristics associated with MCC and then looked at whether an association exists between amount of ePHR use and number of chronic conditions reported in particular. First, older age was significantly associated with a greater number of chronic conditions. As would be expected, the accumulation of chronic conditions mounts over the life span with the majority of Americans aged 75 years and older reporting two or more conditions. This is the same group of Americans that is vulnerable to being left behind as health care becomes increasingly digitally based. In fact, our data showed that the majority of people reporting two or more chronic conditions also reported not using the Internet on a regular basis. Further work is needed, however, to determine what components of ePHRs and patient portals would be of greatest utility in their day-to-day lives.

We also examined which populations reported online access to ePHRs. Here we found strong relationships between ePHR use and education, income, having health insurance, and having a regular provider. Intriguingly, we also identified a significant relationship between ePHR usage and number of chronic conditions such that patients reporting two or more chronic conditions reported twice the frequency of ePHR use compared to those who reported one or no chronic conditions. To control for those influences, we regressed ePHR usage on a set of potential predictors as included in our previous analyses. This revealed an independent, significant relationship between ePHR usage and number of chronic conditions reported regardless of age, education, income, and other variables included in the logistic regression. Specifically, those reporting two or more chronic conditions had more than two-fold increased odds of accessing their ePHRs as compared to those reporting no chronic conditions. Although some studies examining associations between patient portal use included comorbidities as part of their models, none looked specifically at the burden of MCC in association with amount of use [51-55]. One previous study has shown that patients diagnosed with two or more chronic conditions activated ePHR accounts at higher rates than their healthy counterparts [51], but to our knowledge, ours is the first study to examine population data relating use of ePHRs to number of chronic conditions.

This finding is evocative of what some have referred to as "the diagnosis effect": once diagnosed with a disease, patients tend to be more proactive in monitoring the implications of their condition or conditions [56]. For patients with Internet access, ePHRs could potentially be utilized as an extension of the clinical encounter, helping patients with some unmet needs and even the coordination of care across medical specialties. In this sense, patients with MCC may be at the vanguard of ePHR adoption and may provide an illustration of how policies increasing patient access can help create a cultural shift toward patient activation. Indeed, the few studies emerging on the efficacy of online support tools have generally shown strong clinical effects on patient engagement and outcomes [52-55]. One recent investigation examined patient activation and utilization of ePHRs via a telephone survey and found no difference in patient activation; however, these individuals were not necessarily managing MCC nor were they recently diagnosed with a new condition, both of which could affect usage [57].

Limitations of this study are largely related to the nature of cross-sectional surveys; namely, somewhat lower response rates than prospective patient studies, the fact that causation could not be inferred between variables (such as the nature of the relationship between use of online access and ratings of self-efficacy), and the fact that the data, including diagnoses of chronic conditions, were self-reported by respondents. Additionally, the survey administered did not include a comprehensive measure of patient self-efficacy, limiting our analyses to a single item for that construct. Another limitation lies in the nature of the fragmented health care system in the United States in that a patient with MCC may see many different specialists and may have multiple portals and ePHRs, and may answer “yes” to having accessed either ePHR without giving further indication as to which provider they are referring. Individuals with MCCs are more likely to have high levels of contact with their health care providers and, therefore, may have more opportunities to learn about ePHR tools and to be prompted to register for and use them. Additionally, the data on frequency of use were collected in categories rather than counts themselves. We were also unable to examine what patients were doing with their portals (eg, managing appointments, viewing laboratory results). Furthermore, this administration of HINTS did not include items addressing the quality of the ePHR systems and the degree of access that the ePHRs allow (whether patients were able to manage appointments, prescriptions, etc), which can vary between software companies and products. Survey items did not collect data on where care was received (eg, community clinics, private practice), which could have an impact on whether ePHRs were available to respondents; we attempted to control for this using sociodemographic factors. Future administrations of the HINTS instrument will collect those variables.

Despite these limitations, there are some significant strengths associated with national surveillance data, such as those reported through HINTS. These data are reported from the general population rather than being restricted to hospital-only respondent pools, the response rates and coverage results and ability to generalize are more robust than those of online panel surveys, and the sampling weights are carefully derived post stratification to generate nationally representative population estimates. Furthermore, the sampling paradigm for the administration of HINTS presented here oversampled minorities as well as rural residents of Appalachia. Previous investigations have focused on single clinics or health care systems, and many investigations have been qualitative in design. Here, we were able to present the first analyses quantifying differences in offers of access to ePHRs between those with MCCs and those without, which have not previously been examined. Due to the richness of the HINTS data, we were able to include a variety of sociodemographic and health care variables in our models that other investigations have not been able to address, including those reported from the patient’s perspective (ie, self-rated general health and self-efficacy).

Our analyses revealed a strong relationship between the presence of MCC and reports of electronic access to ePHRs as portrayed through a national probability sample of American adults. Our data show that those with two or more conditions have higher odds of accessing their ePHRs more frequently, but the numbers themselves remain low and the differences, although statistically significant, may not be of clinical significance. Our examination of variables related to both clinical and patient characteristics show that, although the system-related components may be in place, additional effort will be necessary to ensure that patients with MCC are equipped to use their ePHRs.

Targeted interventions and emphasis on patient engagement with these tools in MCC populations could greatly impact clinical outcomes. Assessments of those living with MCC have indicated that they are receptive to Web-based and app-based interventions, if the tools and interfaces were tailored to them and addressed any health and technology literacy divides [31,58]. These tools not only have the potential to support patients, but also to allow clinicians to monitor MCC patient behaviors and compliance with the various recommendations and prescribed treatments. Future planned studies involving the next HINTS administration will focus on what patients are using these portals to do; how patients are accessing these records; whether there is an association between MCCs, online access to their ePHRs, and primary care use; how use of ePHRs among those with MCC affect coordination of care; and whether outcomes are improved among those with MCC who engage with their ePHRs.

Acknowledgments

This work was funded in part by NCI Contract HHSN261201600107M.

Conflicts of Interest

None declared.

References

1. Vincent G, Velkoff V. The Next Four Decades: The Older Population in the United States 2010 to 2050. Washington, DC: US Census Bureau; 2010 May. URL: http://www.aarp.org/content/dam/aarp/livable-communities/old-learn/demographics/the-next-four-decades-the-older-population-in-the-united-states-2010-2050-aarp.pdf [accessed 2017-03-15] [WebCite Cache ID 6oyvaUfBN]

2. US Department of Health and Human Services. Multiple Chronic Conditions: A strategic Framework. Optimum Health and Quality of Life for Individuals with Multiple Chronic Conditions. Washington, DC: US Department of Health and Human Services; 2010 Dec. URL: http://www.pined.info/pdf/framework/6.pdf [accessed 2017-03-15] [WebCite Cache ID 6oyvycxIA]
3. Anderson G. National Council on Aging. 2010. Chronic care: making the case for ongoing care URL: https://www.ncoa.org/resources/chronic-care-making-the-case-for-ongoing-care/ [accessed 2017-03-15] [WebCite Cache ID 6oywQF0ae]
4. Thorpe KE, Ogden LL, Galactionova K. Chronic conditions account for rise in Medicare spending from 1987 to 2006. Health Aff 2010 Apr;29(4):718-724 [FREE Full text] [doi: 10.1377/hlthaff.2009.0474] [Medline: 20167626]
5. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007 Dec;22 Suppl 3:391-395 [FREE Full text] [doi: 10.1007/s11606-007-0322-1] [Medline: 18026807]
6. Centers for Medicare and Medicaid Services. Chronic Conditions Among Medicare Beneficiaries, Chartbook 2012 Edition. Baltimore, MD: Centers for Medicare and Medicaid Services; 2012. URL: https://www.scribd.com/document/205843458/Chronic-Conditions-Among-Medicare-Beneficiaries-Chartbook-2012-Edition [accessed 2017-03-15] [WebCite Cache ID 6oywmSdzS]
7. Lochner KA, Cox CS. Prevalence of multiple chronic conditions among Medicare beneficiaries, United States, 2010. Prev Chronic Dis 2013 Apr 25;10:e61 [FREE Full text] [doi: 10.5888/pcd10.120137] [Medline: 23618541]
8. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
9. Weppner WG, Ralston JD, Koepsell TD, Grothaus LC, Reid RJ, Jordan L, et al. Use of a shared medical record with secure messaging by older patients with diabetes. Diabetes Care 2010 Nov;33(11):2314-2319 [FREE Full text] [doi: 10.2337/dc10-1124] [Medline: 20739686]
10. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
11. Friis K, Lasgaard M, Osborne RH, Maindal HT. Gaps in understanding health and engagement with healthcare providers across common long-term conditions: a population survey of health literacy in 29,473 Danish citizens. BMJ Open 2016 Jan 14;6(1):e009627 [FREE Full text] [doi: 10.1136/bmjopen-2015-009627] [Medline: 26769783]
12. Collier T, Steenkamp R, Tomson C, Caskey F, Ansell D, Roderick P, et al. Patterns and effects of missing comorbidity data for patients starting renal replacement therapy in England, Wales and Northern Ireland. Nephrol Dial Transplant 2011 Nov;26(11):3651-3658. [doi: 10.1093/ndt/gfr111] [Medline: 21436380]
13. Edwards ST, Neri PM, Volk LA, Schiff GD, Bates DW. Association of note quality and quality of care: a cross-sectional study. BMJ Qual Saf 2014 May;23(5):406-413. [doi: 10.1136/bmjqs-2013-002194] [Medline: 24287259]
14. Smith PC, Araya-Guerra R, Bublitz C, Parnes B, Dickinson LM, Van Vorst R, et al. Missing clinical information during primary care visits. JAMA 2005 Feb 02;293(5):565-571. [doi: 10.1001/jama.293.5.565] [Medline: 15687311]
15. Cwinn MA, Forster AJ, Cwinn AA, Hebert G, Calder L, Stiell IG. Prevalence of information gaps for seniors transferred from nursing homes to the emergency department. CJEM 2009 Sep;11(5):462-471. [Medline: 19788791]
16. President's Cancer Panel. Improving Cancer-Related Outcomes with Connected Health: A Report to the President of the United States from the President?s Cancer Panel. 2016. URL: https://prescancerpanel.cancer.gov/report/connectedhealth/Letter.html [accessed 2017-03-15] [WebCite Cache ID 6oyxC7fMd]
17. Clauser SB, Wagner EH, Aiello Bowles EJ, Tuzzio L, Greene SM. Improving modern cancer care through information technology. Am J Prev Med 2011 May;40(5 Suppl 2):S198-S207 [FREE Full text] [doi: 10.1016/j.amepre.2011.01.014] [Medline: 21521595]
18. 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]
19. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. The eHealth Enhanced Chronic Care Model: a theory derivation approach. J Med Internet Res 2015;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
20. Gallacher K, May CR, Montori VM, Mair FS. Understanding patients' experiences of treatment burden in chronic heart failure using normalization process theory. Ann Fam Med 2011;9(3):235-243 [FREE Full text] [doi: 10.1370/afm.1249] [Medline: 21555751]
21. Boyd CM, Wolff JL, Giovannetti E, Reider L, Weiss C, Xue Q, et al. Healthcare task difficulty among older adults with multimorbidity. Med Care 2014 Mar;52 Suppl 3:S118-S125 [FREE Full text] [doi: 10.1097/MLR.0b013e3182a977da] [Medline: 24561750]
22. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015;17(6):e137 [FREE Full text] [doi: 10.2196/jmir.4381] [Medline: 26043709]

23. Bratzke LC, Muehrer RJ, Kehl KA, Lee KS, Ward EC, Kwekkeboom KL. Self-management priority setting and decision-making in adults with multimorbidity: a narrative review of literature. Int J Nurs Stud 2015 Mar;52(3):744-755 [FREE Full text] [doi: 10.1016/j.ijnurstu.2014.10.010] [Medline: 25468131]
24. Civan A, Skeels MM, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annu Symp Proc 2006:156-160 [FREE Full text] [Medline: 17238322]
25. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician 2014 Dec;60(12):1123-1133 [FREE Full text] [Medline: 25642490]
26. Ory MG, Ahn S, Jiang L, Lorig K, Ritter P, Laurent DD, et al. National study of chronic disease self-management: six-month outcome findings. J Aging Health 2013 Oct;25(7):1258-1274. [doi: 10.1177/0898264313502531] [Medline: 24029414]
27. Ory MG, Ahn S, Jiang L, Smith ML, Ritter PL, Whitelaw N, et al. Successes of a national study of the Chronic Disease Self-Management Program: meeting the triple aim of health care reform. Med Care 2013 Nov;51(11):992-998. [doi: 10.1097/MLR.0b013e3182a95dd1] [Medline: 24113813]
28. Marks R, Allegrante JP, Lorig K. A review and synthesis of research evidence for self-efficacy-enhancing interventions for reducing chronic disability: implications for health education practice (part I). Health Promot Pract 2005 Jan;6(1):37-43. [doi: 10.1177/1524839904266790] [Medline: 15574526]
29. Leroy L, Bayliss E, Domino M, Miller BF, Rust G, Gerteis J, AHRQ MCC Research Network. The Agency for Healthcare Research and Quality Multiple Chronic Conditions Research Network: overview of research contributions and future priorities. Med Care 2014 Mar;52 Suppl 3:S15-S22. [doi: 10.1097/MLR.0000000000000095] [Medline: 24561753]
30. Townsend A, Adam P, Li LC, McDonald M, Backman CL. Exploring eHealth ethics and multi-morbidity: protocol for an interview and focus group study of patient and health care provider views and experiences of using digital media for health purposes. JMIR Res Protoc 2013;2(2):e38 [FREE Full text] [doi: 10.2196/resprot.2732] [Medline: 24135260]
31. Steele GC, Miller D, Kuluski K, Cott C. Tying eHealth tools to patient needs: exploring the use of eHealth for community-dwelling patients with complex chronic disease and disability. JMIR Res Protoc 2014;3(4):e67 [FREE Full text] [doi: 10.2196/resprot.3500] [Medline: 25428028]
32. Smith SG, Pandit A, Rush SR, Wolf MS, Simon C. The association between patient activation and accessing online health information: results from a national survey of US adults. Health Expect 2015 Dec;18(6):3262-3273. [doi: 10.1111/hex.12316] [Medline: 25475371]
33. Cayton H. The flat-pack patient? Creating health together. Patient Educ Couns 2006 Sep;62(3):288-290. [doi: 10.1016/j.pec.2006.06.016] [Medline: 16876378]
34. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
35. Greene J, Hibbard JH, Sacks R, Overton V. When seeing the same physician, highly activated patients have better care experiences than less activated patients. Health Aff (Millwood) 2013 Jul;32(7):1299-1305. [doi: 10.1377/hlthaff.2012.1409] [Medline: 23836747]
36. Hibbard JH, Greene J, Overton V. Patients with lower activation associated with higher costs; delivery systems should know their patients' 'scores'. Health Aff (Millwood) 2013 Feb;32(2):216-222. [doi: 10.1377/hlthaff.2012.1064] [Medline: 23381513]
37. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res 2015 Feb 20;17(2):e42 [FREE Full text] [doi: 10.2196/jmir.3157] [Medline: 25707036]
38. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-460; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]
39. Hesse BW, Gaysynsky A, Ottenbacher A, Moser RP, Blake KD, Chou WS, et al. Meeting the healthy people 2020 goals: using the Health Information National Trends Survey to monitor progress on health communication objectives. J Health Commun 2014 Dec;19(12):1497-1509 [FREE Full text] [doi: 10.1080/10810730.2014.954084] [Medline: 25491584]
40. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 4;362(9):859-860. [doi: 10.1056/NEJMc0909595] [Medline: 20200398]
41. Westat. Health Information National Trends Survey 4 (HINTS 4): Cycle 4 Methodology Report. Bethesda, MD: National Cancer Institute; 2015. URL: https://hints.cancer.gov/docs/Instruments/HINTS_4_Cycle_4_Methodology_Report_Final.pdf [accessed 2017-03-15] [WebCite Cache ID 6oz34yWBk]
42. Blanch-Hartigan D, Chawla N, Moser RP, Finney Rutten Lila J, Hesse BW, Arora NK. Trends in cancer survivors' experience of patient-centered communication: results from the Health Information National Trends Survey (HINTS). J Cancer Surviv 2016 Dec;10(6):1067-1077. [doi: 10.1007/s11764-016-0550-7] [Medline: 27193357]
43. Chen Y, Feeley TH. Numeracy, information seeking, and self-efficacy in managing health: an analysis using the 2007 Health Information National Trends Survey (HINTS). Health Commun 2014;29(9):843-853. [doi: 10.1080/10410236.2013.807904] [Medline: 24266723]

44. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
45. Chou WS, Wang LC, Finney Rutten LJ, Moser RP, Hesse BW. Factors associated with Americans' ratings of health care quality: what do they tell us about the raters and the health care system? J Health Commun 2010;15 Suppl 3:147-156. [doi: 10.1080/10810730.2010.522692] [Medline: 21154090]
46. Wells S, Rozenblum R, Park A, Dunn M, Bates DW. Personal health records for patients with chronic disease: a major opportunity. Appl Clin Inform 2014;5(2):416-429 [FREE Full text] [doi: 10.4338/ACI-2014-01-RA-0002] [Medline: 25024758]
47. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i8-i12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
48. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010;341:c5814 [FREE Full text] [Medline: 21081595]
49. Taha J, Czaja SJ, Sharit J, Morrow DG. Factors affecting usage of a personal health record (PHR) to manage health. Psychol Aging 2013 Dec;28(4):1124-1139. [doi: 10.1037/a0033911] [Medline: 24364414]
50. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]
51. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
52. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
53. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
54. Lau M, Campbell H, Tang T, Thompson DJ, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
55. Wagner PJ, Dias J, Howard S, Kintziger KW, Hudson MF, Seol Y, et al. Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc 2012 Aug;19(4):626-634 [FREE Full text] [doi: 10.1136/amiajnl-2011-000349] [Medline: 22234404]
56. Fox S, Duggan M. Pew Research Center. 2013 Nov 26. The diagnosis difference: a portrait of the 45% of US adults living with chronic health conditions URL: http://www.pewinternet.org/2013/11/26/the-diagnosis-difference/ [accessed 2017-03-15] [WebCite Cache ID 6oz3evpoZ]
57. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2014 Apr 30;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
58. 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]

Abbreviations

EHR: electronic health records
ePHR: electronic personal health record
HINTS: Health Information National Trends Survey
MCC: multiple chronic conditions
PHI: personal health information

Edited by CL Parra-Calderón, G Eysenbach; submitted 01.02.17; peer-reviewed by Y Liu, W Huang, T Irizarry; comments to author 05.03.17; revised version received 24.03.17; accepted 29.04.17; published 02.06.17.

Please cite as:
Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WYS, Patel V, Moser RP, Hesse BW
Access to Electronic Personal Health Records Among Patients With Multiple Chronic Conditions: A Secondary Data Analysis
J Med Internet Res 2017;19(6):e188
URL: http://www.jmir.org/2017/6/e188/
doi:10.2196/jmir.7417
PMID:28576755

Original Paper

A Qualitative Analysis of How Online Access to Mental Health Notes Is Changing Clinician Perceptions of Power and the Therapeutic Relationship

Lauren M Denneson[1,2], PhD (Psych); Risa Cromer[1], PhD; Holly B Williams[1], BA; Maura Pisciotta[1], MS; Steven K Dobscha[1,2], MD

[1]VA Portland Health Care System, Center to Improve Veteran Involvement in Care (CIVIC), Portland, OR, United States
[2]Oregon Health & Science University, Department of Psychiatry, Portland, OR, United States

Corresponding Author:
Lauren M Denneson, PhD (Psych)
VA Portland Health Care System
Center to Improve Veteran Involvement in Care (CIVIC)
3710 SW US Veterans Hospital Rd
Portland, OR, 97239
United States
Phone: 1 503 220 8262 ext 57351
Fax: 1 503 273 5367
Email: denneson@ohsu.edu

Abstract

Background: As part of the national *OpenNotes* initiative, the Veterans Health Administration (VHA) provides veterans online access to their clinical progress notes, raising concern in mental health settings.

Objective: The aim of this study was to examine the perspectives and experiences of mental health clinicians with OpenNotes to better understand how OpenNotes may be affecting mental health care.

Methods: We conducted individual semi-structured interviews with 28 VHA mental health clinicians and nurses. Transcripts were analyzed using a thematic analysis approach, which allows for both inductive and deductive themes to be explored using an iterative, constant comparative coding process.

Results: OpenNotes is changing VHA mental health care in ways that mental health clinicians perceive as both challenging and beneficial. At the heart of these changes is a shifting power distribution within the patient-clinician relationship. Some clinicians view OpenNotes as an opportunity to better partner with patients, whereas others feel that it has the potential to undo the therapeutic relationship. Many clinicians are uncomfortable with OpenNotes, but acknowledge that this discomfort could both improve and diminish care and documentation practices. Specifically, we found that (1) OpenNotes is empowering patients, (2) OpenNotes is affecting how clinicians build and maintain the therapeutic relationship, and (3) mental health clinicians are adjusting their practices to protect patients and themselves from adverse consequences of OpenNotes.

Conclusions: Our findings suggest that future research should monitor whether OpenNotes notes facilitates stronger patient-clinician relationships, enhancing patient-centered mental health care, or diminishes the quality of mental health care through disruptions in the therapeutic relationship and reduced documentation.

(J Med Internet Res 2017;19(6):e208) doi:10.2196/jmir.6915

KEYWORDS

eHealth; physician-patient relations; mental health; patient-centered care

Introduction

Health care systems across the United States are beginning to allow patients to view their electronic health records, including clinical progress notes, online. These *OpenNotes*, and the OpenNotes initiative more generally [1], respond to recent legislation calling to increase patients' access to their health information [2-4] and are intended to improve health care transparency, facilitate patient sharing of health information with other clinicians, and encourage patient engagement in health care. The few available studies show that patients who read their medical record progress notes are more satisfied with their care, feel more informed about their health, and are more engaged and adherent to care [5-8]. Before a national rollout in 2013, the Veterans Health Administration (VHA) piloted OpenNotes at several sites across the United States to examine patient, clinician, and system-level outcomes [6]. Findings were favorable, with the majority of patients saying that OpenNotes helps them understand their health history and conditions, manage their health, prepare for clinic visits, and take their medications as prescribed [9,10].

Unlike other health care systems implementing OpenNotes, the VHA does not provide note writers the option of preventing selected notes from becoming available to patients online. As such, veterans can read or download any of their VHA progress notes (written after 2013) online, including progress notes detailing their mental health care. This removes logistical barriers for patients who wish to read their mental health notes. Although patients have had the legal right to request paper copies of their medical records for some time [11], the process can be cumbersome and time-consuming, resulting in few patients typically exercising this option [12]. Furthermore, since mental health notes often contain sensitive information about patients' mental illness, they have historically been treated differently from other progress notes, sometimes requiring clinician approval for patients to see or receive paper copies of their notes [13]. Despite promising findings from studies of primary care patients [9,10], some are concerned that this increased ease of access to mental health notes may cause unnecessary worry, confusion, or distress among patients who read their mental health progress notes without guidance or permission from their clinicians [14,15].

In a brief survey of VHA mental health clinicians about OpenNotes, approximately half of clinicians did not feel that mental health OpenNotes was a good idea [14]. Although they thought some positive outcomes might come out of OpenNotes, most expressed concern over potential negative consequences from OpenNotes and reported making changes to their note writing practices, including writing fewer details, changing the tone of the note, and writing less information about diagnoses [14]. However, the impetus for such concern and documentation changes is poorly understood. In this study, we use qualitative methods to further examine mental health clinicians' perspectives on and experiences with OpenNotes, to better understand how mental health clinicians approach care and documentation in the context of OpenNotes.

Methods

Setting and Sample

We conducted this qualitative study at a VHA Medical Center that provides comprehensive care at 11 urban and rural sites. Over 250 mental health clinicians (psychiatrists, psychologists, social workers, nurse practitioners) and nurses (registered nurses and licensed practical nurses) provide mental health care to approximately 18,000 unique patients each year across a variety of services spanning inpatient care to homelessness programs. "MyHealtheVet" is the VHA's online patient portal through which veterans receiving VHA care can access their health care records and progress notes, refill prescriptions, and securely email their clinicians.

All clinicians and nurses providing mental health care at any of the medical center's clinic sites were eligible for study participation. We sent recruitment emails to all eligible staff to describe the purpose of the study and invite interested staff to contact the study team. A total of 28 clinicians and nurses were interviewed between May and October 2014; enrollment was halted when the study team agreed we had reached saturation of themes. Over half (16/28; 57%) of the participants were female, and participants represented a range of disciplines: social workers (10/28; 36%), psychiatrists (7/28; 25%), psychologists (5/28, 18%), mental health nurse practitioners (3/28; 11%), and nurses (3/28; 11%). Participant's length of time working within VHA ranged from 1 to 30 years (mean 11.1 years).

Data Collection

We developed a semistructured interview guide informed by the main research questions and aims of the project, current literature on patient experiences with full health record access [7,16,17], and input from mental health clinicians. The interview guide focused on elucidating clinicians' thoughts across four main domains: (1) general knowledge and attitudes about OpenNotes, including familiarity with OpenNotes, concerns, and benefits; (2) experiences discussing OpenNotes with patients, including responding to patient concerns or initiating conversations; (3) experiences and changes in documentation; and (4) recommendations to other clinicians and education needs regarding clinical practice in the context of OpenNotes. For this analysis, we focused on the first three interview domains. All interviews were conducted in person by nonclinicians with backgrounds in public health and anthropology, and each interview lasted approximately 60 min. Interviews were transcribed and validated for accuracy by an independent reviewer.

Data Analysis

We used ATLAS.ti software (ATLAS.ti Scientific Software Development GmBH) to organize transcripts and facilitate analysis. We used a thematic analysis approach [18,19], which allows researchers to bring preexisting research questions to their analysis of the data while also investigating entirely unanticipated themes. Thus, deductive and inductive codes, respectively, were identified and used in our analysis. We used the main topics from our interview guide to create initial, deductive codes. Then, through an iterative, open coding process, three analysts (RC, HBW, and MP) reviewed transcripts to identify themes emerging from the text to create inductive codes. Together, these codes comprised our codebook. Once all three analysts agreed that the codebook contained the themes emerging from the transcripts, the codebook was considered final. Then, after a calibration period, two analysts applied the codebook independently to all transcripts, with a third analyst arbitrating, as needed. Specific text passages relating to the codes were compiled into code reports for analysis. All authors then reviewed and discussed code reports for thematic interpretations to refine main themes, resulting in our primary findings. In the results presented here, we use participant quotes to illustrate our findings, which are labeled using a numeric code (eg, 1001).

Results

Overview

> *(OpenNotes) really is antithetical to the way that many of us have been, literally, trained and learned to think about our field.* [1008]

Overall, analyses revealed that OpenNotes is changing VHA mental health care in ways that mental health clinicians perceive as both challenging and beneficial. At the heart of these changes is a perception of shifting power distribution within the patient-clinician relationship; OpenNotes provides patients with easier access to information about their health, their health care, and clinicians, resulting in more equitable distribution of power between clinician and patient. This is affecting how clinicians are navigating the therapeutic relationship and making changes to how they practice. Rapport building, which typically relies on carefully delivered conversations to help the patient feel comfortable enough to engage in the therapeutic relationship, is a key concern for clinicians, as progress notes leave room for miscommunication and misinterpretation. Many clinicians are uncomfortable with OpenNotes and want guidance on how to adjust their practices to protect patients and themselves from adverse consequences of OpenNotes. In the paragraphs that follow, we elaborate on these key themes: (1) OpenNotes is *shifting the patient-clinician power distribution*; (2) OpenNotes is affecting how mental health clinicians build and maintain the *therapeutic relationship* and therapeutic process; and ultimately, (3) mental health clinicians are *adjusting their practices in the context of OpenNotes* to protect patients and themselves from potential adverse consequences. Selected quotes illustrate these themes (see Multimedia Appendix 1).

Shifting Patient-Clinician Power Distribution

Clinicians discussed having less control over when and where patients access information contained in progress notes; online access to progress notes was thought to provide patients with more information about their health as well as an increased level of transparency about their health care. Clinicians compared their previously high level of control over the release of information in patients' medical records with their current lack of control (when patients read their notes online, clinicians are not informed). Some clinicians embraced this, whereas others wished to regain some control.

> *Yes it definitely has changed. There was a definite time we used to get a...message saying "can this person read their own record?"* [1009]

Many clinicians perceived the increase in patient access to, and resultant control over, their health information as a change in the distribution of power between patients and clinicians, with power becoming more equally distributed. Clinicians were not necessarily concerned about the loss of power per se, but about how the power shift affected their approach to providing care. Some clinicians viewed this shifting power differential as a move in the right direction, toward "patient-centered" care, creating better opportunities for collaboration with patients and facilitating patient engagement in care.

> *Basically, it lessens that knowledge gap between the treatment team and the patient in terms of what it is we're working towards and how does the treatment plan go about trying to achieve these goals.* [1029]

On the other hand, a few clinicians described feeling that patients could use OpenNotes to dictate to clinicians on how to write their notes and—by extension—direct their care. These clinicians were often referring to a small portion of patients that they described as particularly challenging, such as those diagnosed with borderline personality disorder or schizophrenia.

> *What I'm noticing is that, and I've directly had patients say this to me, "...don't write that in my notes." ...It's just like they're trying to dictate their care and we're trying to provide care...I feel like I'm on the defense.* [1016]

Therapeutic Relationship

Clinicians discussed the idea that developing good rapport and a therapeutic relationship is critical to patient engagement and recovery in mental health care. However, this process can be difficult and requires careful work on the part of clinicians to earn patients' trust.

> *A lot of times with mental health, there is sort of a dance that's done where a patient comes in, drops out and comes in and drops out again, and then finally comes in and feels safe enough and trusts enough to get the help. Anywhere along that line the trust gets hurt, that could be it and they are never seen again. We know that there is a lot of untreated mental illness for a lot of reasons, but that's certainly one. It's very hard to trust people with your most near and dear emotional psychological stuff. Trust is just the main thing we've got to help people in the mental health field, and so that's my real concern is that we run the risk of damaging trust with our patients.* [1008]

Many clinicians expressed concern that providing patients easier access to their notes could damage the therapeutic relationship by exposing a disconnect between the patients' in-person experience with their clinicians and the documentation they read in their notes. Specifically, notes reveal aspects of the therapeutic process—such as clinical formulations and subjective impressions—which clinicians frequently do not communicate to their patients. As such, reading notes could create opportunities for patients to negatively misinterpret clinicians' notation, or increase the likelihood that patients feel judged, stigmatized, or otherwise looked poorly upon by their clinicians. Some clinicians felt that notes had the potential to undo the work they did in session to develop good rapport with their patients.

> *People can feel belittled about something. I had somebody come in not too long ago, within the last few months, saying in a really angry way that "I don't see what my haircut has to do with anything." That's part of the mental status exam. Obviously appearance, grooming and hygiene are something we attend to*

> *see about a person's depression and their hygiene and how are they taking care of things. He felt very criticized by that. I don't think he feels criticized when he's here with me. But reading that caused a separation that I think might not have been disturbing to him if he had not seen that in print.* [1008]

On the other hand, some clinicians saw potential for OpenNotes to benefit the therapeutic relationship by enhancing feelings of trust and transparency, providing opportunity for enhanced communication even when there are disagreements, and showing patients that they are listening and have patients' best interests in mind.

> *I've heard that a couple of times, "from your charting I could see how much work you've put into it, and I could see that you care about me, and the plans that we come up with—you are hearing the things I want out of our goals and plans."* [1027]

Adjusting Practice in the Context of OpenNotes

Mental health clinicians described being careful about what they write in progress notes as a matter of course; they are acutely aware of the clinical, legal, and other audiences of notes within a large integrated health care system that uses a common electronic medical record. However, they expressed increased discomfort with the added complexity of writing notes that their patients can access online, keeping in mind both realized and potential benefits and adverse effects of patients reading their notes.

> *There's another one where someone said, "I smoked meth for 40 years and my wife doesn't know." And I was like, gosh, do I put this in the note? Because I don't know if he is going to give his wife access to his notes and then see something that was delivered in confidence...* [1006]

> *Often we're taught to document things in a particular way in order to cover ourselves for legal concerns and adding the layer of actually having the client also reading these notes just adds an additional layer of complexity to what you have to think about and how you have to phrase things in your documentation.* [1005]

In particular, clinicians felt a strong desire to protect their patients from potential harms, while also feeling vulnerable and exposed themselves.

> *Then again, for me, the onus is on us. We're the ones who are responsible for creating safety. I think that's a big part of this. If OpenNotes were to trigger somebody or create a safety issue, it's still on us to do our best to resolve it in a safe way. It shouldn't be on the person who is sick or war-torn to navigate it.* [1011]

> *So I just feel like it hinders my ability to work without the feeling—sometimes I feel a little threatened, I feel there is going to be really negative consequences if I write what I'm assessing to be clinically accurate.* [1016]

Some clinicians felt this discomfort functioned to help keep clinicians accountable, and would ultimately result in improved care and documentation. Furthermore, some clinicians liked that patients can now review notes and point out inaccuracies, which was viewed as another way to increase clinician accountability and improve the medical record documentation.

> *I think it has this sense of increasing empathy on my part. I really try to see where people are at. And I think when they're coming in here saying, "this hurt, this is what's written in my record," it forces us to kind of be in their shoes a little bit when you know they're clicking on that button and seeing what you wrote.* [1023]

> *When you know that other people are looking at the work that you do, particularly the people who it directly pertains to, then you want to make sure it's the best, it's the most accurate.* [1023]

Many clinicians were hoping for system-level guidance on how to best document care to reduce the potential for negative outcomes for them or their patients.

> *I would appreciate some clarity on who the audience is, on who I'm writing for. And I think in general, training in this more recovery-oriented and strengths-oriented treatment in general. Moving away from thinking about things, like in the medical model, in terms of problems and thinking about things more as this being a collaborative relationship with their clients.* [1005]

> *Some sort of agency/VA-type guidance on what's expected to be in a note. What should and shouldn't be left out in order to minimize the risk of the open note problem, or the potential open note problem...* [1008]

Without such guidance, some clinicians described making their own adjustments to documentation—writing fewer details, using vague terms, striving for increased objectivity, or adjusting how they document patient quotes (some increasing use of quotes and some decreasing)—as well as holding proactive conversations with patients about their notes.

> *How have I adapted? My notes are a lot less detailed now, here. I always have to kind of couch what I'm saying. There's much less detail, much less frankness in my notes now* [1006]

> *I do an informed consent about (OpenNotes). I think it's dangerous, I tell them, "Look, there's this thing called the blue button. You'll hear about it. You may want to push it. If you do you're going to see your clinical notes. That's fine by me, but understand there's stuff here that I'm going to write what I hear and see and it may be upsetting to you, and you may or may not want to do it, but there's risks associated with it."* [1010]

Discussion

Overall, we found that OpenNotes within VHA is affecting how mental health clinicians think about their relationships with patients and the progress notes they write. Primarily, they perceived reduced control over the flow of information pertinent to the therapeutic process; the notes they write are now accessible to patients at any time without clinician approval or other barriers. This change necessarily shifts the patient-clinician power dynamic toward a more equitable distribution.

Interestingly, we found that some clinicians described this power shift as a move toward more "patient-centered" care. Patient-centered care is often characterized by shared patient-clinician power and responsibility, a biopsychosocial orientation, patient and clinician humanity, and a therapeutic alliance [20]. True patient-centered care has been thought to be difficult to implement in mental health settings, where clinicians historically have had role expectations in which the patient is viewed as someone to "protect" and for whom the clinician is "responsible" [21-23]. We saw similar themes in this study. Many clinicians felt that OpenNotes provided benefits such as enhanced opportunity for collaboration, mutual trust, and addressing patients' concerns. However, many were also concerned that OpenNotes could cause unintentional harm for their patients—for whom the clinicians would feel responsible. Indeed, mental health clinicians' changes in note writing reflected a desire to write notes that would mitigate potential harm. This concern could be viewed as stemming from clinicians feeling obligated to protect their patients in the more traditional, paternalistic style of care. In previous work, nearly two-third of VHA mental health clinicians surveyed stated that they had made changes in how they document as a result of OpenNotes, with the majority reporting that they write fewer details in the notes [14]. Together, this suggests that while OpenNotes may help to facilitate care that is more aligned with patient-centered care ideals, mental health clinicians are also often limiting what they write in response to a desire to protect themselves and their patients, which could have unintended negative consequences such as forcing clinicians to rely on their memory more often or reducing clinician-clinician collaboration.

Although most clinicians felt some discomfort with OpenNotes, it is also important to note that some clinicians thought this discomfort might improve care by motivating clinicians to be at their best and providing an impetus to generate "difficult" but important conversations between patients and clinicians. Clinicians surmised that documentation could become more accurate as clinicians pay closer attention to what they write and patients have the opportunity to review and request changes. Generally, clinicians wanted guidance to help them navigate how to document in the context of OpenNotes, suggesting that mental health clinicians are motivated to improve and adapt their documentation.

Our findings should be considered in light of several limitations. We interviewed clinicians and nurses who expressed interest in participating; they may have stronger concerns or views than other clinicians. However, we heard a wide range of opinions and thoughts in response to our questions and we heard a range of experiences in the extent to which clinicians talked about OpenNotes with their patients and their knowledge of patients using OpenNotes. We also note that we do not know the actual potential for the possible outcomes of OpenNotes that clinicians discuss here (eg, diminished care due to reduced documentation). This study was conducted at one VHA medical center in the Pacific Northwest; views on OpenNotes may differ across other regions of the United States or at other VHA facilities. Finally, the VHA is unique in ways that may affect how mental health clinicians think about documentation, limiting our findings' generalizability to other health care settings. For some examples, VHA is an integrated care system in which nonmental health clinicians have access to their patients' mental health notes; some patients receive health benefits as a result of injury during military service, which can be inadvertently affected by clinician documentation; and some veterans eligible for redeployment may worry that contents of their medical record could impact redeployment eligibility.

Findings from this study suggest that online patient access to their mental health progress notes is changing how VHA mental health clinicians think about and document care—in both positive and negative ways. Clinicians perceive a shift in the balance of power between clinicians and patients, primarily resulting from reduced clinician control over the flow of information pertinent to the therapeutic relationship and process. Clinicians often view this as a shift toward more patient-centered care, but many find the change uncomfortable. This discomfort may result in improved documentation and conversations with patients, or could lead to some unintended negative consequences such as reducing what they document. Future research should continue to monitor impacts of OpenNotes in mental health settings and identify methods to reduce potential harms and enhance benefits of OpenNotes.

Acknowledgments

This material is based on work supported by the Department of Veterans Affairs, Veterans Health Administration, and Health Services Research and Development Service Project IIR 13-347. Dr Denneson is a Core Investigator at the VA HSR&D Center to Improve Veteran Involvement in Care at the VA Portland Health Care System. The funders did not give input on study design; the collection, analysis, or interpretation of data; the writing of the article; or the decision to submit it for publication. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or United States government.

Authors' Contributions

SD and LD participated in the conception and design of the study. All authors contributed to the analysis and interpretation of the data. LD and RC lead the drafting of the manuscript, and SD, HW, and MP contributed to critical revisions of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Main themes and supporting clinician interview quotes.

[PDF File (Adobe PDF File), 83KB - jmir_v19i6e208_app1.pdf]

References

1. Open Notes. Patients and clinicians on the same page URL: http://www.opennotes.org/ [accessed 2016-10-21] [WebCite Cache ID 6lQzrWA8h]
2. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 4;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]
3. 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]
4. Centers for Medicare and Medicaid Services. EHR Incentive Programs Centers for Medicare & Medicaid Services URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRIncentivePrograms/
5. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
6. Woods SS, Schwartz E, Tuepker A, Press N, Nazi KM, Turvey C, 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 Mar 17;15(3):e65. [doi: 10.2196/jmir.2356]
7. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
8. Winkelman WJ, Leonard KJ, Rossos PG. Patient-perceived usefulness of online electronic medical records: employing grounded theory in the development of information and communication technologies for use by patients living with chronic illness. J Am Med Inform Assoc 2005;12(3):306-314 [FREE Full text] [doi: 10.1197/jamia.M1712] [Medline: 15684128]
9. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. J Am Med Inform Assoc 2014;21(4):657-663. [doi: 10.1136/amiajnl-2014-002723] [Medline: 24740865]
10. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
11. Office for Civil Rights, HHS. Health Insurance Portability and Accountability Act of 1996 (HIPAA). Public Law 1996:104-191.
12. Fowles JB, Kind AC, Craft C, Kind EA, Mandel JL, Adlis S. Patients' interest in reading their medical record: relation with clinical and sociodemographic characteristics and patients' approach to health care. Arch Intern Med 2004 Apr 12;164(7):793-800. [doi: 10.1001/archinte.164.7.793] [Medline: 15078650]
13. Crichton P, Douzenis A, Leggatt C, Hughes T, Lewis S. Are psychiatric case-notes offensive? Psychiatr Bull R Coll Psychiatr 1992 Nov;16(11):675-677. [Medline: 11659954]
14. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
15. Delbanco T, Walker J, Darer JD, Elmore JG, Feldman HJ, Leveille SG, et al. Open notes: doctors and patients signing on. Ann Intern Med 2010 Jul 20;153(2):121-125. [doi: 10.7326/0003-4819-153-2-201007200-00008] [Medline: 20643992]
16. Gill MW, Scott DL. Can patients benefit from reading copies of their doctors' letters about them? Br Med J (Clin Res Ed) 1986 Nov 15;293(6557):1278-1279 [FREE Full text] [Medline: 3096465]
17. Stein EJ, Furedy RL, Simonton MJ, Neuffer CH. Patient access to medical records on a psychiatric inpatient unit. Am J Psychiatry 1979 Mar;136(3):327-329. [doi: 10.1176/ajp.136.3.327] [Medline: 420331]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Ritchie J, Lewis J. Qualitative Research Practice: A guide for social science students and researchers. London: Sage Publications; 2003.
20. Mead N, Bower P. Patient-centeredness: a conceptual framework and review of the empirical literature. Soc Sci Med 2000 Oct;51(7):1087-1110.

21. Cook J. CMHSRP. “Patient-Centered” and “Consumer-Directed” Mental Health Services URL: http://www.cmhsrp.uic.edu/download/IOMreport.pdf [accessed 2016-10-21] [WebCite Cache ID 6lR1r36vY]
22. Onken S, Dumont J, Ridgway P, Dornan D, Ralph R. Akmhcweb. 2002. Mental health recovery: what helps and what hinders? a national research project for the development of recovery facilitating system performance indicators URL: http://akmhcweb.org/docs/RecoveryPIWebDescription.pdf [accessed 2017-04-26] [WebCite Cache ID 6q0joxaBq]
23. Dobscha SK, Cromer R, Crain A, Denneson LM. Qualitative analysis of US Department of veterans affairs mental health clinician perspectives on patient-centered care. Int J Qual Health Care 2016 Jun;28(3):355-362. [doi: 10.1093/intqhc/mzw028] [Medline: 27029591]

Abbreviations

VA: Veterans Affairs
VHA: Veterans Health Administration

Edited by G Eysenbach; submitted 15.02.17; peer-reviewed by C Turvey, W Sermeus; comments to author 08.03.17; revised version received 21.03.17; accepted 23.03.17; published 14.06.17.

Please cite as:
Denneson LM, Cromer R, Williams HB, Pisciotta M, Dobscha SK
A Qualitative Analysis of How Online Access to Mental Health Notes Is Changing Clinician Perceptions of Power and the Therapeutic Relationship
J Med Internet Res 2017;19(6):e208
URL: http://www.jmir.org/2017/6/e208/
doi:10.2196/jmir.6915
PMID:28615152

Original Paper

The Second Victim Phenomenon After a Clinical Error: The Design and Evaluation of a Website to Reduce Caregivers' Emotional Responses After a Clinical Error

José Joaquín Mira[1], PhD; Irene Carrillo[2], MSc; Mercedes Guilabert[2], PhD; Susana Lorenzo[3], MD, PhD; Pastora Pérez-Pérez[4], PhD; Carmen Silvestre[5], MPH; Lena Ferrús[6], PhD; Spanish Second Victim Research Team[7]

[1]Alicante-Sant Joan Health District, Universidad Miguel Hernández, Elche (Alicante), Spain
[2]Universidad Miguel Hernández, Elche (Alicante), Spain
[3]Hospital Universitario Fundación Alcorcón, Madrid, Spain
[4]Patient Safety Observatory, Andalusian Agency for Health Care Quality, Sevilla, Spain
[5]Servicio Navarro de Salud - Osasunbidea, Pamplona, Spain
[6]Consorci Sanitari Integral, L'Hospitalet de Llobregat, Barcelona, Spain
[7]Spanish Health System, Spain

Corresponding Author:
José Joaquín Mira, PhD
Alicante-Sant Joan Health District, Universidad Miguel Hernández
Avenue Universidad s/n
Elche (Alicante), 03202
Spain
Phone: 34 606433599
Email: jose.mira@umh.es

Abstract

Background: Adverse events (incidents that harm a patient) can also produce emotional hardship for the professionals involved (second victims). Although a few international pioneering programs exist that aim to facilitate the recovery of the second victim, there are no known initiatives that aim to raise awareness in the professional community about this issue and prevent the situation from worsening.

Objective: The aim of this study was to design and evaluate an online program directed at frontline hospital and primary care health professionals that raises awareness and provides information about the second victim phenomenon.

Methods: The design of the Mitigating Impact in Second Victims (MISE) online program was based on a literature review, and its contents were selected by a group of 15 experts on patient safety with experience in both clinical and academic settings. The website hosting MISE was subjected to an accreditation process by an external quality agency that specializes in evaluating health websites. The MISE structure and content were evaluated by 26 patient safety managers at hospitals and within primary care in addition to 266 frontline health care professionals who followed the program, taking into account its comprehension, usefulness of the information, and general adequacy. Finally, the amount of knowledge gained from the program was assessed with three objective measures (pre- and posttest design).

Results: The website earned Advanced Accreditation for health websites after fulfilling required standards. The comprehension and practical value of the MISE content were positively assessed by 88% (23/26) and 92% (24/26) of patient safety managers, respectively. MISE was positively evaluated by health care professionals, who awarded it 8.8 points out of a maximum 10. Users who finished MISE improved their knowledge on patient safety terminology, prevalence and impact of adverse events and clinical errors, second victim support models, and recommended actions following a severe adverse event ($P<.001$).

Conclusions: The MISE program differs from existing intervention initiatives by its preventive nature in relation to the second victim phenomenon. Its online nature makes it an easily accessible tool for the professional community. This program has shown to increase user's knowledge on this issue and it helps them correct their approach. Furthermore, it is one of the first initiatives to attempt to bring the second victim phenomenon closer to primary care.

(J Med Internet Res 2017;19(6):e203) doi:10.2196/jmir.7840

KEYWORDS

patient safety; professionals; hospital; primary care; second victims; clinical error; e-learning

Introduction

Patient safety incidents include both near misses (incidents that do not cause harm) and adverse events (incidents that do). Although the frequency of near misses in clinical practice is difficult to specify, the frequency of adverse events at hospitals in developed countries has been established at approximately 9% [1]; in developing countries, it increases to approximately 10.5% [2]. In ambulatory care, the prevalence of adverse events has been confirmed to be approximately 2% [3-4] and 5%, respectively [5]. One-half of these adverse events are usually considered to be preventable [1]. Most are related to clinical errors, which are defined as the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim [6]. These include system failures and human errors.

Safety incidents associated with clinical errors have a negative emotional impact on patients, but also on the health professionals thought to be involved in them. The term *second victim* is used to describe the experience of the health professional who becomes emotionally overwhelmed as a result of being involved in an incident affecting patient safety [7,8]. The view of health care organizations as *third victims* was introduced because safety incidents may damage the reputation of and reduce trust in health care organizations [9].

Mitigating the impact from these incidents in patients, the health organization, and its professionals is a responsibility of managers and middle managers in the health organizations [10]. One reason is to prevent the same incident from reoccurring [11] and another is to create a proactive culture of safety that creates conditions to alleviate their impact [12].

Impact of Incidents for the Safety of Professionals

Among second victims, fear from legal consequences deriving from the harm done to the patient, fear of damage to their professional reputation, feelings of guilt, doubts about their own abilities for making clinical decisions, anxiety, and mood swings are frequent [7,13-15]. In some cases, these situations can progress toward posttraumatic stress disorder [16].

Among professionals, suspicion—if not fear—persists in disclosing what happened to patients due to the consequences that may result from such conversations [17,18]. Most professionals do not know what to do after an adverse event occurs, nor do they feel prepared for informing the patient [19,20]. They also question the support they would receive from their institution and colleagues [21-23].

Frequency of the Second Victim Phenomenon

Incident severity, its consequences, and individual variability influence the impact of the adverse event in professionals and make the number of second victims vary.

In the United States and Canada, it has been estimated that between 30% [24] and 43% [25] of professionals have experienced a negative emotional response following an incident. In one recent study carried out in Australia, 76% of the professionals involved in either a near miss or an adverse event were seen to be emotionally affected by the incident [26]. As for Spanish hospitals, as much as 69% of nurses and 77% of physicians had, either firsthand or through close colleagues, experienced being the second victim within the preceding five years [27]. In primary care, these figures varied between 55% for nurses and 67% for physicians [27]. In Belgium, Van Gerven et al [28] analyzed the magnitude of the impact among professionals, its evolution over time, and the factors that contribute to minimizing such impact, arriving at the conclusion that health organizations might anticipate this impact and plan for dealing with the second victim phenomenon.

The Help Second Victims Count On

Assistance for second victims is not part of the actions planned to be carried out when an incident affecting patient safety occurs in hospitals [9,29,30], and there are no interventions designed for primary care [29]. Professionals feel unprotected by their institutions [31-33]. Only a few hospitals have developed their own intervention programs so, thus far, the extent of intervention programs in health systems is limited.

Intervention Programs in the Literature

Two approaches in interventions have been described. On the one hand, interventions are centered on the incident [11]. On the other, they focus on dealing with the emotional consequences of the incident [24,34]. These interventions require a positive attitude, empathy toward the second victim, and awareness about the issue of clinical errors that may occur at any moment.

Scott [24], who leads the forYOU program at the University of Missouri Health System, has described the stages that a health professional goes through subsequent to an adverse event. According to this research, only 10% of second victims require specialized mental health services. Above all, most professionals in the first moments after an incident need to talk to a colleague, be relieved of care obligations for the time being, feel respect and empathy from others, and feel supported by their institution [24].

The work group of Wu [34] at Johns Hopkins Hospital has also developed an intervention program to help second victims in adverse events. Theirs is called RISE (Resilience in Stressful Events) and it is based on the fact that most professionals involved in an adverse event need to talk to a colleague, which is usually sufficient for coping with the emotional impact in most cases.

These interventions are meant to be activated after an adverse event occurs. Actions preventive in nature have not been designed for direct care professionals or for middle managers to become aware of the problem and learn how to address it.

The literature has emphasized that professionals do not know how to act after an adverse event and that most health centers do not have protocols in place or provide instructions on how to support second victims [9,27-30]. The second victim phenomenon is unknown to a large number of health professionals and managers, and interventions designed to raise awareness in professionals about the problems presented by near errors and adverse events are nonexistent. There are also not any for providing that first support that second victims need either. A need exists for intervention programs to reinforce the proactive culture of safety at health centers and to promote natural support structures among professionals that would be activated if needed after a patient safety incident.

Study Objective

This study's purpose was to develop and assess an online awareness and information program on the second victim phenomenon directed at health professionals in direct contact with patients at both hospitals and primary care. Such a program should provide demonstrations on how to act with colleagues and patients during the first moments after a severe incident for patient safety. This intervention was initially designed for Spanish frontline health care professionals.

Specific Intervention Objectives

The specific objectives included:

1. Facilitate information and training for a large number of health professionals about the issue of second victims at a reduced cost.
2. Describe emotional reactions and common behavior after being involved in an adverse event and that characterize the second victim phenomenon.
3. Describe correct and incorrect actions of how to act after an adverse effect in order to respect the rights of patients and support the second victim.
4. Act in the area of primary care, expanding the extent of studies that up to now have only taken place at hospitals.

Methods

This is a design and evaluation study of a website devised to mitigate the impact from severe adverse events in hospitals and primary care professionals. This intervention was named Mitigating Impact in Second Victims (MISE).

The phases in the design and evaluation of the website and the preventive intervention program to mitigate the initial impact from an error in health professionals (MISE) are described in Figure 1.

Website Design and Mitigating Impact in Second Victims

A website was designed that hosted an awareness and preventive intervention program (MISE) to mitigate the impact from errors in frontline professionals [35]. The website was structured around two menus: the main menu contained general information on the second victim phenomenon regarding the different actors involved (with sections entitled "Professionals," "Patients and Family," "Health Managers," "Safety Coordinators," and "Insurers"), and a secondary menu with information related to the project and its outcomes, in addition to international studies (sections entitled "Presentation," "Who we Are," "Project Timetable," "Definitions," "News," "Publications of Interest," "Reviews and Comments," and "Project Outcomes"). Access to MISE was gained by clicking on the upper right-hand corner on all website pages [36]

Based on patient safety literature, and that specifically on second victims, a preventive intervention program was designed with informative and demonstrative contents.

A review of review studies relating to open disclosure and second and third victims published in English or Spanish between 2000 and 2015 was conducted. This search was carried out using MEDLINE and Web of Knowledge. Keywords used for the review included a combination of the terms "patient safety" and "adverse event" with the following terms: emotional response, impact, professionals, second victim, third victim, and open disclosure. This yielded 22 possible documents on second victims and 83 on open disclosure. A review of health care organization websites was also conducted. This review included proposals of applied programs, checklists, and algorithms about interventions in the aftermath of an adverse event to support patient and second victims. This was carried out with the Google meta-search engine using the same descriptors. Only websites in English or Spanish were considered. A total of 16 websites were reviewed, two from Europe and the rest from the United States.

A group of 15 health professionals with at least 10 years' experience in quality and safety participated as a promoting group, and they were responsible for identifying content and elements of relevant information and example situations to be included in MISE. Chosen first to be disseminated were elements of patient safety information, specifically on second victims. Then, an index for MISE was created. Third, problem situations were selected for the demonstrative program that were then ultimately acted out by professional actors and recorded on video.

MISE was structured in two packages, one informative and the other demonstrative. The informative package offered information on basic patient safety concepts (incidents for patient safety, incidents without harm, near errors, adverse events), along with the frequency, causality, consequences, avoidability, and other characteristics of adverse events at hospitals and within primary care. Furthermore, it introduced the concepts of second and third victims and the results from research on the impact of adverse events.

The demonstrative intervention package provided a description of the emotional consequences from adverse events in professionals (affective and emotional, in clinical decision making, loss of self-esteem and professional reputation, in relationships with other professionals, with the family, legal implications) and recommendations for action following an adverse event, specifically about how to interact with the patient and their family (open disclosure), how to support a colleague who becomes a second victim, and how to personally cope with the second victim experience. This package included 15 demonstrative videos that showed what and what not to do in different clinical situations linked to errors (Table 1).

Figure 1. Study phases.

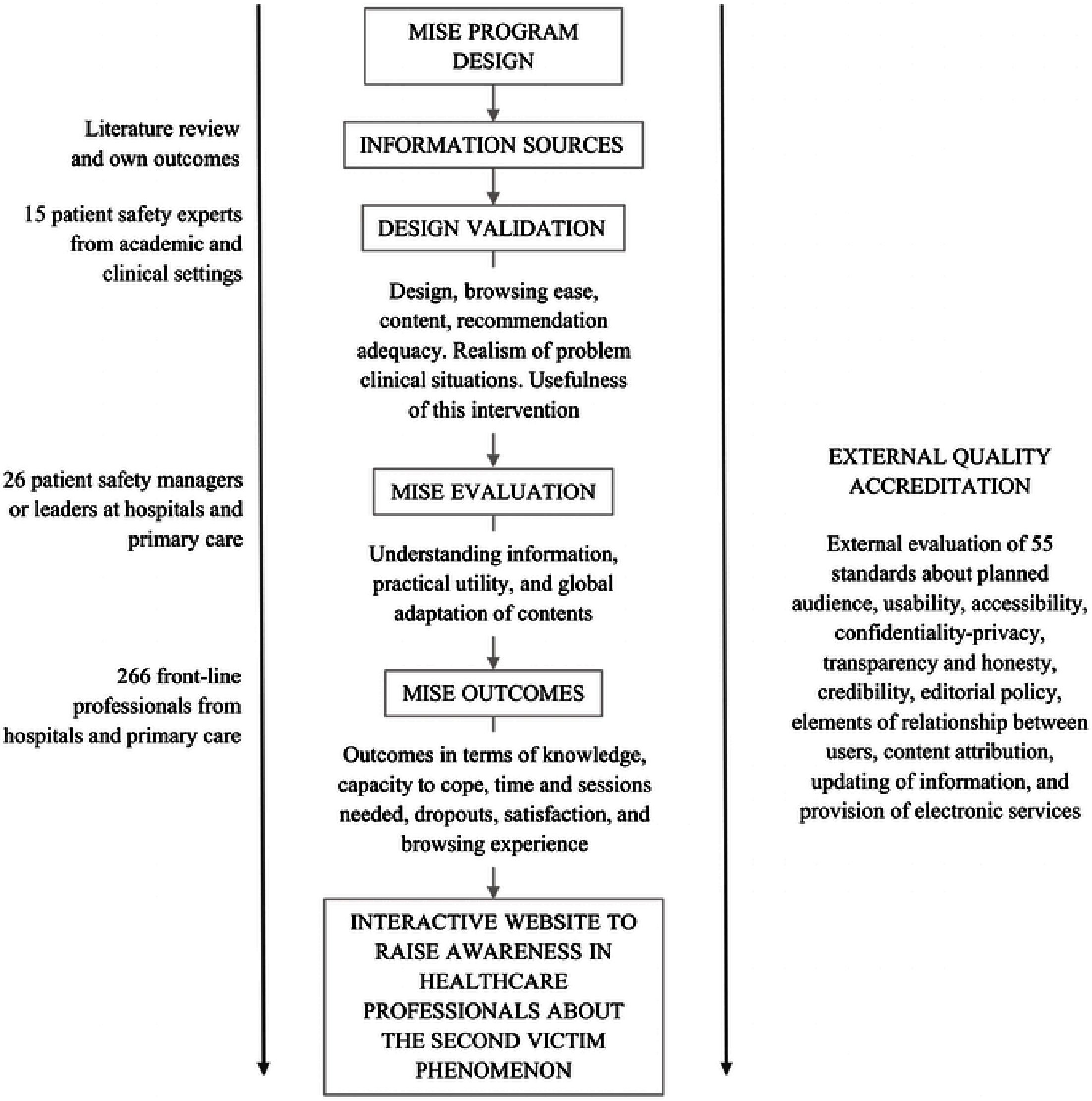


Table 1. Situations represented in the videos included in the demonstrative intervention packet.

Element covered	Situation	Number of videos
Patient information; crisis communication	A group of patients is to be vaccinated without receiving warning that an electrical power outage occurred the night before, which broke the vaccination cold chain (system failure).	2
Information for the patient and the patient’s family	Surgical material is left inside a patient that requires reintervention to extract the forgotten object.	2
Support for the second victim; information for the family of a deceased patient (the person disclosing the information is a professional other than the one involved in the event)	A patient whose condition is severe dies after presenting postintubation laryngeal stenosis. The physician attending the patient exhibits emotional affection after committing an omission mistake during the patient’s resuscitation.	2
Role of peers and supervisors in supporting the second victim; notification of incidents without harm	A nurse is emotionally affected after committing a route of administration error for a medicine that did not result in serious consequences for the patient.	4
Role of managers in supporting the second victim	A physician becomes a second victim after committing a diagnosis error with serious implications for the patient’s health. After identifying the error and having a substitute professional inform the patient, the patient files a complaint.	2
Preventive measures after an error deriving from a system failure	A nurse mistakenly administers to a patient a nonindicated medication due to incorrect storage of said drug.	3

Textbox 1. Index of the contents and their format in the website’s informative and demonstrative packages.

Informative package

What is a patient safety incident? (text)

What is an incident without harm—near error? (text; expert video)

What is an adverse event (AE)? (text; PowerPoint presentation with voice narration)

Description of the research project at primary care and hospitals

What is intended?—project objectives (text)

Why this project? (text; PowerPoint presentation with voice narration)

Who are we? (text; PowerPoint presentation with voice narration)

Sources of information on patient safety (PowerPoint presentation with voice narration)

Second victims (text)

What is a second victim? (text; expert video)

What do we know about second victims? (text; PowerPoint presentation with voice narration)

Outcomes of our study (text; PowerPoint presentation with voice narration)

Institutional help: BICEPS, forYOU, Johns Hopkins Hospital protocol (text; PowerPoint presentation with voice narration)

Spanish Ministry of Health Strategy on Patient Safety (text; link to PDF document on strategy)

What must not be done? (text)

Third victims

What is a third victim? (text; PowerPoint presentation with voice narration)

Outcomes of our study (text; PowerPoing video with voice turned off; link to PDF of questionnaires used in the study)

Crisis communication (PowerPoint presentation with voice narration)

What must not be done? (text)

Demonstrative package

Emotional, family, and work consequences in professionals from adverse events (text; PowerPoint presentation with voice narration)

Recommendations about how to act after an adverse event (link to document guide available in both Spanish and English [37] created by the Second and Third Victims Research Group based on reviews of international papers and adaptation to particularities of the Spanish context)

Role of managers (link to Safety Agenda Mobile app [12]; demonstrative videos on crisis communication)

Care for the first victim (text)

Informing the patient (PowerPoint presentation)

Open disclosure step-by-step guide (PowerPoint presentation with voice narration; demonstrative videos on informing the patient who has suffered an adverse event and the patient’s family)

Support for the second victim (text and images)

Role of second victim’s peers (PowerPoint presentation with voice narration)

Guide on how and how not to act (PowerPoint presentation with voice narration; demonstrative videos on the role of peers, supervisors, and managers in support of the second victim)

How to be prepared and know what happened (text; demonstrative videos on preventive measures after error deriving from system failure)

Root cause analysis (PowerPoint presentation with voice narration)

2.5.2. London protocol (PowerPoint presentation with voice narration)

In order to make MISE contents more dynamic, different formats were used to convey information: text, images, Portable Document Formats (PDFs), PowerPoint presentations with voice narration, videos in which a patient safety expert appears and explains a concept, demonstrative videos (simulations of situations with actors), in addition to a mobile app.

Textbox 1 lists the contents selected by the group of experts for each MISE package and the chosen format in each case for presenting such information.

First, an independent agency specializing in the evaluation of health websites completed the accreditation of its overall design, structure, organizational, and functional quality according to a certification standard [38]. This external evaluation was led by technical personnel from that agency with experienced auditors.

Second, MISE was evaluated by academic and professional safety experts who themselves were responsible for the services of patient quality and safety at hospitals and within primary care in Spain.

Then, MISE was evaluated by a group of professionals who voluntarily followed this program between November 2015 and February 2017. Moreover, they considered the usefulness of MISE for improving information about the second victim phenomenon and what to do after an adverse event. For this purpose, the participants answered a series of knowledge tests.

Website Certification (External Assessment)

The research team assessed the website (self-assessment), following the quality standards of the Andalusian Agency for Healthcare Quality [38]. It was then evaluated externally following the accreditation program for health-related websites of this agency.

This accreditation procedure consisted of 55 standards (31 were required, 10 were recommended, but in order to receive Advanced Accreditation, 8 of these must be met; the remaining 14 are voluntary commitments) that address the following aspects: usability, accessibility, confidentiality-privacy, transparency, credibility, editorial policy, elements related to the Web user, attribution of contents, updating of information, and provision of electronic services.

This evaluation was based on a double procedure of self-assessment and external evaluation. The self-assessment permitted interactive identification of elements from the webpage in need of improvements. By following this system, changes in the design and browsing conveniences were introduced into the website. The subsequent external evaluation ensured compliance with the criteria based on webpage operation, content, and resources.

Suitability of Mitigating Impact in Second Victims by Patient Safety Experts

A group of 26 health professionals who were managers of patient safety services assessed MISE. This professional profile was chosen because their criteria were thought to be the best for assessing the program's focus and content. To complete their assessment, they were allowed to freely explore the program for several weeks. This group included physicians and nurses from the health services at hospitals and within primary care, and the participants had more than three years' professional experience in patient safety. They completed an online questionnaire after being called on the telephone to request their participation and to provide an email address in order to send them an online questionnaire link. Their responses were anonymous and voluntary, and these experts assessed the comprehension of the information, practical usefulness of the contents, and overall suitability on a scale from 0 to 10, with 10 representing the highest possible assessment.

Evaluation of Mitigating Impact in Second Victims

To assess the acquisition of knowledge by intervention program users, three objective tests with pre-established response options at different points in the program were included. Specifically, two tests with pre- and posttest measures were prepared, and these included a total of 20 questions. The first evaluated the additional knowledge gained after completing the informative package (12 items); the second evaluated additional knowledge gained in the demonstrative package (8 items). These test questions consisted of statements with true/false answers. A third series of questions was used, prepared from the demonstrative videos, in which the user had to choose the correct action between two response options. These additional questions also permitted assessing the program's effectiveness in terms of the ability to discriminate between how to act in each situation. These consisted of a total of 25 questions and were answered only after the videos were watched.

Once they finished the program, the participants assessed MISE in terms of comprehending the information, the practical value of its contents, and its overall suitability. Furthermore, the following measures were also considered: the number of connections required to finish the program, total time invested to finish it, average time of each connection, number of program dropouts, and correct answers on the knowledge test (pre- and postmeasures and questions on the situations represented in the demonstrative videos).

Participants in Mitigating Impact in Second Victims Evaluation

Safety professionals from nine autonomous health services in Spain participated in this study. As a country, Spain has 17 autonomous communities, and each has its own health system. The nine participating health services account for 75% of all care activity occurring at hospitals and within primary care in the country.

A sample of 351 professionals from hospitals and primary care within these health services were asked to voluntarily participate. A minimum sampling size of 245 participants was determined, considering a sampling error of 5%, 80% correct answers on the questions, and 70% participation for a 95% confidence level. Quality and safety managers at the centers collaborated in recruiting participants by extending invitations to their hospital and health center staff to participate. To enter the system, the participants had to use a personal password to identify themselves; this way, they could continue participating in the MISE program as time permitted.

Before entering the system, the participants were informed about the study's scope, objectives, and method, in addition to the conditions for their participation. They granted their consent as a requirement for access.

Simple Blind System

Two databases were employed. The first contained the keys used by each participant, separate from the remaining databases. The second contained the anonymized registries of the participants' responses. Only the authorized webmaster had access to the participant databases and no personal data in the response database were linked to the pre-post measures.

Statistics

A student *t* test with repeated measures (intrasubject comparisons) was used to assess the intervention's effectiveness by comparing the pre-post measures. A McNemar test was used to assess the impact of the videos.

Investigation Ethics

This study was approved by the Ethics Committee of the San Juan de Alicante Hospital (Alicante, Spain).

Results

Accreditation

The external evaluation recognized the entire research project website (including MISE) as a health website and awarded it the level of Advanced Accreditation on November 25, 2016. To gain this recognition, 100% of the required standards were satisfied (31/31) along with 80% or more of those recommended (8/10), surpassing the thresholds required by the evaluated standards. Overall, the website complied with 73% (40/55) of those standards. Four standards were not applicable because they referred to patients' rights and the treatment of their health information (the website is directed solely at health professionals and does not allow compiling clinical data of patients), advertising content (absent on the website), virtual health communities (interaction between users is not included among the website's objectives), and the provision of electronic services (the website is not used as a tool for carrying out commercial activities). If these four standards are discounted, the percentage of compliance with the requirements increases from 73% to 78%.

Its strengths were related to identifying the recipients, usability, confidentiality-privacy, transparency and honesty, credibility, attribution of contents, and updating of information. Areas for improvement were related to elements related to website users, accessibility, editorial policy, and usability (Table 2).

Table 2. Results of the website's external accreditation.

Element evaluated (grouping of standards)	Standards, n	Compliance, n (%)
Target audience	1	1 (100)
Usability	11	8 (73)
Accessibility	20	13 (65)
Confidentiality-privacy (privacy and data protection)	4	4 (100)
Transparency and honesty	3	3 (100)
Credibility	2	2 (100)
Editorial policy	6	4 (67)
Elements related to website users	3	0 (0)
Attribution of contents	4	4 (100)
Updating of information	1	1 (100)
Total	55	40 (73)

Table 3. Description of the user sample (N=266).

Demographics	n (%)
Sex	
Male	83 (31.2)
Female	183 (68.8)
Professional profile	
Physicians	114 (42.9)
Nurses	120 (45.1)
Other health care professionals	32 (12.0)
Medical department	
Hospital	211 (79.3)
Physicians from hospitals	174 (82.5)
Surgeons from hospitals	37 (17.5)
Primary care	55 (20.7)
Experience	
<1 year	30 (11.3)
B1 and 3 years	11 (4.1)
>3 years	225 (84.6)

Table 4. MISE evaluation by participating professionals (N=266).

Element	Datum
MISE all pages visited, n (%)	263 (98.9)
MISE dropouts, n (%)	12 (4.5)
Days to complete, mean (SD)	72.8 (40.3)
Number of MISE connections to complete program, mean (SD)	11.4 (8.3)
Connection time per session (minutes), mean (SD)	25 (17)
MISE assessment (scale from 0 to 10), mean (SD)	
Comprehension of the information	8.9 (1.1)
Practical value of the contents	8.8 (1.2)
General assessment	8.8 (1.3)

Evaluation by National Patient Safety Experts

Twenty-six patient safety experts from four autonomous health services assessed MISE (100% response rate). Of these, 92% (24/26) positively assessed the ease of browsing and following the programmed activities, 88% (23/26) positively assessed the comprehension of the contents (mean 8.8, SD 0.9), and 92% (24/26) did likewise for the practical value of the designed intervention (mean 8.7, SD 1.1).

Participation and Evaluation of the Activity

In all, 266 of 351 professionals (75.8% response rate) followed the activities proposed in MISE; of them, 183 were women and 83 were men (Table 3).

Those who completed MISE viewed 99% of its pages (Table 4). On average, two months were required to finish reading its content, watching its videos, and completing the activities. The mean number of connections needed to complete MISE was 11.4 (SD 8.3), and these ranged between 1 and 57. The mean length of each connection was almost 30 minutes. MISE dropouts (those who quit without viewing at least 70% of its pages) were less than 5% (12/266, 4.5%).

Mitigating Impact in Second Victims was highly rated by the professional users, and they awarded it almost nine points out of a maximum of 10 (Table 4). Only two of 266 participants (0.7%) awarded it less than six points for comprehension and usefulness of the program's information.

Postmeasures on the Program's Effectiveness: Pre-Post Comparisons

Participants who completed MISE increased their level of knowledge on patient safety terminology (near misses, adverse events, and sentinel events), prevalence and impact of adverse events and errors (first, second, and third victims), support models for the second victim, and the recommended actions following a severe adverse event. There was a significant difference in the pre- and postmeasures of the knowledge test of information about basic patient safety concepts, prevalence and nature of adverse events, and second victims (informative package). Out of a maximum of 12, the premeasure mean was 6.9 (SD 2.0) and the postmeasure mean was 8.8 (SD 1.6; t_{265}=–10.0, *P*<.001). There was also a significant difference in the pre- and postmeasures of the knowledge test of what to do after an adverse event or error (demonstrative package). Out of a maximum of 8, the premeasure mean was 6.3 (SD 1.5) and the postmeasure mean was 7.2 (SD 1.0; t_{265}=–6.2, *P*<.001).

The correct answers on the knowledge tests did not vary between physicians and nurses in all cases (general knowledge test: *P*=.27; informative test package, MISE: *P*=.13; and demonstrative test package, MISE: *P*=.89).

After watching the problem situations (demonstrative videos), most test questions were answered correctly with the exception of situations representing a system failure. For these, 13.9% (37/266) of the users attributed the event of the hypothetical situation shown in the video to a human error instead of a system failure, and they considered that such failures can always be prevented (Table 5).

Knowledge Test Error Analysis

In the pretest, questions in which the answer given was incorrect more than 50% of the time had to do with the number of professionals involved in this type of event, patient safety concepts (definitions of incidents without harm and the second victim), the preventive ability against system failures, and procedures for crisis communication and open disclosure (who and how). In all these cases, the participants answered these questions as being true when in fact the correct answers were false (Table 6).

Table 5. Number of correct answers after watching demonstrative videos on what and what not to do (N=266; total questions answered=25).

Video content	Possible correct answers, n	Mean (SD)	Participants with all correct answers, n (%)
Forgotten gauze video	3	2.8 (0.4)	211 (79.3)
Extubation error video	4	3.9 (0.5)	244 (91.7)
Crisis communication video	3	3.0 (0.3)	257 (96.6)
Video on support for second victim by peers	6	5.9 (0.5)	234 (88.0)
Video on support for second victim by managers	3	2.8 (0.4)	219 (82.3)
System failure video	2	1.2 (0.7)	97 (36.5)
Human error video	2	1.7 (0.4)	198 (74.4)

Table 6. Analysis and evolution of the errors (>50%) in the knowledge tests.

Item	Error, %		Difference	*P* [a]
	Pretest	Posttest		
Informative package				
If a patient is prescribed a medication that his/her medical record says he/she is allergic to, but on that occasion no harm results, we are talking about a near incident	63.3	52.1	–11.2	.03
According to available data, close to 40% of health professionals are seen as being directly involved in an adverse event every year in our country (Spain)	85.3	70.2	–15.1	<.001
Every health professional who is seen to be directly involved in an adverse event is considered a second victim	90.9	83.5	–7.4	.21
In crisis communication, not disclosing any information during the first 24 hours, until an in-depth analysis of what occurred is completed and detailed information becomes available, is fundamental	57.1	38.0	–19.1	.003
Demonstrative package				
The most appropriate professional for informing the patient who suffered an adverse event is the person seen as being most directly involved because it is this person who knows best what happened	51.4	43.0	–8.4	.30

[a] McNemar test.

Discussion

Principal Results

Mitigating Impact in Second Victims includes a set of contents that has been considered appropriate by patient safety experts. It has also shown to contribute to improving knowledge among health professionals about the second victim phenomenon. The methodology employed for disseminating this knowledge and explaining what and what not to do has been considered appropriate by the MISE participants.

Data from the MISE evaluation confirm that the program increases knowledge about the issue of second victims and how to act with a colleague when either an adverse event or near miss occurs. It also shows how to interact with patients who are victims of an adverse event, providing information on how to act and how to disclose what has occurred.

Five hours was the mean total time dedicated to complete MISE, distributed generally over 12 sessions lasting approximately 30 minutes each. This time demand is reasonable for this group of professionals and is compatible with other care responsibilities and tasks and their personal lives.

Comparison With Previous Studies

Emotional needs immediately following incidents have been analyzed in several studies. The emotional isolation professionals find themselves in, along with the difficulty of talking about what has happened with their colleagues and the lack of protection they feel from their superiors, have been identified as two important gaps that contribute to progressing along the scale of the second victim syndrome [39]. Both aspects have been corroborated by research carried out among participants in benchmark intervention programs (forYOU or RISE) with second victims and that, in turn, pointed out that most of what second victims were searching for and would have liked to receive was support from their colleagues and the management at their centers [24,34].

We know that colleagues of second victims can do much more than what they currently do to prevent the emotional impact from safety incidents for the patient from progressing until manifesting itself as posttraumatic stress [8,23,40,41]. We also know that most professionals do not require specialized intervention to alleviate their initial emotional symptoms because sensing empathy by their colleagues can be sufficient [34]. The MISE intervention program considers these aspects and seeks to act at the base of Scott’s pyramid, where 60% of the professionals who suffer from the impact of incidents for safety as second victims are found [24].

The role of managers is crucial in two senses due to their role as a barrier and their role as a facilitator [10,28,42]. Managers should prepare the organization so that if a severe incident does occur, it is prepared to act, and this includes analysis of what happened and the recovery by and support for the patient and for the second victim as well. Furthermore, they should create a just culture [43] that permits analyzing the incident without prejudging the second victim’s role in it. Likewise, managers should facilitate organizational learning from incidents using formal and informal processes as well as reactive and proactive approaches. Incident reporting is a crucial step for improving patient safety, but frontline professionals identify barriers (ie, lack of training, undesirable repercussions, lack of feedback) that lead to underuse of incident reporting systems [44]. However, Sujan [44] found that professionals use informal processes, such as regular staff meetings, discussions with line managers, and discussions with peers, that facilitate sharing concerns and experiences that can actively contribute to improving patient safety. Organization leadership should be aware of these alternative ways of learning and promote them.

Subsequent to a severe adverse event, the second victim may become helpless and not inform the patient about the incident [45-49]. However, such attention and information on the part of the professionals, when presented in an appropriate manner, usually facilitates resolution of the crisis and prevents litigation [50]. The contents and approach of MISE aim in this direction.

Relevance of This Study

Mitigating Impact in Second Victims is easily accessible to a large number of professionals. It is a low-cost program that can be accessed from work or home with ease.

This program responds to three deficiencies identified in the literature and in daily practice at health centers: (1) the lack of programs raising awareness and providing information about the second victim phenomenon that reaches high numbers of professionals, (2) the issue of second victims in primary care, and (3) the inexistence of structured interventions at most health centers to support professionals and patients following adverse events.

The MISE program provides information to professionals about the second victim phenomenon in nine weeks of online training in which are presented general issues and problem situations based on experiences after committing a clinical error.

The need for second victims to change their care functions could be reduced and less absenteeism linked to this phenomenon [27] could result if these professionals gain greater information and a change in attitude toward second victims. This sought-after attitudinal change would also facilitate a distinct attitude when interacting with patients in order to overcome the traditional difficulties resulting after an adverse event [51].

Limitations

It is possible that those who followed the MISE program were professionals who are more sensitive to the issue of what severe adverse events mean for professionals. We did not possess information about the type of professionals who declined invitations to follow MISE.

A minimum sampling size was defined considering a worst case of 80% correct answers to the questions. The correct answers related to system failure did not match this assumption.

The measure of the resulting impact was based on correct answers to knowledge tests and to self-test questions after watching a set of problem situations on video. Actual adverse events that occur may involve circumstances that are different from what these videos represented.

This study was not designed to assess its effect on secondary prevention of posttraumatic stress; that is something that future research should evaluate. Thus, the effectiveness of this website in contributing to any kind of change was not assessed, and this will be done in the future. In this sense, realistic evaluation, a form of theory-driven evaluation developed by Pawson and Tilley in 1997 [52], may be a good methodology for testing MISE. Realistic explanation refutes the idea that a program works or does not work in an absolute manner and proposes that it is necessary to identify the mechanism (ie, the process of how subjects interpret and act on the intervention) by which the program works for whom and under what particular circumstances [52]. In this way, a program can be effective for achieving some outcomes or changes but not others, always depending on the context.

Recommendations for Practice and Research

The MISE program is designed to assist intervention programs to mitigate the impact of adverse events in professionals. It is not an emotional recovery program for second victims; instead, it responds to the need for the group of professionals to understand what is felt subsequent to an adverse event. MISE also contributes to frontline professionals gaining greater awareness about the emotional needs that are experienced when an error occurs and the importance of speaking about the incident with their colleagues. It also provides ideas about how to act with the patient victim of the adverse event [37]. This way, when an adverse event does occur, MISE contributes by helping the professional affected (second victim) to face the facts and recover his/her clinical capacity and emotional balance early on.

These types of programs, along with other recommendations about what to do after an adverse event occurs, contribute to safer environments at hospitals and primary care.

Future research could analyze whether MISE modifies the frequency, which up until now is relatively low, of patients who are victims of adverse events being informed about the incident. This research could also examine the impact of MISE in the initial hours after an incident. One example of this would be whether the colleagues of the second victims gain a greater ability to listen and act appropriately to prevent the emotional escalation that may result.

Acknowledgments

This study was financed by the Spanish Research Health Agency (FIS) and the European Regional Development Fund (references PI13/01220 and PI13/0473).

Authors' Contributions

JJM and SL conceived of the study. All participated in the design of the study. IC and JJM performed the statistical analysis. SL, PP, CS, and LF coordinated the qualitative and quantitative research, and captured and prepared data to design and improve MISE. IC and MG supervised the website. JJM, IC, and MG prepared a first version of this original. All authors read and approved the final manuscript.

The Spanish Second Victim Research Team is composed of Roser Anglès (Vall d'Hebron University Hospital), Pilar Astier (Servicio Aragonés de Salud), Angélica Bonilla (Complutense University of Madrid), Ana Jesús Bustinduy (Gipuzkoa Primary Care Health District), Clara Crespillo (Alcorcón Foundation University Hospital), Sara Guila Fidel (Vall d'Hebron University Hospital), Álvaro García (Alcorcón Foundation University Hospital), Ana Jesús González (Gipuzkoa Primary Care Health District), María Jesús Gutiérrez, Emilio Ignacio (University of Cadiz), Fuencisla Iglesias (SESCAM, Toledo), Juan José Jurado (Madrid Primary Care Health Centre), Araceli López (Integrated Health Consortium), José Ángel Maderuelo-Fernández (SACYL), Mª Magdalena Martínez (Alcorcón Foundation University Hospital), Isabel María Navarro (Miguel Hernández University), María Cristina Nebot (Department of Health; Regional Government of Valencia), Roberto Nuño-Solinís (Universidad de Deusto), Guadalupe Olivera (SERMAS, Madrid), Antonio Ochando (Alcorcón Foundation University Hospital), Pedro Orbegozo (Gipuzkoa Primary Care Health District), Elene Oyarzabal (Donostia University Hospital), Jesús María Palacio (Muñoz Fernández-Ruiseñores Health Centre), María Esther Renilla (Alcorcón Foundation University Hospital), Carolina Rodriguez-Pereira (Basque Institute for Healthcare Innovation, O+Berri), Sira Sanz (Alcorcón Foundation University Hospital), María Luisa Torijano (Castilla-La Mancha Health Service, SESCAM), Julián Vitaller (Consellería Sanidad Universal y Salud Pública), Elena Zavala (Osakidetza, Donostia).

Conflicts of Interest

None declared.

References

1. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008 Jun;17(3):216-223 [FREE Full text] [doi: 10.1136/qshc.2007.023622] [Medline: 18519629]
2. Aranaz-Andrés JM, Aibar-Remón C, Limón-Ramírez R, Amarilla A, Restrepo FR, Urroz O, IBEAS team. Prevalence of adverse events in the hospitals of five Latin American countries: results of the 'Iberoamerican Study of Adverse Events' (IBEAS). BMJ Qual Saf 2011 Dec;20(12):1043-1051. [doi: 10.1136/bmjqs.2011.051284] [Medline: 21712370]

3. Tsang C, Bottle A, Majeed A, Aylin P. Adverse events recorded in English primary care: observational study using the General Practice Research Database. Br J Gen Pract 2013 Aug;63(613):e534-e542 [FREE Full text] [doi: 10.3399/bjgp13X670660] [Medline: 23972194]
4. Aranaz-Andrés JM, Aibar C, Limón R, Mira JJ, Vitaller J, Agra Y, et al. A study of the prevalence of adverse events in primary healthcare in Spain. Eur J Public Health 2012 Dec;22(6):921-925. [doi: 10.1093/eurpub/ckr168] [Medline: 23180803]
5. Montserrat-Capella D, Suárez M, Ortiz L, Mira JJ, Duarte HG, Reveiz L, AMBEAS Group. Frequency of ambulatory care adverse events in Latin American countries: the AMBEAS/PAHO cohort study. Int J Qual Health Care 2015 Feb;27(1):52-59. [doi: 10.1093/intqhc/mzu100] [Medline: 25609774]
6. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 2000.
7. Wu AW. Medical error: the second victim. The doctor who makes the mistake needs help too. BMJ 2000 Mar 18;320(7237):726-727 [FREE Full text] [Medline: 10720336]
8. Scott SD, Hirschinger LE, Cox KR, McCoig M, Brandt J, Hall LW. The natural history of recovery for the healthcare provider “second victim” after adverse patient events. Qual Saf Health Care 2009 Oct;18(5):325-330. [doi: 10.1136/qshc.2009.032870] [Medline: 19812092]
9. Seys D, Wu AW, Van Gerven E, Vleugels A, Euwema M, Panella M, et al. Health care professionals as second victims after adverse events: a systematic review. Eval Health Prof 2013 Jun;36(2):135-162. [doi: 10.1177/0163278712458918] [Medline: 22976126]
10. Burlison JD, Quillivan RR, Scott SD, Johnson S, Hoffman JM. The effects of the second victim phenomenon on work-related outcomes: connecting self-reported caregiver distress to turnover intentions and absenteeism. J Patient Saf 2016 Nov:1 (forthcoming). [doi: 10.1097/PTS.0000000000000301] [Medline: 27811593]
11. Carrillo I, Mira JJ, Vicente MA, Fernandez C, Guilabert M, Ferrús L, et al. Design and testing of BACRA, a Web-based tool for middle managers at health care facilities to lead the search for solutions to patient safety incidents. J Med Internet Res 2016 Sep;18(9):e257 [FREE Full text] [doi: 10.2196/jmir.5942] [Medline: 27678308]
12. Mira JJ, Carrillo I, Fernandez C, Vicente MA, Guilabert M. Design and testing of the safety agenda mobile app for managing health care managers' patient safety responsibilities. JMIR Mhealth Uhealth 2016 Dec;4(4):e131 [FREE Full text] [doi: 10.2196/mhealth.5796] [Medline: 27932315]
13. Mizrahi T. Managing medical mistakes: ideology, insularity and accountability among internists-in-training. Soc Sci Med 1984;19(2):135-146. [Medline: 6474229]
14. Edrees HH, Paine LA, Feroli ER, Wu AW. Health care workers as second victims of medical errors. Pol Arch Med Wewn 2011 Apr;121(4):101-108 [FREE Full text] [Medline: 21532531]
15. Lander LI, Connor JA, Shah RK, Kentala E, Healy GB, Roberson DW. Otolaryngologists' responses to errors and adverse events. Laryngoscope 2006 Jul;116(7):1114-1120. [doi: 10.1097/01.mlg.0000224493.81115.57] [Medline: 16826044]
16. Rassin M, Kanti T, Silner D. Chronology of medication errors by nurses: accumulation of stresses and PTSD symptoms. Issues Ment Health Nurs 2005 Oct;26(8):873-886. [doi: 10.1080/01612840500184566] [Medline: 16203641]
17. Varjavand N, Nair S, Gracely E. A call to address the curricular provision of emotional support in the event of medical errors and adverse events. Med Educ 2012 Dec;46(12):1149-1151. [doi: 10.1111/medu.12074] [Medline: 23171256]
18. Mazor KM, Simon SR, Gurwitz JH. Communicating with patients about medical errors: a review of the literature. Arch Intern Med 2004 Aug;164(15):1690-1697. [doi: 10.1001/archinte.164.15.1690] [Medline: 15302641]
19. Manser T, Staender S. Aftermath of an adverse event: supporting health care professionals to meet patient expectations through open disclosure. Acta Anaesthesiol Scand 2005 Jul;49(6):728-734. [doi: 10.1111/j.1399-6576.2005.00746.x] [Medline: 15954950]
20. Mira JJ, Lorenzo S, en nombre del Grupo de Investigación en Segundas Víctimas. [Something is wrong in the way we inform patients of an adverse event]. Gac Sanit 2015 Sep;29(5):370-374 [FREE Full text] [doi: 10.1016/j.gaceta.2015.04.004] [Medline: 26026725]
21. Waterman AD, Garbutt J, Hazel E, Dunagan WC, Levinson W, Fraser VJ, et al. The emotional impact of medical errors on practicing physicians in the United States and Canada. Jt Comm J Qual Patient Saf 2007 Aug;33(8):467-476. [Medline: 17724943]
22. Wu AW, Boyle DJ, Wallace G, Mazor KM. Disclosure of adverse events in the United States and Canada: an update, and a proposed framework for improvement. J Public Health Res 2013 Dec;2(3):e32 [FREE Full text] [doi: 10.4081/jphr.2013.e32] [Medline: 25170503]
23. Ferrús L, Silvestre C, Olivera G, Mira JJ. Qualitative study about the experiences of colleagues of health professionals involved in an adverse event. J Patient Saf 2016 Nov:1 (forthcoming). [doi: 10.1097/PTS.0000000000000309] [Medline: 27811596]
24. Scott SD, Hirschinger LE, Cox KR, McCoig M, Hahn-Cover K, Epperly KM, et al. Caring for our own: deploying a systemwide second victim rapid response team. Jt Comm J Qual Patient Saf 2010 May;36(5):233-240. [Medline: 20480757]
25. Wolf ZR, Serembus JF, Smetzer J, Cohen H, Cohen M. Responses and concerns of healthcare providers to medication errors. Clin Nurse Spec 2000 Nov;14(6):278-287. [Medline: 11855445]

26. Harrison R, Lawton R, Stewart K. Doctors' experiences of adverse events in secondary care: the professional and personal impact. Clin Med (Lond) 2014 Dec;14(6):585-590. [doi: 10.7861/clinmedicine.14-6-585] [Medline: 25468840]
27. Mira JJ, Carrillo I, Lorenzo S, Ferrús L, Silvestre C, Pérez-Pérez P, Research Group on Second and Third Victims. The aftermath of adverse events in Spanish primary care and hospital health professionals. BMC Health Serv Res 2015 Apr;15:151 [FREE Full text] [doi: 10.1186/s12913-015-0790-7] [Medline: 25886369]
28. Van Gerven E, Bruyneel L, Panella M, Euwema M, Sermeus W, Vanhaecht K. Psychological impact and recovery after involvement in a patient safety incident: a repeated measures analysis. BMJ Open 2016 Aug;6(8):e011403 [FREE Full text] [doi: 10.1136/bmjopen-2016-011403] [Medline: 27580830]
29. Mira JJ, Lorenzo S, Carrillo I, Ferrús L, Pérez-Pérez P, Iglesias F, Research Group on Second and Third Victims. Interventions in health organisations to reduce the impact of adverse events in second and third victims. BMC Health Serv Res 2015 Aug;15:341 [FREE Full text] [doi: 10.1186/s12913-015-0994-x] [Medline: 26297015]
30. Gerven EV, Seys D, Panella M, Sermeus W, Euwema M, Federico F, et al. Involvement of health-care professionals in an adverse event: the role of management in supporting their workforce. Pol Arch Med Wewn 2014 Apr;124(6):313-320 [FREE Full text] [Medline: 24781784]
31. Pratt SD, Jachna BR. Care of the clinician after an adverse event. Int J Obstet Anesth 2015 Feb;24(1):54-63. [doi: 10.1016/j.ijoa.2014.10.001] [Medline: 25499810]
32. Christensen JF, Levinson W, Dunn PM. The heart of darkness: the impact of perceived mistakes on physicians. J Gen Intern Med 1992 Jul;7(4):424-431. [Medline: 1506949]
33. Ullström S, Andreen Sachs M, Hansson J, Ovretveit J, Brommels M. Suffering in silence: a qualitative study of second victims of adverse events. BMJ Qual Saf 2014 Apr;23(4):325-331 [FREE Full text] [doi: 10.1136/bmjqs-2013-002035] [Medline: 24239992]
34. Edrees H, Connors C, Paine L, Norvell M, Taylor H, Wu AW. Implementing the RISE second victim support programme at the Johns Hopkins Hospital: a case study. BMJ Open 2016 Sep;6(9):e011708 [FREE Full text] [doi: 10.1136/bmjopen-2016-011708] [Medline: 27694486]
35. Second and Third Victim Research Group. Second victims research project URL: http://www.segundasvictimas.es/ [accessed 2017-05-30] [WebCite Cache ID 6qqfuiwjW]
36. Second and Third Victim Research Group. Access to Mitigating Impact in Second Victims (MISE) online program URL: http://www.segundasvictimas.es/acceso.php [WebCite Cache ID 6qqhB9VUM]
37. Mira JJ, Lorenzo S, Carrillo I, Ferrús L, Silvestre C, Astier P. Lessons learned for reducing the negative impact of adverse events on patients, health professionals and healthcare organisations. Int J Qual Health Care 2017:00 (forthcoming).
38. Andalusian Agency for Health Care Quality. ACSA certification of health websites and blogs URL: http://www.juntadeandalucia.es/agenciadecalidadsanitaria/en/acsa-accreditation/accreditation-of-health-websites-and-blogs/ [accessed 2017-05-09] [WebCite Cache ID 6qqK5vuHW]
39. White AA, Waterman AD, McCotter P, Boyle DJ, Gallagher TH. Supporting health care workers after medical error: considerations for health care leaders. J Clin Outcomes Manag 2008 May;15(5):240-247.
40. Aasland OG, Førde R. Impact of feeling responsible for adverse events on doctors' personal and professional lives: the importance of being open to criticism from colleagues. Qual Saf Health Care 2005 Feb;14(1):13-17 [FREE Full text] [doi: 10.1136/qshc.2002.003657] [Medline: 15691998]
41. Wu AW, Folkman S, McPhee SJ, Lo B. How house officers cope with their mistakes. West J Med 1993 Nov;159(5):565-569 [FREE Full text] [Medline: 8279153]
42. MacLeod L. “Second victim” casualties and how physician leaders can help. Physician Exec 2014 Jan;40(1):8-12. [Medline: 24575696]
43. Dekker S. Just Culture: Balancing Safety and Accountability. Aldershot, UK: Ashgate; 2007.
44. Sujan M. An organisation without a memory: a qualitative study of hospital staff perceptions on reporting and organisational learning for patient safety. Reliab Eng Syst Safe 2015 Dec;144:45-52. [doi: 10.1016/j.ress.2015.07.011]
45. Vincent JL. Information in the ICU: are we being honest with our patients? The results of a European questionnaire. Intensive Care Med 1998 Dec;24(12):1251-1256. [Medline: 9885876]
46. Blendon RJ, DesRoches CM, Brodie M, Benson JM, Rosen AB, Schneider E, et al. Views of practicing physicians and the public on medical errors. N Engl J Med 2002 Dec;347(24):1933-1940. [doi: 10.1056/NEJMsa022151] [Medline: 12477944]
47. Gallagher TH, Waterman AD, Ebers AG, Fraser VJ, Levinson W. Patients' and physicians' attitudes regarding the disclosure of medical errors. JAMA 2003 Feb;289(8):1001-1007. [Medline: 12597752]
48. Kaldjian LC, Jones EW, Wu BJ, Forman-Hoffman VL, Levi BH, Rosenthal GE. Disclosing medical errors to patients: attitudes and practices of physicians and trainees. J Gen Intern Med 2007 Jul;22(7):988-996 [FREE Full text] [doi: 10.1007/s11606-007-0227-z] [Medline: 17473944]
49. Gallagher TH, Waterman AD, Garbutt JM, Kapp JM, Chan DK, Dunagan WC, et al. US and Canadian physicians' attitudes and experiences regarding disclosing errors to patients. Arch Intern Med 2006 Aug;166(15):1605-1611. [doi: 10.1001/archinte.166.15.1605] [Medline: 16908793]
50. Wu AW, Marks CM. Close calls in patient safety: should we be paying closer attention? CMAJ 2013 Sep;185(13):1119-1120 [FREE Full text] [doi: 10.1503/cmaj.130014] [Medline: 23589434]

51. Garbutt J, Waterman AD, Kapp JM, Dunagan WC, Levinson W, Fraser V, et al. Lost opportunities: how physicians communicate about medical errors. Health Aff (Millwood) 2008 Jan;27(1):246-255 [FREE Full text] [doi: 10.1377/hlthaff.27.1.246] [Medline: 18180501]
52. Pawson R, Tilley N. Realistic Evaluation. London: Sage; 1997.

Abbreviations

MISE: Mitigating Impact in Second Victims
PDF: Portable Document Formats
RISE: Resilience in Stressful Events

Edited by G Eysenbach; submitted 11.04.17; peer-reviewed by K Vanhaecht, C Fernández, W Sermeus, M Sujan; comments to author 17.05.17; revised version received 23.05.17; accepted 23.05.17; published 08.06.17.

Please cite as:
Mira JJ, Carrillo I, Guilabert M, Lorenzo S, Pérez-Pérez P, Silvestre C, Ferrús L, Spanish Second Victim Research Team
The Second Victim Phenomenon After a Clinical Error: The Design and Evaluation of a Website to Reduce Caregivers' Emotional Responses After a Clinical Error
J Med Internet Res 2017;19(6):e203
URL: http://www.jmir.org/2017/6/e203/
doi:10.2196/jmir.7840
PMID:28596148

Original Paper

The Gap in Medical Informatics and Continuing Education Between the United States and China: A Comparison of Conferences in 2016

Jun Liang[1], MS; Kunyan Wei[2], MD; Qun Meng[3], PhD; Zhenying Chen[4], MPH; Jiajie Zhang[5], PhD; Jianbo Lei[6,7], MD, PhD

[1]Second Affiliated Hospital, School of Medicine,Zhejiang University, Hangzhou, China
[2]Department of Gastroenterology, Affiliated Hospital of Southwest Medical University, Luzhou, China
[3]Center for Statistics and Information, National Health and Family Planning Commission of China, Beijing, China
[4]Library of Zhejiang University, Hangzhou, China
[5]School of Biomedical Informatics, University of Texas Health Sciences Center, Houston, TX, United States
[6]Center for Medical Informatics, Peking University, Beijing, China
[7]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

Corresponding Author:
Jianbo Lei, MD, PhD
Center for Medical Informatics
Peking University
38 Xueyuan Rd.
Haidian District
Beijing, 100191
China
Phone: 86 (10) 8280 5901
Fax: 86 (10) 8280 5900
Email: jblei@hsc.pku.edu.cn

Abstract

Background: China launched its second health reform in 2010 with considerable investments in medical informatics (MI). However, to the best of our knowledge, research on the outcomes of this ambitious undertaking has been limited.

Objective: Our aim was to understand the development of MI and the state of continuing education in China and the United States from the perspective of conferences.

Methods: We conducted a quantitative and qualitative analysis of four MI conferences in China and two in the United States: China Medical Information Association Annual Symposium (CMIAAS), China Hospital Information Network Annual Conference (CHINC), China Health Information Technology Exchange Annual Conference (CHITEC), China Annual Proceeding of Medical Informatics (CPMI) versus the American Medical Informatics Association (AMIA) and Healthcare Information and Management Systems Society (HIMSS). The scale, composition, and regional distribution of attendees, topics, and research fields for each conference were summarized and compared.

Results: CMIAAS and CPMI are mainstream academic conferences, while CHINC and CHITEC are industry conferences in China. Compared to HIMSS 2016, the meeting duration of CHITEC was 3 versus 5 days, the number of conference sessions was 132 versus 950+, the number of attendees was 5000 versus 40,000+, the number of vendors was 152 versus 1400+, the number of subforums was 12 versus 230, the number of preconference education symposiums and workshops was 0 versus 12, and the duration of preconference educational symposiums and workshops was 0 versus 1 day. Compared to AMIA, the meeting duration of Chinese CMIAAS was 2 versus 5 days, the number of conference sessions was 42 versus 110, the number of attendees was 200 versus 2500+, the number of vendors was 5 versus 75+, and the number of subforums was 4 versus 10. The number of preconference tutorials and working groups was 0 versus 29, and the duration of tutorials and working group was 0 versus 1.5 days.

Conclusions: Given the size of the Chinese economy and the substantial investment in MI, the output in terms of conferences remains low. The impact of conferences on continuing education to professionals is not significant. Chinese researchers and professionals should approach MI with greater rigor, including validated research methods, formal training, and effective continuing education, in order to utilize knowledge gained by other countries and to expand collaboration.

(J Med Internet Res 2017;19(6):e224) doi:10.2196/jmir.8014

KEYWORDS

medical informatics; conferences; continuing education; Sino-American comparison

Introduction

Background

China, the world's second largest economy, launched its second health reform in 2010 with considerable investments in medical informatics (MI) [1]. An important contribution of this reform is its definition of MI as one of the "four constructs and eight pillars," or foundations of health care reform [2]. Under the guidance and incentives provided in a series of policies, the Chinese government and health care informatics industry have invested a huge amount of money in hospital informatics and population health informatics. However, current MI in China can be characterized as "hot in industry application, and cold in academic research" [3]. To the best of our knowledge, research on the outcomes of this ambitious health reform has been limited.

Conferences including tutorial sessions, talks, product exhibitions, and conference proceedings are potentially important channels of continuing education, allowing professionals to update their knowledge and learn about industry developments. Therefore, this study focused on the characteristics of mainstream MI professional conferences in China and compared them to those in the United States, in order to promote the international exchange of knowledge to improve MI in China and potentially other nations, where progress in the discipline is perceived to be lagging.

Objectives

We focused on four national mainstream medical informatics conferences in China—the China Medical Information Association Annual Symposium (CMIAAS 2015) [4], the China Hospital Information Network Annual Conference (CHINC 2016) [5], the China Health Information Technology Exchange Annual Conference (CHITEC 2016) [6], and the China Annual Proceeding of Medical Informatics (CPMI 2016) [7]). These conferences are also the only four academic and industry conferences that are certified to offer continuing education credits by the National Commission of Health and Family Planning. We then compared the characteristics of those four conferences with two international MI conferences based in the United States—the American Medical Informatics Association (AMIA) and the Healthcare Information and Management Systems Society (HIMSS) 2016 conferences—with the aim of identifying gaps and summarizing lessons learned to encourage greater international cooperation.

Methods

Selection of Mainstream Medical Informatics Conferences in China

Following the eHealth Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) methodology [8], we chose four MI conferences as the mainstream conferences in mainland China: (1) CMIAAS hosted by the China Medical Informatics Association (the only country representative of International Medical Informatics Association); (2) CPMI hosted by the Medical Informatics Branch, Chinese Medical Association, which is the most authoritative organization for Chinese medicine; (3) CHINC hosted by the Committee on Information Management, Chinese Hospital Association; and (4) CHITEC hosted by the China Institutes of Health Information (the largest two industry associations in MI).

Data Collection

The three largest Chinese literature databases were searched: the VIP database of Chinese scientific and technical journals [9], the Chinese National Knowledge Infrastructure [10], and the Wanfang data retrieval platform [11]. The databases were searched for "conference proceedings". Keywords included the Chinese name of the conferences; the Chinese organizer of the conferences (China Medical Informatics Association, the Committee on Information Management, Chinese Hospital Association, China Institutes of Health Information, and the Medical Informatics Branch, Chinese Medical Association); and the English acronym of each conference (CMIAAS, CHINC, CHITEC, and CPMI). Conference proceedings between 1985 and 2015 were retrieved. In addition, we searched for relevant supporting information and biographical references, including the purpose, scope of business, and "call for papers" related to each conference.

Data for AMIA and HIMSS—the two American MI conferences—were extracted from literature reviews [12,13], along with the "letter of welcome" for AMIA and HIMSS and website information regarding conference organizers [14,15].

Data Extraction

We exported information on target conferences from the literature and databases, including the following:

The scale of the conference, including the name of the conference, the organizer, its year of inception, its schedule, the duration of the conference, and the number of attendees.

General information regarding the proceedings, including title, publication year, author affiliation, the name of the conference, and the author's geographic area (extracted using a self-developed tool).

Topics of conference proceedings, which we collected and analyzed, as well as the “call for papers” of the CMIAAS, CHINC, CPMI, and CHITEC (2000-2016) and the organizer’s scope of business.

We used EndNote X2, MS Excel 2011, and Python 2.7 for data processing and analysis.

Results

To further investigate the current level of MI academic research and health information technology (HIT) applications in China, we compared the number of conference sessions, number of attendees, topics presented in conference proceedings, academic research, real-world applications, presentations of academic achievement, number of participating companies, and review criteria among four Chinese mainstream medical informatics conferences, as well as two US conferences. The results are shown in Tables 1 and 2, and specific details are presented in Table 3. The topics of the conference proceedings indicated that these conferences focused on highly similar topics: among them, 10 research fields, including but not limited to, hospital informatics; regional, public, and grassroots health informatics; and telemedicine.

Table 1. Comparison of major characteristics of MI conferences (academic) in United States and China, 2016.

Characteristics	CMIAAS 2015	CPMI 2016	AMIA 2016
Days of meeting, n	2	2	5
Conference sessions, n	42	83	110
Attendees, n	200	300	2500
Participating companies, n	5	17	75
Overall volume of academic achievements[a]	71	248	1349
Presentations of academic achievements, n	1	2	6
Subforums, n	4	4	10
Topics in conference proceedings, n	6	5	20
Pre-symposium tutorials & work group sessions, n	0	7	29
Days of pre-symposium tutorials & working group, n	0	0.5	1.5

[a] Since CPMI 2016 conference papers were still not available from the three major Chinese literature databases (the VIP database of Chinese scientific and technical journals, the Chinese National Knowledge Infrastructure, and the Wanfang database) at the time of writing, CPMI 2015 conference papers were cited here instead.

Table 2. Comparison of major characteristics of MI conferences (industry) in United States and China, 2016.

Characteristics	CHINC 16	CHITEC 16	HIMSS 16
Days of meeting, n	3	3	5
Sessions, n	178	132	950
Attendees, n	3500	5000	40,000
Participating companies, n	138	152	1400
Overall volume of conference proceedings[a]	853	667	0
Concurrent education sessions, n	64	40	300
Subforums, n	23	12	230
Topics in conference proceedings, n	6	10	20
Pre-symposium tutorials & workshops, n	4	0	12
Days of pre-symposium tutorials & working groups, n	1	0	1

[a] Since the conference papers of CHINC 2016 and CHITEC 2016 were still not available from the three major Chinese literature databases (the VIP database of Chinese scientific and technical journals, the Chinese National Knowledge Infrastructure, and the Wanfang database) at the time of writing, the conference papers of CHINC 2015 and CHITEC 2015 were cited here instead.

Table 3. Summary of CMIAAS 2015, CHINC 2016, CHITEC 2016, CPMI 2016, AMIA 2016, and HIMSS 2016.

Characteristics	CMIAAS 2015	CPMI 2016	AMIA 2016	CHITEC 2016	CHINC 2016	HIMSS 2016
Organizer	China Medical Informatics Association	Medical Informatics Branch, Chinese Medical Association	American Medical Informatics Association	China Institutes of Health Information	Committee on Information Management, Chinese Hospital Association	Healthcare Information and Management Systems Society of the US
Inception	1981	1993	1977	2004	1997	1962
Schedule	Every 3 years	Annual	Annual	Annual	Annual	Annual
Meeting duration	1 day	2 day	5 days	3 days	3 days	5 days
Attendees, n	200+	300+	2500+	5000	3500+	40,000+
Composition of attendees	Medical institutions, universities, research institutes	Medical institutions, research institutes, universities	Medical institutions, universities, research institutions, companies	Medical institutions, research institutes, government authorities, enterprises, some universities	Medical institutions, enterprises, some universities and research institutes	Companies, medical institutions, research institutions
Academic achievement	4 forums, 71 conference papers	4 forums, half-day preconference seminar, 248 conference papers, including 13 papers presented at general conference and 48 papers at forums[a]	10 forums, one and half-day preconference seminar, 29 continuing education sessions, and 110 lectures; 151 full-text papers, 50 student articles, 1050 posters, 11 system presentations, 80 abstracts, 7 contests of student-led project design. Rigorous peer review mechanism.	12 forums, 40 continuing education lectures; 667 conference papers, of which 37 were nominated as “outstanding papers”[a]	23 forums, 64 continuing education lectures, 1-day preconference seminar; 853 conference papers, of which 71 were nominated as “outstanding papers”[a]	230 forums and presentations, 300 concurrent education sessions, 1-day preconference seminar
Paper review mechanism	Format review only	Format review only	Rigorous peer review mechanism	Format review only	Format review only Presenters required to submit abstracts and PPT for peer review	
Participating companies, n	5	17	75	152	138	1400+, with review performed between 12 regions
Role of participants	Institution IT staff (48%), university research staff (33%), MI specialists (7%)	Institution IT staff (34%), university research staff (27%), MI specialist (24%)	MDs (27%), research staff (19%), institution IT and administrative staff (17%)	Institution IT staff (57%), governmental personnel, students (29%), MI specialists (6%)	Institution IT staff (78%), software providers, R&D staff (14%), university research staff (5%)	Medical institution leaders (25%), software providers, R&D staff (23%), governmental personnel, students (13%)
Fields covered[b]	B, D, I, F, P, T	L, D, B, F, P	A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T	M, N, K, C, D, I, F, H, Q, T	B, D, I, F, P, T	A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T

[a] Since the conference papers of CHINC 2016, CHITEC 2016 and CPMI 2016 were still not available from the three major Chinese literature databases (the VIP database of Chinese scientific and technical journals, the Chinese National Knowledge Infrastructure, and the Wanfang database) at the time of writing, the conference papers of CHINC 2015, CHITEC 2015, and CPMI 2015 were cited here instead.

[b] A: consumer health informatics; B: clinical information management; C: decision support system; D: electronic medical records; E: medical language processing; F: nursing informatics; G: achievement evaluation; H: public health informatics: I: information retrieval; J: medical cognitive science; K: clinical project management; L: computer-based training; M: coding, classification, and terminology; N: clinical guidelines for computerization; O:

image, robotics, virtual medical treatment; P: signal processing; Q: standards, social, and legal issues; R: dental informatics; S: artificial intelligence; T: telemedicine.

We briefly analyzed the data shown in Table 3 and reached the following primary conclusions regarding the current status of MI development in China and the United States.

First, for academic exchange conferences, we compared the relatively large CPMI conference with AMIA (the annual symposium 2016) and observed that the number of attendees and number of conference papers of the CPMI were approximately only 8% of the numbers of AMIA conference papers. For presentation formats of academic achievements, at AMIA 2016, academic achievements were presented through various means—including full-text papers, posters, student papers, abstracts, design contests, system demonstrations, and descriptions—whereas, at CPMI 2015, academic achievements were presented only as full-text and exchange papers. Of the topics, only 7 fields were covered at CPMI 2015, of which 70.6% of papers were related to information retrieval and hospital informatics. Conversely, all 20 subfields related to MI were covered at AMIA 2016. At the same time, there was remarkable gap between the two conferences in terms of number, size, and degree of attention of the special pre-symposium continuing education sessions. AMIA 2016 scheduled 29 high-level tutorials and working groups in the day and a half before the official symposium, and the average time of each session lasted about 53 minutes. The CPMI 2016 scheduled only 7 tutorials in the afternoon 1 day before the official symposium (ie, less than a quarter of AMIA), and the average time of each session lasted only 25 minutes (ie, only half of AMIA). Therefore, Chinese professionals had fewer continuing education opportunities to attend tutorials offered by conferences. In terms of topics included in the academic conference, AMIA 2016 contained 20 themes, while CMIAAS 2015 and CPMI 2016 contained only 6 and 5, respectively. In terms of participant distribution, AMIA 2016 was mainly composed of medical doctors, researchers, and medical institutions information technology and administrative personnel, while CMIAAS 2015 and CPMI 2016 comprised mainly information technology management of medical institutions, university researchers, and professional MI researchers. Medical doctors were seldom involved.

We then examined the data for HIT application conferences where the gap was also significant. Our comparison of the relatively large CHITEC 2016 and HIMSS 2016 revealed that the number of attendees at CHITEC 2016 was only 12.5% of the number at HIMSS 2016 (5000+ compared to 40,000+), and the number of participating companies at CHITEC 2016 was only 10.9% of the number at HIMSS 2016 (152 compared to 1400+). For academic value-related measures, CHITEC 2016 included only 12 subforums and 40 continuing education lectures, which were, respectively, 5% and 13% of the forums and lectures at HIMSS 2016 (230+ forums and 300+ concurrent education sessions over 5 days). For topics, 10 fields were covered at CHITEC 2015; however, 73.8% of the conference papers were related to hospital informatics or population health informatics, whereas all 20 subfields related to MI were covered at HIMSS 2016. In addition, there was a similar gap between the two conferences in terms of number, size, and degree of attention of pre-symposium continuing learning opportunities. HIMSS 2016 scheduled 12 high-quality preconference education symposiums and workshops, while CHINC 2016 scheduled only 4 (ie, less than half of HIMSS) for 1 day before the official meeting. In terms of subjects, industry conference HIMSS 2016 contained 20 themes, while CHITEC 2016 and CHINC 2016 contained only 10 and 6, focusing mainly on hospital and clinical informatics. In terms of participant distribution, HIMSS 2016 had a similar composition of participants to that of CHITEC 2016 and CHINC 2016, but with more uniform and reasonable distribution.

Third, we noted a significant gap between the academic and continuing education value of Chinese and US MI conference proceedings papers. Specifically, at Chinese HIT conferences, only format review was conducted, due to the lack of well-rounded peer reviewers who had professional training in MI courses. As a consequence, the academic value of the Chinese conference papers was fairly low. In contrast, rigorous peer review was implemented for AMIA 2016. According to data published on its website, the acceptance rate for AMIA's proceedings papers is less than 30%. Thus, AMIA has been recognized by the China Computer Federation Academic Committee as "an important internationally recognized conference" [16].

Overall, we believe that the huge gap between the US and Chinese MI conferences is closely related to their economic strength, population, and health care system. According to 2015 statistics, the US gross domestic product is 1.78 times the Chinese (US $18 trillion and 40 billion [17] vs US $10 trillion and 140 billion [18]). The US population is only 23% of China (314 million [19] vs 1 billion 370 million [20]), but US spending on health care accounted for 16.9% of its gross domestic product [21], which is only 5.95% for China [22]. That is to say, in the same year the health expense per capita was $9451 in the United States [21] and was only $438 in China [22]. The former is 21.6 times the latter. In addition, a huge difference between the Chinese and US health systems is that in the United States, new inventions will soon be put into use [23]. MI is a new interdisciplinary field that has a wide market prospect and naturally has been attracting a lot of investment funds support. Investors want to reduce medical cost through the use of new technologies, forcing medical institutions' cost control under the premise of quality assurance. Therefore, it is not surprising that the number of attendees of HIMSS 2016 is 8 times that of CHITEC 2016, and AMIA 2016's academic achievements are 8.3 times those of CPMI 2016.

In addition, this phenomenon of focusing more on application in HIT and less on research on MI has led to repeated HIT construction in China and a huge resource waste. Contrasted to the United States, there exists a difficulty in applying theoretical research to real-world problems in China, the foundation of theoretical research in medical informatics is weak, and few such studies have been published in China. At present, only a few Chinese teaching institutions have graduate courses or research institutes focused on MI. Moreover, very little investment from the Chinese government has been devoted to basic MI research compared with the huge investments into HIT application, especially MI. We searched and compared the medical information projects funded by Natural Science Foundation of China (NSFC) [24]. As the main channel from the Chinese government to support basic and applied basic research, NSFC is targeted at the whole country for researchers from universities, colleges, research institutions, and enterprises. NSFC is more or less similar to the US National Institutes of Health (NIH). The funded area directly related to MI in NSFC was very limited, with only one domain code set up for independent medical information from 2005, which is code H1814 (medical information system and telemedicine). The number of MI projects funded by NSFC was extraordinarily limited compared to NIH. In 2015, only 3 MI projects (under code H1814) over 3 institutions were funded, with a total amount of USD $284,000; the amount per project was USD $4700. This was in sharp contrast to NIH where 122 MI projects over 82 institutions were funded, with a total of USD $6,514,279. The amount per project was USD $545,199, which were 40.7 times, 27.3 times, 234.2 times, and 5.76 times the Chinese counterparts respectively in 2015. In summary, compared with their American counterparts, funding from government in terms of quantity and amount in basic MI research was very poor in China. The results are shown in Table 4.

Table 4. Comparison of funding support for the academic areas of MI by major funders: NSFC (China) and NIH (US), 2015.

Organization	Funded projects, n	Funded institutions, n	Average amount (USD)	Total amount of funded projects (USD)
NSFC	3	3	94,700	284,000
NIH	122	82	545,199	66,514,279

Discussion

Principal Findings

This cross-sectional study of the quantitative measures of Chinese and US academic and industry exchange conferences showed a large gap between China and the United States in both MI academic research and HIT applications. In addition, there exists a gap in the effect of conferences themselves and proceedings on continuing education to MI researchers and professionals.

Our research showed that MI conferences in China exhibit a significant gap compared to their US counterparts in terms of scale, quality, timeliness, breadth, and depth, from the perspective of discipline development and continuing education. The value of conferences as a continuing education channel to MI professionals in China is also far less than their US counterparts. The training tutorials, lectures, forums, and proceedings of MI conferences in China are more focused on HIT applications and less on the level of academic research.

Strengths and Limitations

This study compared the development of Sino American medical informatics from the perspective of relevant conferences in the two countries. However, the measures to evaluate an overall discipline or industry development are complex. Conferences and meetings are only one part of the field. Furthermore, some information was publically unavailable, such as the demographic information of actual attendees at the Chinese MI conferences or the Chinese participants in the US MI conferences. In addition to conferences, future research should examine capital investment, published papers, patent applications, safety improvement, patient satisfaction, etc, to evaluate the outcomes of MI development in China.

Conclusion

This study demonstrated a significant gap in MI development and continuing education between China and the United States from the perspective of conferences and analyzed the underlying reasons for this gap. There is an urgent need to elevate MI in China in the area of academic discipline and research. Relevant MI professionals need to strengthen mutual understanding and develop pertinent and effective cooperation between China and the United States as well as other global colleagues.

Acknowledgments

This study was supported in part by the National Natural Science Foundation of China (NSFC) (grants #81171426 and #81471756) and the Medical and Health Planning Project of Zhejiang Province of China (grant #2017KY386).

Conflicts of Interest

None declared.

References

1. Yip WC, Hsiao WC, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. The Lancet 2012 Mar;379(9818):833-842. [doi: 10.1016/s0140-6736(11)61880-1]
2. Lei J, Wen D, Zhang X, Li J, Lan H, Meng Q, et al. Enabling Health Reform through Regional Health Information Exchange: A Model Study from China. Journal of Healthcare Engineering 2017 Jan(1):1-9. [doi: 10.1155/2017/1053403]

3. Lei J, Meng Q, Li Y, Liang M, Zheng K. The evolution of medical informatics in China: A retrospective study and lessons learned. Int J Med Inform 2016 Aug;92:8-14. [doi: 10.1016/j.ijmedinf.2016.04.011] [Medline: 27318067]
4. China Medicine Information Association Annual Symposium. URL: http://www.cmia.info/cn/news_detail.asp?id=12227 [accessed 2017-05-02] [WebCite Cache ID 6qA1OkIy9]
5. Chinese Hospital Informatics Network Annual Conference. URL: http://www.chinc.org.cn/ [accessed 2017-05-03] [WebCite Cache ID 6qA1zsuPA]
6. Chinese Health Information Technology Exchange Annual Conference. 2016. URL: http://www.chia-moh.org.cn/zgwsxxxh/jldh15/bjnh_2015.shtml [accessed 2017-05-03] [WebCite Cache ID 6qA21KGmK]
7. China Annual Proceedings of Medical Informatics. 2016. URL: http://csmi2016.medmeeting.org/cn [accessed 2017-06-16] [WebCite Cache ID 6rF3Wbvny]
8. Moher D, Liberati A, Tetzlaff J, Altman DG. Reprint--Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Phys Ther 2009 Sep;89(9):873-880. [Medline: 19723669]
9. VIP China Science and Technology Journal Database. URL: http://www.cnki.net/ [accessed 2017-06-16] [WebCite Cache ID 6rF3sSQcc]
10. Chinese National Knowledge Infrastructure. URL: http://www.cnki.net/ [accessed 2017-06-16] [WebCite Cache ID 6rF3sSQcc]
11. Wanfang Data. 2016. URL: http://g.wanfangdata.com.cn/ [accessed 2017-05-02] [WebCite Cache ID 6qA2cqS8I]
12. Ravvaz K, Kuziemsky C, Koppel R, Kaplan B, Adams SA, Adams MB. AMIA members' "vital signs": what the HIT implementation listserv says about goals for AMIA and for medical informatics. AMIA Annu Symp Proc 2015;2015:1067-1075 [FREE Full text] [Medline: 26958245]
13. Middleton B. President's column: AMIA's policy priorities for 2014. J Am Med Inform Assoc 2014 May;21(3):574 [FREE Full text] [doi: 10.1136/amiajnl-2014-002809] [Medline: 24718742]
14. American Medical Informatics Association 2016 Annual Symposium. 2016. URL: https://www.amia.org/amia2016 [accessed 2017-05-03] [WebCite Cache ID 6qA2qFsBa]
15. HIMSS Annual Conference 2016. URL: http://www.himss.org/event/himss15 [accessed 2017-05-03] [WebCite Cache ID 6qA2skZRz]
16. Chinese Computer Federation Core Journal and Conference Lists. URL: http://www.cnblogs.com/guolei/p/3576366.html [accessed 2017-05-03] [WebCite Cache ID 6qA43lQax]
17. OECD. Gross domestic product (GDP) (indicator). 2017. URL: https://data.oecd.org/gdp/gross-domestic-product-gdp.htm [accessed 2017-05-03] [WebCite Cache ID 6qA3dU99w]
18. Chinese National Database. Gross domestic product (GDP). URL: http://data.stats.gov.cn/ks.htm?cn=C01&zb=A0501 [accessed 2017-05-03] [WebCite Cache ID 6qA468ocO]
19. OECD. Population (indicator). 2017. URL: https://data.oecd.org/pop/population.htm [accessed 2017-05-03] [WebCite Cache ID 6qA48PgPU]
20. Chinese National Database. Population. URL: http://data.stats.gov.cn/ks.htm?cn=C01&zb=A0501 [WebCite Cache ID 6qA468ocO]
21. OECD. Health spending (indicator). 2017. URL: https://data.oecd.org/healthres/health-spending.htm [accessed 2017-05-03] [WebCite Cache ID 6qA4jiBE8]
22. Chinese National Database. Health spending. URL: http://data.stats.gov.cn/easyquery.htm?cn=C01 [accessed 2017-05-03] [WebCite Cache ID 6qA4mCCbR]
23. Deaton A. The great escape: health, wealth, and the origins of inequality. Princeton, NJ: Princeton University Press; 2015.
24. The Natural Science Foundation of China. URL: http://www.nsfc.gov.cn/publish/portal1/ [accessed 2017-05-03] [WebCite Cache ID 6qA4qg1Xn]

Abbreviations

AMIA: American Medical Informatics Association
CHINC: China Hospital Information Network Annual Conference
CHITEC: China Health Information Technology Exchange Annual Conference
CMIAAS: China Medicine Information Association Annual Symposium
CPMI: China Annual Proceeding of Medical Informatics
HIMSS: Healthcare Information and Management Systems Society
HIT: health information technology
MI: medical informatics
NIH: National Institutes of Health
NSFC: Natural Science Foundation of China
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

Edited by G Eysenbach; submitted 11.05.17; peer-reviewed by F Wang, E Moosavinasab; comments to author 31.05.17; revised version received 03.06.17; accepted 08.06.17; published 21.06.17.

Please cite as:
Liang J, Wei K, Meng Q, Chen Z, Zhang J, Lei J
The Gap in Medical Informatics and Continuing Education Between the United States and China: A Comparison of Conferences in 2016
J Med Internet Res 2017;19(6):e224
URL: http://www.jmir.org/2017/6/e224/
doi:10.2196/jmir.8014
PMID:28637638

Original Paper

The Effectiveness of Information Technology-Supported Shared Care for Patients With Chronic Disease: A Systematic Review

Laura Kooij[1], MSc; Wim G Groen[1], PhD; Wim H van Harten[1,2,3], MD, PhD

[1]The Netherlands Cancer Institute, Division of Psychosocial Research and Epidemiology, Amsterdam, Netherlands
[2]University of Twente, Department of Health Technology and Services Research, Enschede, Netherlands
[3]Rijnstate hospital, Arnhem, Netherlands

Corresponding Author:
Wim H van Harten, MD, PhD
The Netherlands Cancer Institute
Division of Psychosocial Research and Epidemiology
Plesmanlaan 121
Amsterdam, 1066CX
Netherlands
Phone: 31 88 005 75
Fax: 31 51 223 22
Email: w.v.harten@nki.nl

Abstract

Background: In patients with chronic disease, many health care professionals are involved during treatment and follow-up. This leads to fragmentation that in turn may lead to suboptimal care. Shared care is a means to improve the integration of care delivered by various providers, specifically primary care physicians (PCPs) and specialty care professionals, for patients with chronic disease. The use of information technology (IT) in this field seems promising.

Objective: Our aim was to systematically review the literature regarding the effectiveness of IT-supported shared care interventions in chronic disease in terms of provider or professional, process, health or clinical and financial outcomes. Additionally, our aim was to provide an inventory of the IT applications' characteristics that support such interventions.

Methods: PubMed, Scopus, and EMBASE were searched from 2006 to 2015 to identify relevant studies using search terms related to shared care, chronic disease, and IT. Eligible studies were in the English language, and the randomized controlled trials (RCTs), controlled trials, or single group pre-post studies used reported on the effects of IT-supported shared care in patients with chronic disease and cancer. The interventions had to involve providers from both primary and specialty health care. Intervention and IT characteristics and effectiveness—in terms of provider or professional (proximal), process (intermediate), health or clinical and financial (distal) outcomes—were extracted. Risk of bias of (cluster) RCTs was assessed using the Cochrane tool.

Results: The initial search yielded 4167 results. Thirteen publications were used, including 11 (cluster) RCTs, a controlled trial, and a pre-post feasibility study. Four main categories of IT applications were identified: (1) electronic decision support tools, (2) electronic platform with a call-center, (3) electronic health records, and (4) electronic communication applications. Positive effects were found for decision support-based interventions on financial and health outcomes, such as physical activity. Electronic health record use improved PCP visits and reduced rehospitalization. Electronic platform use resulted in fewer readmissions and better clinical outcomes—for example, in terms of body mass index (BMI) and dyspnea. The use of electronic communication applications using text-based information transfer between professionals had a positive effect on the number of PCPs contacting hospitals, PCPs' satisfaction, and confidence.

Conclusions: IT-supported shared care can improve proximal outcomes, such as confidence and satisfaction of PCPs, especially in using electronic communication applications. Positive effects on intermediate and distal outcomes were also reported but were mixed. Surprisingly, few studies were found that substantiated these anticipated benefits. Studies showed a large heterogeneity in the included populations, outcome measures, and IT applications used. Therefore, a firm conclusion cannot be drawn. As IT applications are developed and implemented rapidly, evidence is needed to test the specific added value of IT in shared care interventions. This is expected to require innovative research methods.

(J Med Internet Res 2017;19(6):e221) doi:10.2196/jmir.7405

KEYWORDS

review; integrated healthcare systems; health information systems; chronic disease

Introduction

In Europe, 77 % of the disease burden is attributable to chronic diseases. For example, 60 million people live with diabetes [1] and 4-10% suffer from chronic obstructive pulmonary disease (COPD) [2]. Cancer is the leading cause of death in Europe with at least 3 million new cases each year, and cancer survivors are increasingly considered as having a chronic disease [3]. Many health care professionals and various providers are involved during treatment and follow-up of patients with these chronic diseases [3,4]. This inevitably increases fragmentation and can lead to suboptimal care [3]. Coordination of care between multiple professionals caring for patients with chronic disease is essential to guarantee quality of care [4,5]. However, coordination and integration of different professionals is often lacking [3,4]. Shared care is a means to improve integration and is defined as “the joint participation of GPs and hospital consultants in the planned delivery of care for patients with a chronic condition, informed by an enhanced information exchange over and above routine discharge and referral letters” [6]. Shared care can improve care delivery, since it involves a collaboration between primary and specialty care professionals, and this delivery of care is expected to be better than the separation of specialty and primary care [7]. Optimal information exchange between health care professionals is very important for the coordination and continuity of care [8,9]. However, oftentimes information exchange between professionals caring for the same patient is suboptimal [9,10], since professionals lack information [9] or the information is not exchanged on time [10].

The use of information technology (IT) seems promising [10] and is increasingly used to support information exchange [6]. IT can improve information accessibility [4,11-13] and can have a positive effect on safety [14,15]. Additionally, IT can support health care processes and has the potential to improve quality [16] and efficiency of care processes [15,16]. For example, electronic referral can improve the quality of care, access to a professional, and decrease costs [17], and electronic reminders can improve efficiency [4].

An overview of the characteristics and effectiveness of IT-supported shared care interventions is lacking. Previous systematic reviews, such as by Smith et al, provided a total overview of shared care interventions for chronic disease including IT support. They found shared care to be a promising approach but only three IT-supported shared care interventions were reported on. Therefore, there is a need for more evidence, especially as the selected studies were of low methodological quality [7,18]. We presume that since previous reviews [7,18], considerably more IT-supported shared care interventions have been developed and reported on in the literature. Also, IT applications in health care are being developed and implemented at a rapid pace and involve considerable costs. Therefore, we aim to systematically review the state-of-the-art regarding the effectiveness of IT-supported shared care interventions on the care of patients with chronic diseases: diabetes, chronic obstructive pulmonary disease (COPD), (congestive) heart failure, cardiovascular disease (CVD), hypertension, asthma, or cancer. More specifically, we aim to provide an inventory of the effects of shared care, supported by IT, on the care of patients with chronic diseases and to describe the characteristics of the IT applications that support such interventions.

Methods

Information Sources and Search Strategy

Studies were identified by searching the literature in EMBASE, Scopus, and PubMed from January 2006 to September 2015. The search consisted of three concepts: (1) shared care, (2) chronic disease, and (3) IT. Several mesh terms were used for these concepts. The full search string is provided in Multimedia Appendix 1. We also checked the reference lists of included articles to detect other relevant studies focusing on (other) chronic diseases (“snowballing method”). As we wanted to provide a total overview of IT-supported shared care interventions, we selected relevant studies from before 2006 from 2 excellent previous reviews (that searched up until 2006) [7,18].

Eligibility Criteria

For the selection, we used the following eligibility criteria: (1) English-language studies describing a randomized controlled trial (RCT), nonrandomized controlled study or a single-group before and after study; (2) included a shared care intervention; (3) supported by IT; (4) developed specifically for people with a chronic disease: diabetes, COPD, congestive heart failure, CVD, hypertension, or asthma, or cancer; (5) involved health care providers were both primary care physicians (PCPs) operating outside hospitals or physician practices and specialty health care professionals; and (6) study included outcome measures focusing on at least health or clinical, process, provider or professional and financial outcomes.

Study Selection

The first and second authors independently assessed titles and abstracts focusing on the concepts of shared care, type of disease, and study type. IT was not a criterion for the abstract rejection because it was assumed that IT might only be described in the full texts. In the case of ambiguity or when there was no consensus about the abstracts, the full publication was reviewed by the 2 authors. Disagreement was resolved by discussion; when an issue remained unresolved, the decision of a third reviewer (WvH) was decisive. This selection process was similar for the further selection of full texts.

Data Extraction

From the selected studies, we report on study characteristics (year, design, measurement time points, and country), patient population (number and type of disease), intervention characteristics (content), IT characteristics (type of application), outcome measures, and effects. The latter were structured according to provider or professional (proximal), process (intermediate), health or clinical and financial (distal) outcomes. These data items were extracted independently by 2 researchers (LK and WG) and disagreement was resolved by discussion.

Risk of Bias Assessment

We assessed the risk of bias of the included (cluster) RCTs by using the Cochrane risk of bias tool.

The risk of bias was independently assessed by 2 researchers (LK and WG). Disagreement was solved by discussion until consensus was reached. Each aspect and the overall risk of bias of the Cochrane risk of bias tool was graded as high, low, or unclear according to the criteria in the Cochrane handbook [19].

Synthesis of Results

For the reporting of this systematic review, we used the PRISMA guidelines [20]. Results were synthesized in a qualitative way as there were large differences in the types of intervention, target populations, and outcome measures. Due to the diversity of intervention characteristics and outcomes measures, we could not conduct a meta-analysis.

Results

Study Selection

The primary search yielded 4167 results. After title and abstract selection and the removal of duplicates, 29 papers were read in full text. Nine articles met our inclusion criteria. One additional study was found by reviewing the reference lists, and we identified 3 additional studies from the previous systematic review of Smith et al [7,18]. Reasons for excluding studies were inappropriate study design, no available full text, lack of a shared care intervention, and/or lack of IT support. Figure 1 gives a detailed overview of the study selection procedure.

Study Characteristics

In total, we included 8 RCTs, 3 cluster RCTs, 1 controlled trial, and 1 pre-post feasibility study. The 13 manuscripts described 11 unique studies. Two papers by Casas et al [21] and Garcia-Aymerich et al [22] described the same intervention but with different patient populations and outcome measures. Lalonde et al [23] and Santschi et al [24] both described the same intervention but assessing different outcome measures.

The included studies were conducted in Canada (n=2) [23,24], Italy (n=2) [25,26], Scotland (n=3) [27-29], United States (n=2) [30,31], Australia (n=1) [32], Denmark (n=1) [33], Spain (n=1) [22], and Spain and Belgium (n=1) [21]. The intervention groups were mostly compared with a group receiving usual care [21-25,27,29,30,32,33], with a specialist outpatient and a nurse practitioner clinic [28] or in one case through general correspondence by email [31].

Figure 1. Flowchart of the search and selection procedure.

Patient Population Characteristics

Patient populations included patients with COPD (n=2) [21,22]; chronic kidney disease (CKD; n=2) [23,24]; diabetes (n=3) [25,27,31]; hypertension (n=1) [28]; asthma (n=1) [29]; and multiple conditions, such as heart failure, diabetes, (risk for) CVD (n=1) [26], and cancer (n=2) [32,33]. One study did not specify the target population but considered hospital discharges in general, which included all conditions [30].

Intervention Characteristics

The intervention characteristics are presented in Multimedia Appendix 2. There was a large variation in the nature of the interventions, IT applications, and the professionals involved. The primary health care providers who participated in the interventions were PCPs or general practitioners (GPs) (n=11) [21,22,25-33] and pharmacists [23,24]. Specialty care professionals included case managers [21,22,26] and specialists [23,24,28,29,31,33]. However, in 4 interventions the type of specialty care professional was not specified [25,27,30,32].

The objectives varied among the included studies. The majority of the interventions aimed to assess the effectiveness of shared care interventions on the level of distal and/or intermediate outcomes. This included (clinical) patient outcomes [22,24,25,31], sometimes in combination with social and economic settings [27,29]. Other objectives were to study the effects on the number of readmissions, GP contacts with the hospital [21,30], or (diabetes) care outcomes [31]. The impact of a pharmaceutical training and communication network on both distal (pharmaceutical opinions and refusals, clinical outcomes) and proximal outcomes (knowledge and satisfaction of pharmacists) were assessed [24]. Proximal outcomes were also assessed, including tailored information provision to GPs [32] and hospital-based case management [33]. One study aimed to evaluate the feasibility, acceptability, and cost-effectiveness of shared care in comparison with other follow-up approaches [28].

Information Technology (IT) Characteristics

Four types of IT applications can be distinguished: electronic decision support [26,31], electronic health records (EHRs) [25,27-30], an IT platform combined with a call center [21,22], and electronic communication applications [23,24,32,33]. These will be described in more detail in the next section.

Electronic Decision Support

The electronic decision support tools were mainly used for care management, specifically for patients with diabetes [31] and (at risk of) CVD, diabetes, or heart failure [26]. A diabetes electronic management system was used to provide PCPs with decision support aimed at reducing cardiovascular risk in diabetes. PCPs received patient-specific and evidence-based information from endocrinologists via secure-email. Based on this information, PCP and patient discussed how to further continue treatment [31]. Decision support was also used to improve care coordination for patients with diabetes, heart failure, and (at risk of) CVD. Therefore, their care managers were provided with notifications and monitoring instruments [26].

Electronic Health Records

In one nonrandomized controlled study, PCPs and hospital professionals exchanged information via a connected EHR in care for diabetes patients [25]. In a RCT, a connected EHR provided GPs with information regarding their elderly patients' hospital discharge [30]. In 3 cases, the EHRs were "synchronized" and therefore used to store information, which was shared between professionals without technology involved (ie, hardcopies were sent via surface mail). GPs send information to secondary care providers, who add this to their EHR. Consequently GPs periodically receive back the latest updated version [27-29].

IT Platform Including a Web-Based Call Center

An IT platform was used by case managers to manage COPD patients' health records. This platform was connected to a call center that was accessible to PCPs and patients to allow them to contact the case manager. This was part of an intervention aimed at improving health or clinical related outcomes [22] and preventing or reducing of hospitalization [21].

Electronic Communication Applications

IT applications were used to provide (one-way) electronic communication using text, for example, fax and electronic messaging. This information was provided by specialty care professionals to inform primary care physicians about their patients.

Fax was used to inform GPs about chemotherapy and patient specifics [32]. To improve community pharmacists' control over medication-related problems related to CKD, the predialysis clinic provided them with medication and clinical information by fax [23,24]. Case managers, specially trained nurses, aimed to improve the coordination and continuity of care for patients with colorectal cancer. They used electronic messaging to inform GPs about their patients, including contact information [33].

Outcome Measures and Effects

The most striking proximal (professional or provider) [23,32,33], intermediate (process) [21,23,30,31,33], and distal (health or clinical and financial) [22-26,31] results are described for each IT category, and a comprehensive overview is presented in Multimedia Appendix 3.

Electronic Decision Support

A decision support tool described in an RCT was used with the aim to improve metabolic and cardiovascular risk factor control, process of care, and costs for diabetes patients [31]. In a pre-post feasibility study, electronic decision support was used to support care managers in their care of patients with CVD or heart failure [26].

Health or Clinical and Financial Outcomes

Electronic decision support for case management in a pre-post feasibility study showed multiple statistically significant outcomes, for example, days of physical activity per week increased from 2.5 to 4.2 days (*P*<.01) and time from 19.9 to 32.9 min each time, self-monitoring increased by 20-27%. Body mass index (BMI), low-density lipoprotein (LDL), systolic blood pressure (BP), and total cholesterol decreased by 10-20%. Additionally, survey results indicate high levels of satisfaction among physicians, care managers, and patients [26]. However, Smith et al [31] found a significant difference between intervention and usual care for smoking cessation (96.0%, 343/358 in the intervention; 93.0%, 257/277 in the control group; *P*=.04) and aspirin use (66.0%, 238/358 in the intervention; 52.0%, 145/277 in the control group; *P*=.001). A significant effect on metabolic outcomes was not detected. Lower costs were reported benefiting the intervention group. The total mean costs of the intervention were US $6252 compared with US $8564 for the control group (*P*=.02); the outpatient costs for the intervention were US $1842 and US $2129 for the control group (*P*=.04). However, these costs were not specifically related to diabetes care [31].

Electronic Health Records

EHRs were used to (1) share (real-time) data by connecting primary and secondary EHRs [25,30], and (2) synchronize records by collecting professionals' input and storing patients data [27-29].

Provider or Professional Outcomes

Use of an EHR for hypertension patients was compared with specialists' outpatient- and nurse practitioner (NP) follow-up. Sixty-one percent (90/147) of the GPs had a preference to continue shared care and 32% (47/147) preferred shared care over the usual, outpatient- or NP care [28].

Process Outcomes

EHRs were used to inform GPs about hospital discharges. This had no significant effect on the number of PCP visits after discharge nor on rehospitalization rates (18.77%, 351/1870) compared with the control group (19.88%, 356/1791) [30]. The use of “synchronized” EHRs did not seem to affect the number of consultations [27], admissions [27,29], or GP consultations [29] compared with usual care. However, significant effects were noted for the number of patients receiving a complete (medical) review after 2 years (82.4%, 220/267) in comparison with outpatients (54.1%, 146/270) and with nurse practitioner (74.8%, 202/270) follow-ups [28].

Health or Clinical and Financial Outcomes

Clinical information about diabetes patients was shared between GPs and hospital professionals. This had a significant positive effect on various clinical outcomes—for example, glycated hemoglobin (HbA1c), BMI, LDL, and cholesterol [25]. However, the use of “synchronized” health records showed no difference with usual care for most patient-related outcomes, such as metabolic control, psychosocial problems [27], or sleep disturbance [29].

IT platform and Web-Based Call Center

COPD patients’ care managers were accessible for PCPs and patients via a call center that was an integral part of an IT platform in which care managers could also manage health records [21,22].

Health or Clinical and Financial Outcomes

A significant effect on the number of patients without readmissions was detected: 55% (36/65) of patients in the intervention group compared with 33% (30/90) of patients in the control (*P*=.03) [21].

The intervention was also evaluated on a range of clinical, health-related, quality of life and lifestyle aspects; and on self-management medical treatment and patients’ satisfaction. Only statistically significant improvements in dyspnea and BMI were detected. Patients in the intervention had better knowledge of the name of their disease (81%, 17/21 vs 44%, 18/41 in usual care group; *P*=.005), awareness of identification of COPD exacerbations (81%, 17/21 vs 22%, 9/41 in usual care group; *P*<.001), and of exacerbations in early COPD treatment (90%, 19/21 vs 66%, 27/41 in usual care group *P*=.04) than patients receiving usual care—without support from a case manager [22].

Electronic Communication Applications

Information was transferred from secondary to primary care using electronic communication applications, for example, fax [23,24,32,33].

Provider or Professional Outcomes

Overall, PCPs were satisfied about the interventions and information [23,32,33]. For example, GPs receiving extra information about their chemotherapy patients were more confident (7% difference with usual care, *P*=.03) and more satisfied than GPs receiving only the usual correspondence (10% difference with usual care, *P*=.002) [32]. Jefford et al [32] found no effect for GP knowledge, whereas Lalonde et al [23] found that the knowledge of pharmacist in the intervention group increased by more than 30%.

Process Outcomes

The majority of process-related outcomes improved significantly in the included interventions. For example, training combined with a communication network for pharmacists had positive effects on the number of pharmaceutical recommendations [23,24]. GPs were informed by electronic messaging in a care management intervention for patients with colorectal cancer. In the 9 months follow-up period, the case manager intervention showed a decrease in GPs contacting the hospital (*P*=.008), and fewer patients contacted GPs during out-of-hours service (that is not daytime) (*P*=.02) compared with the control group [33].

Health or Clinical and Financial Outcomes

An effect on systolic BP, but not on diastolic or BP control, was reported in one study [24].

Risk of Bias

An overview of the risk of bias is provided in Multimedia Appendix 4. No study was free from the risk of bias. Inherent to the type of intervention blinding either the participants or professionals was not possible. Of the 11 included (cluster) RCTs, 6 studies had adequate random sequence generation; in most cases, computer-generated systems were used. More than half of the studies had a low risk of bias for allocation assessment, mainly because of the use of numbered sealed envelopes. Other aspects that were rated for risk of bias were (1) selective reporting, (2) blinding of outcome assessment, and (3) incomplete outcome data. These items were often not reported, and therefore, score as an unclear risk of bias according to the Cochrane handbook [19].

Discussion

Summary of Evidence

We have systematically reviewed 13 studies focusing on IT-supported shared care for patients with a chronic disease. Overall, there seems to be much merit in IT supported shared care interventions.

The reviewed interventions were supported by four main categories of IT applications: (1) electronic decision support systems, (2) electronic platform and call center, (3) EHR, and (4) electronic communication applications. The main findings of these studies are (1) electronic decision support-based interventions showed a significant positive effect on reducing costs; (2) connected EHRs resulted in more PCP visits, less rehospitalization and better clinical outcomes; and (3) the use of an IT platform resulted in fewer readmissions and positive effects on some health or clinical outcomes. However, it failed to show positive effects on quality of life or doctor visits. Additionally, (4) the use of electronic communication applications showed positive results in terms of PCPs’ satisfaction, confidence [32], and the lower number of GPs contacting the hospital [33]. However, effects on GPs’ knowledge were inconsistent [23,32].

As IT often was only a small part of the intervention, it is hard to determine its real added value in shared care. The reviewed studies varied considerably with regard to the type of intervention, the studied patient population, the IT applications used, and the various outcome measures. As a result of this great variation, and because no study was free from the risk of bias, it is difficult to reliably compare the effects found between the various studies or to make valid generalizations about outcomes that hold true for most chronic patients.

The level of advancedness of included IT applications varied and they have evolved over time. The intervention studies conducted in 1994 [27-29] all used an EHR to manage clinical information and shared this (nonelectronically) between professionals. EHRs have evolved into connected systems that ensure real-time information exchange. Examples are the EHRs used in the studies of Gurwitz et al [30] and Carallo et al [25]. Surprisingly, in 2008 and 2011, fax was still used to transfer information from secondary to primary care, and on the other hand innovative electronic decision support systems were used as well [26,31]). Such “intelligent” systems support professionals in their care of patients, for example, by sending automatic alerts or providing tailored advice. Based on this review we regard this as the most advanced IT application to support shared care.

Comparison With Previous Work

The findings of our review are comparable with previous reviews on shared or integrated care, in the way that these also reported mixed overall results. For example, Smith et al reviewed the effectiveness of shared care studies for patients with chronic disease [7,18]. The results of the included studies were mixed, and therefore, they pose that it was not possible to draw conclusions about the effectiveness of the interventions. Also the reviewed interventions were complex and consisted of multiple elements that precluded attribution of the effects to the different elements. Additionally, in line with our review, the studies were of low methodological quality [7,18].

Ouwens et al [34] reviewed integrated care interventions and also found heterogeneity in patient populations, outcomes, and interventions. Although integrated care appears to be an effective approach, this heterogeneity may lead to incorrect conclusions [34]. A similar conclusion was drawn in the review of Aubin et al on the effects of interventions to improve continuity of follow-up care for cancer patients. In this review, a shared care model was used in 14 of 63 studies, and even though some effects in separate studies were found, no clear conclusions could be drawn because the results were too mixed [35]. Again, just as in the review of Smith et al [7,18], the interventions were complex, which makes it hard to determine which elements of the intervention were effective and which were not. Overall, it seems difficult to determine the real added value of shared care as a result of mixed results and heterogeneity in the included populations and intervention elements.

The use of IT based interventions in these previous reviews was minimal and also a description of the applications and their effects was lacking [7,18]. We found several IT-supported share care interventions but unfortunately, we were unable to draw firm conclusions about the added value of IT because it is not evaluated as a single component.

Future Research

Nowadays, many IT applications have been or are being developed to support health care processes [16], but despite this, we only found a surprising small number of publications analyzing their effectiveness in a controlled study. The rapid development of IT applications for shared care purposes is currently not underpinned by rigorous studies showing its added value. Although in evidence-based medicine the RCT is regarded as the gold standard design, there may be drawbacks in using this design for evaluating health care IT applications. RCTs are, by nature, time and cost intensive and may not be able to keep up with fast developing technologies. In other words, when the results of a RCT are finally available, the IT may be outdated. Other research designs could provide more information and save time [36] and may better keep up with the rapid development of IT. Another approach to reflect the rapid development of IT is to measure the feasibility of an IT intervention in a smaller population within a larger RCT [37].

The assessment of the risk of bias of the studies indicates that there is room for improvement in several areas. For example, concealment of intervention allocation and the lack of blinding of participants were not clearly described. This can mean that the effects are overestimated, and it may also be due to the type of intervention. In future research, researchers should provide estimates (as blinding is seldom possible) about how likely it is that this will influence the outcomes. The measurements should also be described more accurately and preferably distinguish proximal or intermediate or distal outcomes because the exact mechanism of intervention and effects is often unclear. Also better standardization on outcome assessments by using a framework, such as the chronic care model (CCM) may be useful. This is a framework to improve clinical and functional outcomes for patients suffering from a chronic disease, and IT can support that model. Key elements are clinical information systems, including databases and care protocol systems. But other applications are also increasingly used to share data with patients, such as patient portals and PHRs. These are applications to provide patients with their clinical information and the ability to share this information [38,39]. Patients’ needs are important, and care should be focused on patients’ preferences to improve quality of care [40]. Professionals should work together, by means of a shared care model, to meet the needs of patients [41]. In line with this, the definition of shared care may be open to discussion or other care models may be increasingly relevant.

Future research must adapt to these aspects and developments. It is also relevant to examine the processes and time points for which IT will be most valuable in supporting shared care.

Limitations

A limitation of this study is the inclusion of “IT” as a search term in the initial search (title or abstract selection). We therefore might have missed studies that were supported by IT but did not mention this in the title or abstract. Furthermore, although we included a broad range of terms in our search, we may not have retrieved all studies that in fact are a shared care intervention. Our search was conducted from 2006 to January 2015, and we added IT-supported shared care studies from before 2006 from the review of Smith et al [7,18] Although unlikely, we might miss relevant studies from before 2006 that were not reviewed by Smith et al [7,18] because they used slightly different search terms.

Conclusions

Despite the potential benefits of using IT to support shared care in chronic diseases, we found surprisingly few—whether controlled or uncontrolled—studies that substantiated these anticipated benefits. Studies showed a large heterogeneity in the study populations, outcome measures, and IT applications. The reviewed interventions reported many positive effects on (proximal) provider or professionals outcomes (such as GPs' satisfaction and confidence). To a lesser extent, positive effects on intermediate (GPs contacting the hospital) and distal outcomes (costs and readmissions) were also reported. Nonetheless, a firm conclusion cannot be drawn on the effect of IT-supported shared care — especially its clinical effect. As IT applications for shared care are developed and implemented rapidly, we are in need of more and better evidence on the specific added value of IT in shared care interventions, and this is expected to require innovative research methods.

Acknowledgments

This research was sponsored by charities.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy in PubMed.

[PDF File (Adobe PDF File), 31KB - jmir_v19i6e221_app1.pdf]

Multimedia Appendix 2

Study and information technology (IT) characteristics.

[PDF File (Adobe PDF File), 287KB - jmir_v19i6e221_app2.pdf]

Multimedia Appendix 3

Outcome measures and effects; + indicates a positive effect and indicates a negative effect.

[PDF File (Adobe PDF File), 170KB - jmir_v19i6e221_app3.pdf]

Multimedia Appendix 4

Risk of bias.

[PDF File (Adobe PDF File), 17KB - jmir_v19i6e221_app4.pdf]

References

1. World Health Organization, Regional office for Europe Diabetes. WHO. Data and statistics URL: http://www.euro.who.int/en/health-topics/noncommunicable-diseases/diabetes [accessed 2016-01-22] [WebCite Cache ID 6nhwKHM8o]
2. European COPD Coalition. Key facts URL: http://www.copdcoalition.eu/about-copd/key-facts [accessed 2016-01-22] [WebCite Cache ID 6nhwRRbpz]
3. Hewitt M, Greenfield S, Stovall E. From cancer patient to cancer survivor: lost in transition. Washington, DC: National Academies; 2006.
4. Corrigan JM. NAP. 2005. Crossing the quality chasm URL: https://www.nap.edu/login.php?record_id=11378 [accessed 2017-01-22] [WebCite Cache ID 6nhx4kPIV]
5. 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]
6. Hickman M, Drummond N, Grimshaw J. A taxonomy of shared care for chronic disease. J Public Health Med 1994 Dec;16(4):447-454. [Medline: 7880576]
7. Smith SM, Allwright S, O'Dowd T. Does sharing care across the primary-specialty interface improve outcomes in chronic disease? a systematic review. Am J Manag Care 2008 Apr;14(4):213-224 [FREE Full text] [Medline: 18402514]

8. Branger PJ, van der Wouden JC, Schudel BR, Verboog E, Duisterhout JS, van der Lei J, et al. Electronic communication between providers of primary and secondary care. BMJ 1992 Oct 31;305(6861):1068-1070 [FREE Full text] [Medline: 1467688]
9. van Walraven C, Taljaard M, Bell CM, Etchells E, Zarnke KB, Stiell IG, et al. Information exchange among physicians caring for the same patient in the community. CMAJ 2008 Nov 04;179(10):1013-1018 [FREE Full text] [doi: 10.1503/cmaj.080430] [Medline: 18981442]
10. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007 Feb 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
11. Cruz-Correia RJ, Vieira-Marques PM, Ferreira AM, Almeida FC, Wyatt JC, Costa-Pereira AM. Reviewing the integration of patient data: how systems are evolving in practice to meet patient needs. BMC Med Inform Decis Mak 2007 Jun 12;7:14 [FREE Full text] [doi: 10.1186/1472-6947-7-14] [Medline: 17565667]
12. Graetz I, Reed M, Shortell SM, Rundall TG, Bellows J, Hsu J. The next step towards making use meaningful: electronic information exchange and care coordination across clinicians and delivery sites. Med Care 2014 Dec;52(12):1037-1041 [FREE Full text] [doi: 10.1097/MLR.0000000000000245] [Medline: 25304020]
13. García GM, Valenzuela MM, Martínez OJ, Otero LM, Ponz CE, López AT, et al. Results of a coordination and shared clinical information programme between primary care and nephrology. Nefrologia 2011;31(1):84-90 [FREE Full text] [doi: 10.3265/Nefrologia.pre2010.Oct.10615] [Medline: 21270918]
14. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med 2003 Jun 19;348(25):2526-2534. [doi: 10.1056/NEJMsa020847] [Medline: 12815139]
15. Shekelle PG, Morton SC, Keeler EB. Costs and benefits of health information technology. Evid Rep Technol Assess (Full Rep) 2006 Apr(132):1-71. [Medline: 17627328]
16. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [Medline: 16702590]
17. Bodenheimer T. Coordinating care-a perilous journey through the health care system. N Engl J Med 2008 Mar 6;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]
18. Smith SM, Allwright S, O'Dowd T. Effectiveness of shared care across the interface between primary and specialty care in chronic disease management. Cochrane Database Syst Rev 2007 Jul 18(3):CD004910. [doi: 10.1002/14651858.CD004910.pub2] [Medline: 17636778]
19. Higgins JP, Green S. Cochrane. Cochrane handbook for systematic reviews of interventions URL: http://handbook.cochrane.org/part_2_general_methods_for_cochrane_reviews.htm [accessed 2016-01-22] [WebCite Cache ID 6nhxPJqth]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
21. Casas A, Troosters T, Garcia-Aymerich J, Roca J, Hernández C, Alonso A, Members of the CHRONIC Project. Integrated care prevents hospitalisations for exacerbations in COPD patients. Eur Respir J 2006 Jul;28(1):123-130 [FREE Full text] [doi: 10.1183/09031936.06.00063205] [Medline: 16611656]
22. Garcia-Aymerich J, Hernandez C, Alonso A, Casas A, Rodriguez-Roisin R, Anto JM, et al. Effects of an integrated care intervention on risk factors of COPD readmission. Respir Med 2007 Jul;101(7):1462-1469 [FREE Full text] [doi: 10.1016/j.rmed.2007.01.012] [Medline: 17339106]
23. Lalonde L, Normandeau M, Lamarre D, Lord A, Berbiche D, Corneille L, et al. Evaluation of a training and communication-network nephrology program for community pharmacists. Pharm World Sci 2008 Dec;30(6):924-933. [doi: 10.1007/s11096-008-9253-0] [Medline: 18802782]
24. Santschi V, Lord A, Berbiche D, Lamarre D, Corneille L, Prud'homme L. Impact of collaborative and multidisciplinary care on management of hypertension in chronic kidney disease outpatients. J Pharm Health Serv Res 2011 Mar 22;2(2):79-87. [doi: 10.1111/j.1759-8893.2011.00038.x]
25. Carallo C, Scavelli FB, Cipolla M, Merante V, Medaglia V, Irace C, et al. Management of type 2 diabetes mellitus through telemedicine. PLoS One 2015;10(5):e0126858 [FREE Full text] [doi: 10.1371/journal.pone.0126858] [Medline: 25974092]
26. Ciccone MM, Aquilino A, Cortese F, Scicchitano P, Sassara M, Mola E, et al. Feasibility and effectiveness of a disease and care management model in the primary health care system for patients with heart failure and diabetes (Project Leonardo). Vasc Health Risk Manag 2010 May 06;6:297-305 [FREE Full text] [Medline: 20479952]
27. DICE. Integrated care for diabetes: clinical, psychosocial, and economic evaluation. Diabetes integrated care evaluation team. BMJ 1994 May 07;308(6938):1208-1212 [FREE Full text] [Medline: 8180540]
28. McGhee SM, McInnes GT, Hedley AJ, Murray TS, Reid JL. Coordinating and standardizing long-term care: evaluation of the west of Scotland shared-care scheme for hypertension. Br J Gen Pract 1994 Oct;44(387):441-445 [FREE Full text] [Medline: 7748631]

29. Drummond N, Abdalla M, Buckingham J, Beattie J, Lindsay T, Osman L, et al. Integrated care for asthma: a clinical, social, and economic evaluation. Grampian asthma study of integrated care (GRASSIC). BMJ 1994 Feb 26;308(6928):559-564 [FREE Full text] [Medline: 8148678]
30. Gurwitz JH, Field TS, Ogarek J, Tjia J, Cutrona SL, Harrold LR, et al. An electronic health record-based intervention to increase follow-up office visits and decrease rehospitalization in older adults. J Am Geriatr Soc 2014 May;62(5):865-871 [FREE Full text] [doi: 10.1111/jgs.12798] [Medline: 24779524]
31. Smith SA, Shah ND, Bryant SC, Christianson TJ, Bjornsen SS, Giesler PD, et al. Chronic care model and shared care in diabetes: randomized trial of an electronic decision support system. Mayo Clin Proc 2008 Jul;83(7):747-757. [doi: 10.4065/83.7.747] [Medline: 18613991]
32. Jefford M, Baravelli C, Dudgeon P, Dabscheck A, Evans M, Moloney M, et al. Tailored chemotherapy information faxed to general practitioners improves confidence in managing adverse effects and satisfaction with shared care: results from a randomized controlled trial. J Clin Oncol 2008 May 10;26(14):2272-2277. [doi: 10.1200/JCO.2007.14.7710] [Medline: 18467717]
33. Wulff CN, Vedsted P, Søndergaard J. A randomized controlled trial of hospital-based case management in cancer care: a general practitioner perspective. Fam Pract 2013 Feb;30(1):5-13. [doi: 10.1093/fampra/cms050] [Medline: 22952209]
34. Ouwens M, Wollersheim H, Hermens R, Hulscher M, Grol R. Integrated care programmes for chronically ill patients: a review of systematic reviews. Int J Qual Health Care 2005 Apr;17(2):141-146. [doi: 10.1093/intqhc/mzi016] [Medline: 15665066]
35. Aubin M, Giguère A, Martin M, Verreault R, Fitch MI, Kazanjian A, et al. Interventions to improve continuity of care in the follow-up of patients with cancer. Cochrane Database Syst Rev 2012 Jul 11(7):CD007672. [doi: 10.1002/14651858.CD007672.pub2] [Medline: 22786508]
36. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
37. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten Wim H. Development of MijnAVL, an interactive portal to empower breast and lung cancer survivors: an iterative, multi-stakeholder approach. JMIR Res Protoc 2015;4(1):e14 [FREE Full text] [doi: 10.2196/resprot.3796] [Medline: 25614924]
38. Kahn JS, Aulakh V, Bosworth A. What it takes: characteristics of the ideal personal health record. Health Aff (Millwood) 2009;28(2):369-376 [FREE Full text] [doi: 10.1377/hlthaff.28.2.369] [Medline: 19275992]
39. 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]
40. Institute of Medicine, National Academy of Engineering. In: Reid PP, Grossman JH, Fanjiang G, editors. Building a Better Delivery System: A New Engineering/Health Care Partnership. Washington, DC: National Academies Press; 2005.
41. Bazzell JL, Spurlock A, McBride M. Matching the unmet needs of cancer survivors to resources using a shared care model. J Cancer Educ 2015 Jun;30(2):312-318. [doi: 10.1007/s13187-014-0708-9] [Medline: 25103849]

Abbreviations

BMI: body mass index
BP: blood pressure
CKD: chronic kidney disease
COPD: chronic obstructive pulmonary disease
CVD: cardiovascular disease
EHR: electronic health record
GP: general practitioner
IT: information technology
LDL: low-density lipoprotein
NP: nurse practitioner
PCP: primary care physician
PHR: personal health record
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 25.01.17; peer-reviewed by P Gee, J Apolinário-Hagen; comments to author 26.03.17; revised version received 10.04.17; accepted 24.04.17; published 22.06.17.

Please cite as:
Kooij L, Groen WG, van Harten WH
The Effectiveness of Information Technology-Supported Shared Care for Patients With Chronic Disease: A Systematic Review
J Med Internet Res 2017;19(6):e221
URL: http://www.jmir.org/2017/6/e221/
doi:10.2196/jmir.7405
PMID:28642218

Original Paper

Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research

Laura Sbaffi[1*], MA, PhD; Jennifer Rowley[2*], MSc, PhD

[1]Information School, Department of Social Sciences, University of Sheffield, Sheffield, United Kingdom

[2]Information Interaction Research Group, Department of Languages, Information and Communication, Manchester Metropolitan University, Manchester, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Laura Sbaffi, MA, PhD
Information School
Department of Social Sciences
University of Sheffield
211 Portobello
Sheffield, S1 4DP
United Kingdom
Phone: 44 114 2222 ext 686
Fax: 44 114 2222686
Email: L.Sbaffi@sheffield.ac.uk

Abstract

Background: Internet sources are becoming increasingly important in seeking health information, such that they may have a significant effect on health care decisions and outcomes. Hence, given the wide range of different sources of Web-based health information (WHI) from different organizations and individuals, it is important to understand how information seekers evaluate and select the sources that they use, and more specifically, how they assess their credibility and trustworthiness.

Objective: The aim of this study was to review empirical studies on trust and credibility in the use of WHI. The article seeks to present a profile of the research conducted on trust and credibility in WHI seeking, to identify the factors that impact judgments of trustworthiness and credibility, and to explore the role of demographic factors affecting trust formation. On this basis, it aimed to identify the gaps in current knowledge and to propose an agenda for future research.

Methods: A systematic literature review was conducted. Searches were conducted using a variety of combinations of the terms WHI, trust, credibility, and their variants in four multi-disciplinary and four health-oriented databases. Articles selected were published in English from 2000 onwards; this process generated 3827 unique records. After the application of the exclusion criteria, 73 were analyzed fully.

Results: Interest in this topic has persisted over the last 15 years, with articles being published in medicine, social science, and computer science and originating mostly from the United States and the United Kingdom. Documents in the final dataset fell into 3 categories: (1) those using trust or credibility as a dependent variable, (2) those using trust or credibility as an independent variable, and (3) studies of the demographic factors that influence the role of trust or credibility in WHI seeking. There is a consensus that website design, clear layout, interactive features, and the authority of the owner have a positive effect on trust or credibility, whereas advertising has a negative effect. With regard to content features, authority of the author, ease of use, and content have a positive effect on trust or credibility formation. Demographic factors influencing trust formation are age, gender, and perceived health status.

Conclusions: There is considerable scope for further research. This includes increased clarity of the interaction between the variables associated with health information seeking, increased consistency on the measurement of trust and credibility, a greater focus on specific WHI sources, and enhanced understanding of the impact of demographic variables on trust and credibility judgments.

(J Med Internet Res 2017;19(6):e218) doi:10.2196/jmir.7579

KEYWORDS

literature review; trust; health information; information retrieval; web

Introduction

People are increasingly seeking health information and advice online. Statistics from the Pew Research Center show that one in three adults in the United States go online to try to identify a diagnosis or to know more about a health complaint [1]. For the United Kingdom, a report from Oxford Internet Surveys indicates that the number of people going online to seek health information has doubled since 2005, from 37% to 69% [2]. Hence, the Internet is an important source of health information and advice, and the information obtained may have a significant effect on health care decisions and outcomes [3,4] and reduce anxiety and depression while increasing feelings of self-efficacy and empowerment [5-7]. Powell et al [5] suggest that online information is used to educate, reassure, and to sometimes challenge information received from health professionals. However, health information seekers encounter a plethora of Web-based and other sources of health information from a variety of organizations and individuals, and of varying quality, accuracy, and reliability [8-10]. This presents individuals with significant challenges in evaluating and selecting the sources to use, and more specifically, in assessing the credibility and trustworthiness of those sources [11-14]. Yet, in health information seeking, source evaluation is especially important because the information or advice gleaned may have a significant effect on health-related behavior and decisions [10,15]. Furthermore, research suggests that “meagre information evaluation skills” or low health literacy enhance consumers’ vulnerability [16,17], and that individuals with higher eHealth literacy gain more positive outcomes from health information searching including improved self-management of health care needs and more effective interactions with their physician [18]. In addition, Stvilia et al [19] found that consumers may lack the motivation or literacy skills to evaluate the information quality of health Web pages, and Chenet et al [20] suggest that digital inequalities may influence the extent of an individual’s health information repertoires. Hence, research that enhances understanding of the factors that influence the evaluation and selection processes associated with digital health information is important, and can inform the design of information literacy programs, health information content, health information systems, and the design of the interaction between patients and health care professionals.

Given the importance of the evaluation of WHI, and more specifically the role of trust and credibility judgments, there is a growing body of research in this area, and therefore, a continuing need to develop coherent reviews of the field as a basis for further research and to inform practice. Hence, this article undertakes a systematic literature review of the research that features the concepts of trust and credibility in WHI seeking. More specifically, it seeks to answer the following research questions:

RQ1: What is the profile of the research conducted on trust and credibility in WHI seeking?

RQ2: Which factors have been identified as impacting on judgments of trustworthiness and credibility in WHI seeking?

RQ3: Which factors, alongside trust and credibility, have been identified as influencing WHI seeking?

RQ4: What demographic factors affect trust formation in WHI seeking?

Given the importance of the trustworthiness of the health information gathered from digital sources, other authors have conducted literature reviews on this and related topics. Most of these were conducted a few years ago and, while retaining significant reference value, require updating [16,21-25]. There are also more recent reviews that focus on specific aspects of WHI assessment [17,26-29].

Another unique and important aspect of this review is its scope, in that it embraces both trust and credibility. Most prior reviews, and indeed much of the research, distinguishes between trust and credibility. Furthermore, for some authors, trust is defined as an antecedent to credibility (eg, [30-32]), but by other authors (eg, [6,25,33]) trust is viewed as the end result of a process in which credibility is only one of many components. To further add to the ambiguity, other authors regard trust and credibility as interchangeable (eg, [34]), or believe that trustworthiness is one of only two primary dimensions of credibility [26]. Hence, given the interweaving of the concepts of credibility and trust, it is appropriate to include research on both of these aspects in this review.

As suggested above, there are many definitions of trust and credibility in literature. This section provides some examples to orientate the reader and give context to this research. Tseng and Fogg [35] have argued that trust and credibility should not be used interchangeably, nor be considered synonyms; according to these authors, trust “indicates a positive belief about the perceived reliability of, dependability of, and confidence in a person, object or process” (p. 41). Rowley and Johnson [36] stated that trust is “a precursor to successful and effective adoption, interaction and ongoing commitment in the digital space” (p. 494). On the other hand, credibility can be defined as “a characteristic defined by reader judgments, (...) not necessarily equivalent to the actual quality of the information, such as its accuracy or truthfulness” ([37], p. 240). Self [38] regarded credibility as “believability, trust, perceived reliability, and dozens of other concepts and combinations” (p. 421). This paper considers “trust” and “credibility” as two aspects of the same concept, without entering the debate about their different nature [23,25,39,40] because the articles analyzed below use both terms without much discrimination, demonstrating that the issue of defining the true relationship between trust and credibility is still very much unresolved.

Methods

A systematic literature review was carried out to highlight and explore the various aspects and applications of the concept of “trust” in digital health information. The review protocol selected was that proposed by Tranfield et al [41], which advocates an evidence-based approach (ie, the appraisal and synthesis of research evidence). The main advantage of a systematic literature review over a more traditional one (eg, overview or narrative) is the adoption of a "replicable, scientific and transparent process" ([41], p. 209). Key to Tranfield et al's approach are three main stages, defined as planning, conducting, and reporting; this is also consistent with the guidelines proposed by the NHS Centre for Review and Dissemination [42]. The first stage involves the identification of the need for a study on a particular topic. In this study, prior research conducted by the authors in the area of trust formation in health-related Internet searches highlighted the spread and variety of trust and related concepts and uses in the academic literature, making it difficult, at times, to locate and select relevant and targeted research. Exploratory searches were conducted to identify the initial relevant search terms and strings. This process was then refined and reapplied throughout the entire search phase every time new search strings were recognized, in order to maximize coverage. The second stage (ie, the actual building of the dataset) involved the selection of suitable databases. Four multidisciplinary and four health-oriented databases were selected (Table 1). The search was conducted in the article title, abstract, and keywords fields.

Table 1. Review protocol: databases.

Type of database	Database name	Search fields	Number of final records
Multidisciplinary	Scopus	Title and abstract and keywords	932
	ScienceDirect	Title and abstract and keywords	117
	Web of Science	Title and topic	787
	ProQuest	All-except full text	1208
Health-focus	Medline	Title and keywords	313
	PubMed	Title and abstract	254
	PsycINFO	Title and abstract	211
	Cochrane Library	Title and abstract and keywords	5
Total number of records			3827

An exhaustive series of search strings was employed in each database, accounting for synonyms, plurals, hyphenations, and multiple word combinations (eg, "information quality" or "quality of information," and "ehealth" and its variants "e-Health" or "e-health"). Numerous combinations of words and strings were applied with Boolean operators "AND" and "OR" to broaden the search. Over 20 searches were conducted. Examples include:

[online health information] AND [trust]

[digital health information] AND [credibility]

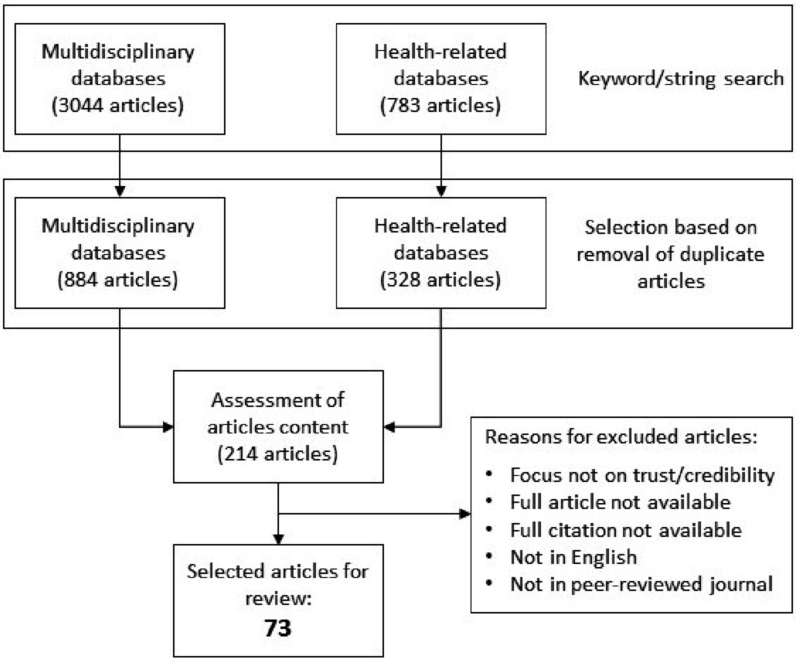
[web health information] AND [information quality]

[health information] AND [trust] AND [online] OR [electronic]

The search exercise, conducted in July 2016 on academic, peer-reviewed literature written in English from the year 2000 onward, identified a total of 3827 records (Table 1). All search results were exported to Microsoft Excel, collated, and all duplicates removed; this reduced the number of records to 1212 unique entries (Figure 1).

Next, an iterative process of refinement and exclusion was carried out on the records to optimize the emphasis on the proposed research topic. Each record (ie, titles, keywords, and abstracts) was scanned for relevance and source, and all articles off-topic, without a full citation, and written in languages other than English were discarded. All conference proceedings papers, books, and book chapters were also discarded, except for 2 conference papers that were retained because of their high citation rate ([43], 434 citations; [44], 180 citations). The final dataset comprised 73 journal articles that were downloaded and fully reviewed by the authors.

Figure 1. Search procedure for articles on trust and credibility in Web-based health information.



Results

This section summarizes key aspects of the bibliographic profile of the literature (RQ1) before further elaboration on the topics covered by this research in response to the remaining research questions (RQ2, RQ3, and RQ4), which are discussed further below.

RQ1: What Is the Profile of the Research Conducted on Trust and Credibility in Web-Based Health Information Seeking?

This section profiles the dataset in terms of the research methodologies used, and the distribution of articles over time, by discipline and by country.

Research Methodology

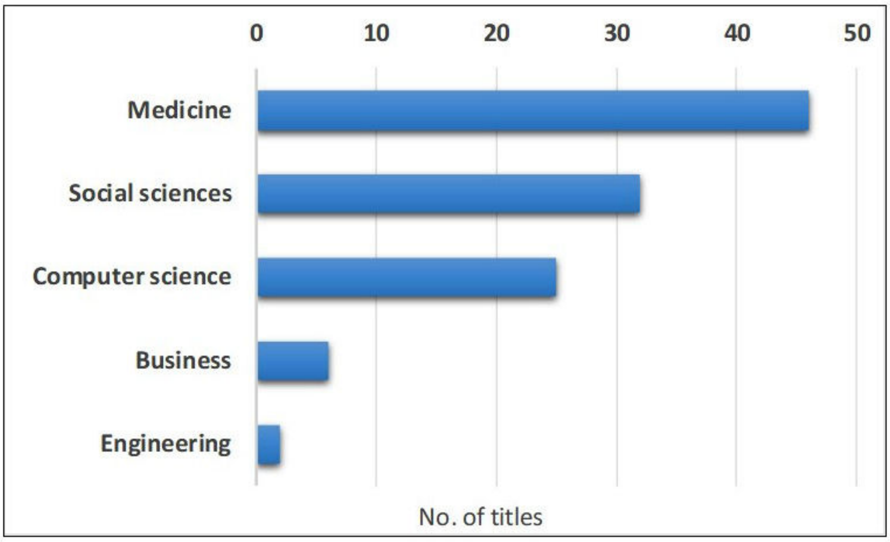
The majority of the titles (71 articles) in the final dataset were empirical studies largely conducted by means of quantitative research instruments (55 articles); nine studies were qualitative, seven used mixed-method approaches, and two were conceptual articles. The participants for the empirical studies were, in almost half of the cases (37/71 articles), adults between 18 and 65 years of age. Undergraduate students were the subject of the research in 20 articles and older people (usually 50-55 years and over) were studied in eight articles. Two studies [31,45] focused on both adults and undergraduate students, two more on high school students aged 11-19 years [12,13], one study was a comparative analysis between older (62+ years) and younger people (<26 years, nonstudents) [46], and one involved content analyses of websites, and therefore, did not require human participants [47].

In addition, there is no agreement on how trust and credibility are measured among different authors. For example, in 2007, Flanagin and Metzgen [31] adopted a 22-item scale to measure the credibility of health websites as a whole. In this scale, the authors included aspects like “colourful,” “aggressive,” “bold,” or “sophisticated” that have not been encountered in prior literature or reutilized since. In the same year, Sillence et al [48] adopted a 25-item scale to measure trust formation that included both information content and website design aspects. A few years later, Corritore et al [11] expanded on the concept of trust formation and acknowledged the complexity of the issue. These authors proposed a 34-item scale which encapsulated concepts of honesty, expertise, predictability, reputation, ease of use, and risk, but they did not include the visual or design aspect of the WHI experience. Recently, Johnson et al [49] attempted to merge previous measurement tools by creating a 55-item scale, including both design and content aspects of WHI. However, in contrast to other research [6,11,16,50], these authors did not explicitly include the concept of the risk associated with information seeking. Therefore, although progress has been made since 2000 in addressing and measuring trust (and credibility) formation, coherence and comprehensiveness are still to be achieved.

Discipline

The final dataset was then categorized in terms of journal subject area, defined according to Scopus’ Scimago Journal and Country Rank (SJR) website (Figure 2). The main disciplines identified were medicine, accounting for more than half of all publications (46 articles), followed by social sciences (32 articles), and computer science (25 articles). The two most recurrent journal titles, with 8 articles each, were the Journal of Health Communication and the Journal of Medical Internet Research (medicine), followed by Social Science and Medicine (medicine or social sciences) and Decision Support Systems (business or computer science), with 3 articles.

Figure 2. Article distribution by journal area of research. This chart has been informed by the Scopus-defined journal subject areas. The overall number of titles exceeds 73 because many journals have more than one subject area.



Country

All articles were also categorized on the basis of the country where the bulk of the research was carried out (Figure 3). he research conducted in the United Stated represents more than half of the entire dataset (40/73 articles, 54.8% of dataset). Research conducted in the United Kingdom followed in second place, but only accounting for 13.7% of the total dataset (10/73 articles). The other eight countries identified make up the remaining 27.4% (20/73 articles) of the dataset. In addition, three articles were the result of collaboration between two countries: Gray et al [12,13], between United Kingdom and United States, and Kitchens et al [10], between China and United States. Grouping the titles in terms of distribution by continent, the Americas account for 61.6% of the research (45/73 articles), Europe for 24.7% (18/73 articles), and Australasia for 13.7% (10/73 articles). This analysis highlights the wide gap in research output between the United States and the rest of the world in this area, indicating that there is still considerable scope for the study of trust in WHI before an exhaustive picture of the situation can be produced, particularly in those developing countries where access to technologies is less well established.

Timeline

There is an established acknowledgment of the importance of research into trust and credibility in WHI seeking, and the number of publications has increased over the last 15 years, although slowly and with setbacks. The number of publications in the dataset never exceeded nine articles in any given year (Figure 4).

Figure 3. Article distribution by journal area of country.

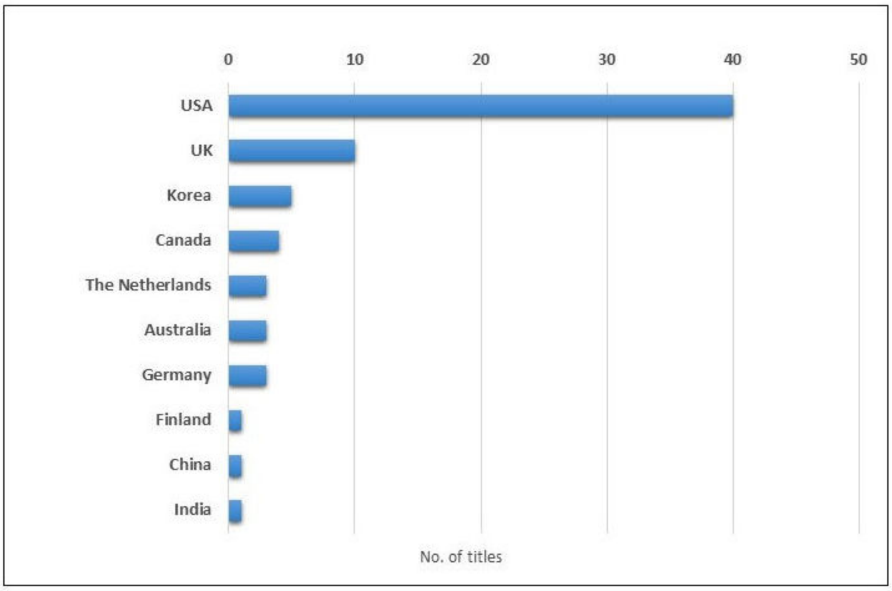
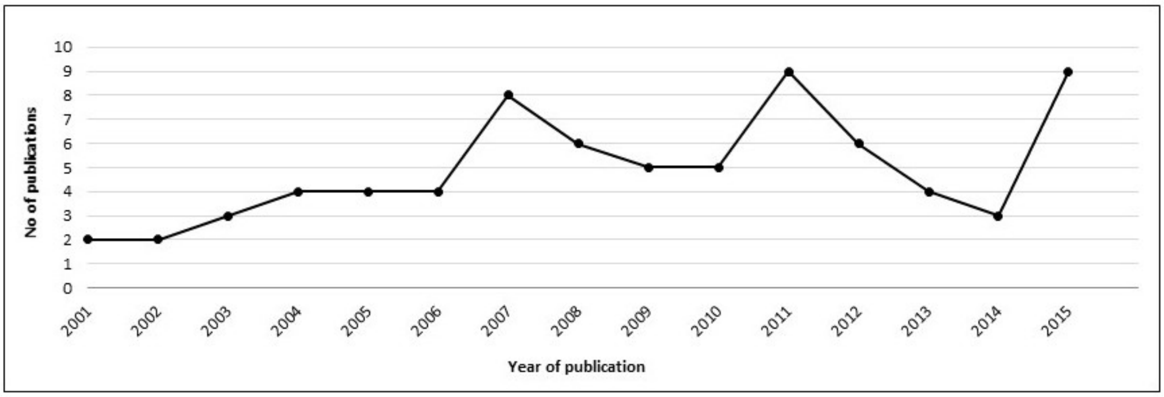


Figure 4. Publication of articles over time.



Discussion

RQ2: Which Factors Have Been Identified as Impacting on Judgments of Trustworthiness and Credibility in Web-Based Health Information Seeking?

The 34 articles in this category argue that trust (or credibility) can be defined as the end result of a series of judgments people apply during their online search processes. Such judgments are usually exercised on both the design features of websites and the content features of the information found. Tables 2 and 3 summarize the factors influencing trust formation. They are coded to show whether they have a positive or negative effect on trust formation, and whether the authors are using the terms "credibility" or "trust" in their research. In two instances [51,52], "Evaluation" ("E") is used throughout, although the research is contextualized with reference to trust and credibility. In tables 2 and 3, factors are listed in order of decreasing number of citations.

There might be some overlap between the features identified as design and those as content, due to the personal interpretation that some authors provide of certain aspects. For example, discussion groups could be classed as a content feature, implying access to information created by other peers rather than professionals, and indeed, Sillence et al's [44] discussion groups are classified as content features. However, in this review, the distinction is based on whether the website offers a link to a discussion forum or group to its users (design feature) or the material discussed in the forum or group is used for study purposes (content feature). In addition, personalization (here reported as design feature) could also encompass both design and content aspects as, for example, cookies could help remember preferences (design), and include opportunities to interact with other people on the site (content). However, in the articles analyzed here, only design aspects were discussed under the umbrella "personalization."

When discussing design features, many authors agree that a clear layout of the website is a strong positive factor influencing trust formation; the presence of contact details and the authority of the owner of the website follow as the second and third most cited factors to consider when addressing trust or credibility issues. The availability of interactive features is another desirable trait of Web design, but it should be noted that the concept of "interactive" has been evolving with time. For example, Walther et al [53] defined a website as "dynamic" if it had changing features such as graphics and fonts. However, Diviani et al [52], more than 10 years later, considered "interactive features" the presence of chat-rooms and fora linked to the website. Interestingly, Fogg et al [43] were the only authors to take into consideration elements such as functionality, customer service, and affiliations.

Much less work has been conducted on negative aspects of design features (Table 2). There is consensus among authors that the presence of explicit advertising on a health website is the least desirable visual feature. Other than this, the majority of negative website design features discussed in the literature emerge from the work of Sillence et al conducted in early 2000s [6,44,48,54]. This is an interesting finding as, although the positive aspects can, in principle, be argued in opposite terms, very little explicit distinction is made between what is good and what is bad in WHI.

Table 2. Factors influencing trust judgments with trust or credibility as dependent variable—design features (C=credibility, T=trust, and E=evaluation).

Features	Impact	Factor	Articles using this factor
Website design features	Positive	Clear layout/design	[6] (T); [8] (C); [31] (C); [33] (T); [43] (C); [44] (T); [46] (C); [48] (T); [49] (T); [51] (E); [54] (T); [55] (T); [56] (C); [57] (T); [58] (T); [59] (T); [60] (C)
		Contact details	[8] (C); [30] (C); [32] (C); [34] (C & T); [52] (E); [61] (T); [62] (C)
		Authority of owner	[8] (C); [30] (C); [32] (C); [34] (C & T); [48] (T); [52] (E); [53] (C)
		Interactive features	[6] (T); [48] (T); [52] (E); [53] (C); [63] (T); [61] (T)
		Brand/logo	[33] (T); [49] (T); [54] (T); [55] (T); [61] (T)
		External links	[8] (C); [32] (C); [34] (C & T); [64] (C)
		Quality seal/endorsement	[8] (C); [30] (C); [52] (E); [58] (T)
		Navigation aids	[6] (T); [30] (C); [32] (C)
		Pictures	[6] (T); [32] (C); [52] (E)
		Discussion groups	[6] (T); [48] (T); [61] (T)
		Privacy policy	[30] (C); [32] (C)
		Identity of sponsor	[34] (C & T); [43] (C)
		Health on the Net (HON) network	[30] (C); [52] (E)
		Personalisation	[57] (T); [59] (T)
		Functionality	[43] (C)
		Customer service	[43] (C)
		Affiliations	[43] (C)
		Easy to access	[57] (T)
		FAQ section	[6] (T)
	Negative	Advertising	[6] (T); [8] (C); [30] (C); [43] (C); [44] (T); [52] (E); [53] (C); [54] (T); [64] (C); [58] (T)
		Slow	[6] (T); [44] (T); [48] (T); [54] (T)
		Complex layout/design	[6] (T); [31] (C); [44] (T)
		Boring layout/design	[6] (T); [31] (C); [44] (T)
		Inappropriate name	[6] (T); [31] (C); [44] (T)
		No navigation aids	[6] (T); [31] (C); [44] (T)
		No/poor search facility	[6] (T); [31] (C); [44] (T)
		Commercial domain	[53] (C); [65] (T)
		Uncaring/unconcerned	[66] (C)
		Textual deficit	[64] (C)

The most widely discussed positive aspect of content features in the literature (Table 3) is the authority of the author, representing the level of expertise of the person or persons writing the information; this is followed by the credibility or trustworthiness of the information (in articles where the main focus was either trust or credibility). For example, Freeman and Spyridakis [37] in their study on the measure of Web-based credibility, described credibility itself as being defined by two main components: trustworthiness and expertise, but Corritore et al [11] found that credibility was one of the direct predictors of trust. The objectivity of the information is also equally important and usually defined by authors in terms of how impartial and unbiased the source is. Ease of use, the fourth most common factor identified in the literature, is here seen as an aspect of content and how the information is made intelligible, and is based on a user's perceptions [11].

Table 3. Factors influencing trust judgments with trust or credibility as dependent variable—content features (C=credibility, T=trust, and E=evaluation).

Feature	Impact	Factor	Articles using this factor
Information content features	Positive	Authority of author	[8] (C); [11] (T); [12] (C); [30] (C); [32] (C); [33] (T); [34] (C & T); [43] (C); [48] (T); [49] (T); [50] (T); [52] (E); [55] (T); [67] (T); [66] (C); [58] (T); [61] (T); [62] (C); [65] (T); [68] (C); [69] (T)
		Credibility/trustworthiness	[11] (T); [12] (C); [33] (T); [49] (T); [53] (C); [54] (T); [64] (C); [55] (T); [66] (C); [61] (T); [62] (C); [69] (T)
		Objectivity	[6] (T); [11] (T); [33] (T); [34] (C & T); [44] (T); [46] (C); [48] (T); [49] (T); [54] (T); [55] (T); [66] (C); [59] (T); [61] (T)
		Ease of use	[11] (T); [33] (T); [48] (T); [49] (T); [51] (E); [52] (E); [64] (C); [63] (T); [55] (T); [57] (T); [61] (T)
		Readability	[6] (T); [8] (C); [43] (C); [44] (T); [52] (E); [54] (T); [63] (T)
		Familiarity	[31] (C); [43] (C); [49] (T); [67] (T); [56] (C); [61] (T); [65] (T)
		Currency (up-to-date)	[8] (C); [30] (C); [34] (C & T); [52] (E); [70] (C); [58] (T)
		Triangulation	[33] (T); [34] (C & T); [49] (T); [55] (T); [58] (T); [65] (T)
		Usefulness	[33] (T); [43] (C); [49] (T); [63] (T); [55] (T); [71] (T)
		References	[30] (C); [52] (E); [70] (C); [65] (T)
		Relevance	[6] (T); [31] (C); [43] (C); [44] (T); [48] (T); [54] (T); [63] (T); [71] (T)
		Recommended by others	[33] (T); [49] (T); [55] (T)
		Accuracy	[43] (C); [52] (E); [71] (T)
		Quality	[50] (T); [57] (T); [59] (T)
		Clarity/understandability	[48] (T); [63] (T); [71] (T)
		Adequacy	[51] (E); [63] (T); [68] (C); [69] (T); [71] (T)
		Quotations	[30] (C); [70] (C)
		Comprehensiveness	[52] (E); [66] (C)
		Statistics	[30] (C); [70] (C)
		Empathy	[12] (C)
	Negative	Risk	[11] (T); [50] (T); [63] (T)
		Inappropriate information	[6] (T); [44] (T); [54] (T)
		Irrelevant information	[6] (T); [44] (T); [54] (T)
		Complex information	[52] (E)
		Bias of information	[43] (C)

Readability, familiarity, currency, triangulation, and usefulness are other factors that have fuelled researchers' interest and which have been the subject of a number of studies. Less common positive content aspects, mentioned only marginally in literature, include the presence of quotations, statistics, and empathy.

The discussion around the negative aspects of the content characteristics of WHI has been much more limited and fragmented. Very few authors have dedicated time to assess what hinders trust (or credibility) in WHI sources. Some researchers have discussed the concept of risk being associated with trust, particularly from a philosophical perspective, stressing that every transaction that requires trust has a degree of associated risk [11,50,63]. The detrimental effects on trust formation deriving from the information being inappropriate or irrelevant, have only been discussed by Sillence and coauthors [6,44,54], with one work focusing on the complexity of the information [52] and another on the bias of the information [43].

RQ3: Which Factors, Alongside Trust and Credibility, Have Been Identified as Influencing Web-Based Health Information Seeking?

Nine articles (Table 4) use trust as an independent variable in their research model and view it as an antecedent to (1) the evaluation of information "quality," or (2) to the intention to "use" the information found. In addition, Escoffery et al [72] report on the ranking that college students give to a number of criteria when looking for WHI.

Information Quality

Bates at al [73] showed how trustworthiness, together with truthfulness, readability, and completeness can influence the quality of information; in addition, more readable health websites can improve the quality perception of the information, but this does not seem to have an effect on the overall trust [74]. Stvilia et al [19] highlighted how the quality of the information is informed by many aspects and trustworthiness is only one of them, together with accuracy, reliability, credibility, and clarity to mention only the most relevant. In a recent article, Kitchens et al [10] showed that the quality of health information is dependent on trust as well as accuracy, but also on the relevance of so-called "referral links" (ie, other websites that link to the main one).

Table 4. Factors, alongside trust or credibility (independent variable), influencing online health information seeking.

Outcome variable	Related article	Major findings
Quality of information	[10]	The quality of health information is dependent on information accuracy and trustworthiness. Quality is then linked to website importance via the number and importance of referral links (ie, links to the website and importance of those websites that link to it)
	[19]	The quality of information is informed by many factors; the first five, in decreasing order of importance, are: accuracy, reliability, credibility, trustworthiness and clarity
	[73]	Trustworthiness, truthfulness, readability and completeness are the main factors influencing the quality of information
	[74]	Making a health website more readable improves quality perception of the information, but there is no effect on trust
Use of the information	[75]	Trust, together with the importance given to written media, concerns for one's own health, importance given to the opinion of HCPs and perceived usefulness, is an antecedent of the intention to use the information
	[76]	Perceived benefit, high interactivity and trust positively affect health information use, as well as satisfaction and long-term loyalty
	[77]	Trust, together with demographics, experience, salience of info and health beliefs, positively influences the intention to use
	[78]	Older people have concerns about the credibility of online health information and the less they trust it, the less they discuss it with their doctors
	[57]	Usability and usefulness contribute to trust formation which, in return, is key to return and reuse a source of information
Factual list	[72]	College students have ranked a series of criteria to consider when looking for online health information and accuracy, credibility and currency of the information are the top three

Intention to Use the Information

Five articles discuss the factors that affect intention to use health information found online. Lemire et al [75] proposed how trust is linked to the use of the information, but only in conjunction with the importance given to the opinion of health care professionals (HCPs) and the perceived usefulness of the information. Lee et al [76] argued that perceived personal benefit, highly interactive websites, and long-term loyalty to specific resources can, together with trust, affect health information use. In two more recent empirical studies, Sheng and Simpson [77] and Pannor Silver [78] discussed the issue of information use from the perspective of older people. Sheng and Simpson [77] claimed that some demographic factors can still influence health information seeking in senior users, particularly age, education, and income, but such factors only bear weight if considered in association with one's own experience with the resources and health beliefs. Pannor Silver [78] highlighted a number of barriers that prevent aging people from trusting and, therefore, using digital health information, particularly in relation to poor e-literacy skills and lack of critical judgment of the quality of the information. Fisher et al [57] conducted a study in Australia on how usability and usefulness of medical websites are crucial to build trust in users and how such trust is then applied to return to and reuse a specific resource.

RQ4: What Demographic Factors Affect Trust Formation in Web-Based Health Information Seeking?

Table 5 lists the 24 papers analyzing demographic aspects in connection with the concept of trust in WHI.

Table 5. Demographic factors influencing trust formation in Web-based health information seeking.

Factor	Hypothesis	Related articles
Gender	Women go online/trust online info more than men	[45,67,79-82]
	No difference between genders	[30,45,83-85]
Education	People with higher education levels go online/trust online info more	[45,79-81,83,85,86]
	No differences due to education level	[87]
Health status	People with poor health go online/trust online info more	[80,86,88]
	People with good health look for offline health info resources more	[80]
	People with good health go online/trust online info more	[79,83]
	Positive relationship between trust and self-efficacy belief in taking care of one's health	[89,90]
	No differences due to health status	[85,87]
Income	People with higher income go online/trust online info more	[45,79-81,83,85,91]
	No differences due to income	[87]
Age	Younger people (25-55 years) go online/trust online info more	[45,79,81,85,86]
	Younger people (25-55 years) trust online info less than older people	[37]
	Older people (usually 55+) do not trust online info and prefer offline resources	[83,84,87,91]
	Articles discussing how young adults (from teenagers to college students) judge and trust online info	[12,13,33,34,49,55,60,71,72,92]
	Articles discussing how elderly people judge and trust online info	[58,77,78,84,93-96]
	Articles comparing young adults versus elderly online behaviour and trust	[46,97]
Health literacy	High health literacy and seeing HCPs often promote online trust	[12,13,98]
	No differences due to health literacy	[85]
Race	White people go online/trust online info more than black people	[81,86]
Parental status	Parents, regardless of gender, behave similarly online	[82]
High/low skilled Web users	The higher the skills the lower the trust in the info	[37,67,97]

Widely discussed in literature are the two extremes of the age spectrum, the young and the old, and their alleged profound differences in selecting, evaluating, and trusting WHI. Old age has been associated with an overall low trust in Web resources [83,84,87,91], as people in this age group rely more on interpersonal relations with physicians, pharmacists, friends, and family [84]; in this respect, the better the quality of doctor-patient communications or other health care providers, the less people tend to go online to look for alternative health resources [86]. Medlock et al [93] found that, indeed, older people depend more on face-to-face interactions with doctors (first) and pharmacists (second), but the use of the Internet is their third chosen source on health information, and aging people who use the Internet more than their peers tend to use all other health information sources as well. McMillan and Macias [94] made the distinction between "health technologists," who are younger seniors using online resources frequently and reporting higher trust in the information retrieved and "health traditionalists," who are the older segment of seniors using the Internet seldom and, therefore, trusting its information much less. Distrust in Web-based resources is associated with difficulties in navigating through large amounts of often confusing information. Zulman et al [95] noted that health websites reporting clearer features that identify easily the source and authorship of the information would promote the use of the Internet among more senior people. This age group acknowledges the importance of selecting trustworthy, credible information but lacks the experience in identifying what trust indicators should be used, hence, simpler and clearer layouts would be easier to navigate. The use of the Internet is associated with an intrinsic trust in the information found and with the perception that searches are easy to carry out, but such perception and the associated feeling of trust decrease with age [63]. A comparative study conducted in Germany by Feufel and Stahl [97] on young and elderly people, emphasized how highly skilled Internet users (identified as young, with high levels of education and more Internet experience) are more confident about the quality of the information retrieved, achieve more focused results, and conduct searches to objectively inform themselves as opposed to low skilled people (identified as older, with lower education levels and patchy Internet experience), who perform less effective searches only to confirm their own preexisting opinions on a topic. Better Internet skills have also been correlated with less trust in the health information [37].

Not surprisingly, at the opposite end of the age scale, young adults, in particular teenage users, seem to experience the same lack of judgmental skills of their much older peers when evaluating health information on the Web. In studies conducted on UK and US adolescents, Gray et al [12,13] reported that young people have difficulties in evaluating online information, which are further exacerbated by low functional, critical, and interactive literacy skills. Very young adults base their assessments of WHI on aesthetics prompts, how familiar they are with a certain website [92], and on how easy it is to access the information [55]. The confidence in their own search strategies contributes to increasing the trust in the information [92]. However, it might be that such confidence is entirely subjective and unjustified, particularly if other older and more expert people are asked to review the same sources [12,13]. Johnson et al [55] found that assessment skills become deeper and more content-oriented with age, so that university students in their third year of study show more discerning judgments than their peers in the first year.

The intermediate age group, here generally defined as comprising people between 25-55 years, shows an online behavior that is more influenced by other demographic factors, particularly education and income levels. As reported in Table 5, several studies have confirmed that people with higher incomes [79-81,83,85,91] and a higher education level [79-81,83,85,86] trust and use digital health information more than people in lower socioeconomic groups. A study conducted in Australia by Dart [45] showed how people from high socioeconomic and university-based backgrounds used the Internet for health information more than disadvantaged people but, in spite of considerable differences in the Internet use, neither group particularly trusted online information. Only Ye [87] reports conflicting findings; according to this research, neither personal capital (ie, income, age, education, and health status) nor social capital (one's network of social interactions) have an impact on trust judgments of health digital information. A study by Dutta-Bergman [99] described how, at the dawn of the Internet era, trust in WHI was segmented according to the sources accessed; for example, younger people with strong health beliefs would trust local doctors' websites the most, whereas less educated people with weaker health beliefs would trust in hospitals more; people with higher income and education would prefer medical universities' websites for their information needs. A more recent trend, due to the widespread use of social media, is the willingness to share health information with others online.

The perception of one's own health status is another determining factor in health information evaluation; however, research has identified conflicting trends. Ye [89] reported that trust in WHI is associated with the ability to assess and look after one's own health (self-efficacy) and with negative emotions due to perceived poor health. This finding is in agreement with Atkinson et al [80] and Hou and Shim [86], who have also demonstrated how people in good health tend to use offline information sources more. This contrasts with findings from Cotten and Gupta [79] and Soederberg et al [107] who assert that a good health status is indicative of more online activity. Other authors [85,87] have shown that there is no relationship between perceived health status and trust in health information.

A considerable number of the studies reviewed cover the issue of gender, but mostly as part of a larger set of demographic attributes (see Table 5); some authors agree that women use and trust the Internet more than men when it comes to health problems [45,67,79-82]. However, a number of studies showed no differences [66,83-85], leaving the gender debate open and in need of further research. Only two studies focus on gender differences in online health searches behavior: one from Korea [67]and one from the United States [82]. In contrast, the role of gender in influencing online trust judgments, in contexts other than health, has received more attention (eg, [100,101]).

The Remaining Themes

Four articles focused on information sources and trust or credibility. LaValley et al [102] reported that almost 3/4 of Americans use commercially sponsored websites to satisfy their health information requirements, but stressed how different website types have different reasons for sponsoring health information, which may affect the website's content and design. Hu and Sundar [103] showed that websites are preferred to bulletin boards, home pages, and blogs, and credibility was strongly associated with users' perceptions of the relevance of the message associated to a certain source. Stoerger [47] found that websites with lower credibility levels were associated with a lot of interactive features and advertising. In an earlier article by Rains [70], based on the Health Information National Trends Survey (HINTS), the author demonstrated a link between a person's trust in mass media and one's health care provider and an increased use in WHI resources.

Two articles discussed trust from a theoretical perspective. Sillence and Briggs [104] explained how ubiquitous computing [105] has long-term and still unknown implications for the health care sector because it produces a shift of people's trust from physicians to artificial agents (ie, computers). Singal and Shruti [106] claimed that there is no standard that exemplifies how to make trust decisions in health; however the authors envisaged to develop a technique to rerank search results using trust as a determining factor so that the more trustworthy a website, the higher its position in a result list.

Toward a Future Research Agenda

This article reports on a systematic literature review of the peer reviewed literature exploring the concept of trust and credibility formation in WHI seeking. The review demonstrates that there is still no consensus of the relationship between the terms "trust" and "credibility". This review also demonstrates that trust and credibility have been investigated both as the dependent variable, representing the end product of a series of cues and factors influencing the process of information seeking and as independent variable, alongside other variables associated with the quality or use of information. In addition, other research has examined the impact of demographic variables such as gender and age, on trust and credibility judgments in WHI seeking. Nevertheless, while there is a growing body of research in this area, given the importance of the trustworthiness of WHI, there is considerable scope for future research and theoretical development in this area. This includes:

Conceptual or theoretical: It would be of considerable benefit to be able to arrive at a consensus on the relationships between the various variables associated with research into WHI seeking. For example, there needs to be further consideration as to whether trust and credibility or information quality are the most important outcome variables. We propose that consideration of the context may be important in differentiating between trust and credibility, with, for example, trust being the appropriate term to use when information is used to inform a decision or action. This stance would also necessitate the development of an improved understanding of the relationship between trust and risk in digital health information seeking. Having established, or at least, further explored the relationships between trust and credibility, it will be important to increase understanding of the key influencing factors, and the extent to which context might impact on these.

Methods and measurement: Most research on trust and credibility in WHI seeking has adopted a quantitative approach. The purpose of quantitative studies is typically to test theory. However, in the absence of a consensus on definitions of variables and the dominant relationships between them, such studies are unlikely to lead to an integrated and coherent body of knowledge. This is further undermined by the considerable variability in the measurement scales used for trust and credibility and related variables. This needs to be addressed by a much greater number of qualitative studies that offer deeper insights into the context, processes, and judgments associated with WHI seeking and the relationship between these.

Topics: Most studies have investigated the factors that influence trust judgment in relation to WHI in general. Hence, there is scope for more studies that take into consideration judgments on specific health information sources, including specific websites and social media platforms, and the role of the owner and community associated with these platforms in influencing trust judgments. This review has not embraced research on trust regarding health information received in social support groups; this would also be an important agenda for future research. In addition, most prior research has privileged factors that have a positive effect on trust judgments, with few reporting on those factors (such as advertising) that might undermine trust. Finally, the dynamic between trust and risk deserves greater attention, particularly with regard to patients’ perceptions of the seriousness of their complaint.

Impact of demographic variables: There is evidence that various demographic variables (eg, age, income, and gender) may influence WHI-seeking behaviors, but the evidence that this also impacts on their trust judgments is scant. Further research is needed in this area. In particular, the research on disadvantaged groups has focused on identifying their needs, but little work has been done on how these needs and the ability of members of these groups to discriminate between trustworthy and untrustworthy information can be enhanced. In addition, whereas some research has been conducted on the role of gender on trust formation in information seeking, which makes links to the role of the psycho-social context, there is considerable scope for further research into the role of this context in trust formation in health information seeking. Finally, the current research base focuses on health information seekers in the United States, United Kingdom, and Australia. There is a need for research in countries where not only technological differences can play a role in information seeking, but also culture and, more specifically, trust formation and relationships with health organizations and professionals are likely to differ from non-western countries.

Conflicts of Interest

None declared.

References

1. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2017-06-03] [WebCite Cache ID 6qwgSf3xS]
2. Dutton W, Blank G. Worldinternetproject. 2013. Cultures of the Internet: The Internet in Britain URL: http://www.worldinternetproject.net/_files/_Published/23/820_oxis2011_report.pdf [accessed 2017-06-03] [WebCite Cache ID 6qwgnglaX]
3. Fox S. Pew Research Center. 2011 May 12. The Social Life of Health Information, 2011 URL: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2017-06-12] [WebCite Cache ID 6r9pbO61u]
4. Xiao N, Sharman R, Rao HR, Upadhyaya S. Factors influencing online health information search: an empirical analysis of a national cancer-related survey. Decis Support Syst 2014;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
5. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
6. 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]
7. Ybarra M, Suman M. Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. Health Educ Res 2008 Jun;23(3):512-521 [FREE Full text] [doi: 10.1093/her/cyl062] [Medline: 16880222]
8. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
9. Fergie G, Hunt K, Hilton S. What young people want from health-related online resources: a focus group study. J Youth Stud 2013 Aug;16(5):579-596 [FREE Full text] [doi: 10.1080/13676261.2012.744811] [Medline: 24748849]

10. Kitchens B, Harle CA, Li S. Quality of health-related online search results. Decis Support Syst 2014 Jan;57:454-462. [doi: 10.1016/j.dss.2012.10.050]
11. Corritore C, Wiedenbeck S, Kracher B, Marble R. Online trust and health information websites. Int J Hum Comput Interact 2012;8(4):92-115. [doi: 10.4018/jthi.2012100106]
12. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
13. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. The Internet: a window on adolescent health literacy. J Adolesc Health 2005 Sep;37(3):243. [doi: 10.1016/j.jadohealth.2004.08.023] [Medline: 16109345]
14. Metzger M, Flanagin A. Credibility and trust of information in online environments: the use of cognitive heuristics. J Pragmat 2013;59:210-220. [doi: 10.1016/j.pragma.2013.07.012]
15. Zhang Y. Searching for specific health-related information in MedlinePlus: behavioral patterns and user experience. J Assoc Inf Sci Tec 2013 Oct 04;65(1):53-68. [doi: 10.1002/asi.22957]
16. 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]
17. Diviani N, van den Putte B, Giani S, van Weert JCM. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
18. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
19. Stvilia B, Mon L, Yi YJ. A model for online consumer health information quality. J Am Soc Inf Sci 2009 Sep;60(9):1781-1791. [doi: 10.1002/asi.21115]
20. Chen W, Lee K, Straubhaar JD, Spence J. Getting a second opinion: social capital, digital inequalities, and health information repertoires. J Assoc Inf Sci Technol 2014 Apr 30;65(12):2552-2563. [doi: 10.1002/asi.23130]
21. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [Medline: 15667044]
22. O'Grady L. Future directions for depicting credibility in health care web sites. Int J Med Inform 2006 Jan;75(1):58-65. [doi: 10.1016/j.ijmedinf.2005.07.035] [Medline: 16125999]
23. Rieh SY, Danielson DR. Credibility: a multidisciplinary framework. Ann Rev Info Sci Technol 2007;41(1):307-364. [doi: 10.1002/aris.144.v41:1]
24. Adams SA. Revisiting the online health information reliability debate in the wake of “web 2.0”: an inter-disciplinary literature and website review. Int J Med Inform 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006] [Medline: 20188623]
25. Vega L, Montague E, DeHart T. Trust in health websites: a review of an emerging field. 2010 Presented at: Proc 1st ACM Int Health Informatics Symp; 2010; Arlington p. 700-709. [doi: 10.1145/1882992.1883100]
26. Choi W, Stvilia B. Web credibility assessment: conceptualization, operationalization, variability, and models. J Assoc Inf Sci Technol 2015 May 13;66(12):2399-2414. [doi: 10.1002/asi.23543]
27. Shah A, Ravana S, Hamid S, Ismail M. Umexpert.um. 2015. Web credibility assessment: affecting factors and assessment techniques URL: https://umexpert.um.edu.my/file/publication/00006249_125919.pdf [accessed 2017-06-03] [WebCite Cache ID 6qwkMPMgb]
28. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assoc Inf Sci Technol 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
29. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun;22(2):355-369. [doi: 10.1177/1460458214559432] [Medline: 25518944]
30. Hong T. The influence of structural and message features on Web site credibility. J Am Soc Inf Sci 2005 Jan 01;57(1):114-127. [doi: 10.1002/asi.20258]
31. Flanagin AJ, Metzger MJ. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. New Media Soc 2007 Apr;9(2):319-342. [doi: 10.1177/1461444807075015]
32. Rains SA, Karmikel CD. Health information-seeking and perceptions of website credibility: examining Web-use orientation, message characteristics, and structural features of websites. Comput Human Behav 2009 Mar;25(2):544-553. [doi: 10.1016/j.chb.2008.11.005]
33. Rowley J, Johnson F, Sbaffi L. Students' trust judgements in online health information seeking. Health Informatics J 2015 Dec;21(4):316-327. [doi: 10.1177/1460458214546772] [Medline: 25193449]
34. Hargittai E, Fullerton L, Menchen-Trevino E, Thomas K. Trust online: young adults' evaluation of web content. Int J Commun 2010;4(1):448-494.
35. Tseng S, Fogg BJ. Credibility and computing technology. Commun ACM 1999;42(5):39-44. [doi: 10.1145/301353.301402]
36. Rowley J, Johnson F. Understanding trust formation in digital information sources: the case of Wikipedia. J Info Sci 2013 Aug;39(4):494-508. [doi: 10.1177/0165551513477820]
37. Freeman K, Spyridakis J. An examination of factors that affect the credibility of online health information. Tech Comm 2004;51(2):239-263 [FREE Full text]

38. Self C. Credibility. In: An integrated approach to communication theory and research. Mahwah, NJ: Lawrence Erlbaum Associates; 1996:421-441.
39. Marsh S, Dibben MR. The role of trust in information science and technology. Ann Rev Info Sci Tech 2005 Jan 31;37(1):465-498. [doi: 10.1002/aris.1440370111]
40. Vega LC, Montague E, Dehart T. Trust between patients and health websites: a review of the literature and derived outcomes from empirical studies. Health Technol (Berl) 2011 Nov 18;1(2-4):71-80 [FREE Full text] [doi: 10.1007/s12553-011-0010-3] [Medline: 22288026]
41. Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Man 2007 Jan 23;14(3):207-222. [doi: 10.1111/1467-8551.00375]
42. NHS Centre for Reviews and Dissemination. CRD Report Number 4 2nd ed, York. 2009. Systematic Reviews: CRD's guidance for undertaking reviews in health care URL: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf [accessed 2017-06-12] [WebCite Cache ID 6r9qfAElB]
43. Fogg B, Soohoo C, Danielson D, Marable L, Stanford J, Tauber E. How do users evaluate the credibility of Web sites? A study with over 2,500 participants. 2003 Presented at: Conference on Designing for user experiences; 2003; San Francisco p. 1-15. [doi: 10.1145/997078.997097]
44. Sillence E, Briggs P, Fishwick L, Harris P. Trust and mistrust of online health sites. 2004 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2004; Vienna p. 663-670. [doi: 10.1145/985692.985776]
45. Dart J. The internet as a source of health information in three disparate communities. Aust Health Rev 2008 Aug;32(3):559-569. [Medline: 18666885]
46. Liao Q, Fu W. Age differences in credibility judgments of online health information. 2012 Presented at: 2nd ACM SIGHIT International Health Informatics Symposium; 2012; Miami p. 353-362. [doi: 10.1145/2110363.2110404]
47. Stoerger S. I'm not a doctor, but I play one on the web: credibility, funding, and interactivity features on health organization websites. Proc Am Soc Info Sci Tech 2008 Oct 24;44(1):1-5. [doi: 10.1002/meet.1450440314]
48. Sillence E, Briggs P, Harris P, Fishwick L. Going online for health advice: changes in usage and trust practices over the last five years. Interact Comput 2007 May;19(3):397-406 [FREE Full text] [doi: 10.1016/j.intcom.2006.10.002]
49. Johnson F, Rowley J, Sbaffi L. Modelling trust formation in health information contexts. J Info Sci 2015 Aug;41(4):415-429. [doi: 10.1177/0165551515577914]
50. Yi MY, Yoon JJ, Davis JM, Lee T. Untangling the antecedents of initial trust in Web-based health information: the roles of argument quality, source expertise, and user perceptions of information quality and risk. Dec Sup Sys 2013 Apr;55(1):284-295. [doi: 10.1016/j.dss.2013.01.029]
51. Thielsch MT, Blotenberg I, Jaron R. User evaluation of websites: from first impression to recommendation. Interact Comput 2013 Jun 04;26(1):89-102. [doi: 10.1093/iwc/iwt033]
52. Diviani N, van den Putte B, Meppelink CS, van Weert J. Exploring the role of health literacy in the evaluation of online health information: insights from a mixed-methods study. Patient Educ Couns 2016 Jun;99(6):1017-1025. [doi: 10.1016/j.pec.2016.01.007] [Medline: 26817407]
53. Walther JB, Wang Z, Loh T. The effect of top-level domains and advertisements on health web-site credibility. J Med Internet Res 2004 Sep 03;6(3):e24 [FREE Full text] [doi: 10.2196/jmir.6.3.e24] [Medline: 15471750]
54. Sillence E, Briggs P, Harris P, Fishwick L. A framework for understanding trust factors in web-based health advice. Int J Hum Comput Stud 2006;64(8):697-713.
55. Johnson F, Sbaffi L, Rowley J. Students' approaches to the evaluation of digital information: insights from their trust judgments. Brit J Educ Technol 2015 Jul 07;47(6):1243-1258. [doi: 10.1111/bjet.12306]
56. Toms EG, Latter C. How consumers search for health information. Health Informatics J 2007 Sep;13(3):223-235. [doi: 10.1177/1460458207079901] [Medline: 17711883]
57. Fisher J, Burstein F, Lynch K, Lazarenko K. "Usability + usefulness = trust": an exploratory study of Australian health web sites. Int Res 2008;18(5):477-498.
58. Czaja SJ, Sharit J, Nair SN, Lee CC. Older adults and internet health information seeking. Proc Human Factors Ergon Soc Ann Meeting 2009 Oct 01;53(2):126-130. [doi: 10.1177/154193120905300204]
59. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
60. Pariera K. Information literacy on the web how college students use visual and textual clues to assess credibility on health websites. Comm Inf Lit 2012;6(1):34-48 [FREE Full text] [doi: 10.7548/cil.v6i1.184]
61. Sillence E, Briggs P. Please advise: using the Internet for health and financial advice. Comp Human Behav 2007 Jan;23(1):727-748. [doi: 10.1016/j.chb.2004.11.006]
62. Freeman KS, Spyridakis JH. Effect of contact information on the credibility of online health information. IEEE Trans Profess Commun 2009 Jun;52(2):152-166. [doi: 10.1109/TPC.2009.2017992]
63. Song J, Zahedi F. Trust in health infomediaries. Dec Sup Sys 2007 Mar;43(2):390-407. [doi: 10.1016/j.dss.2006.11.011]
64. Dubowicz A, Schulz PJ. Medical information on the internet: a tool for measuring consumer perception of quality aspects. Interact J Med Res 2015;4(1):e8 [FREE Full text] [doi: 10.2196/ijmr.3144] [Medline: 25835333]

65. Huntington P, Nicholas D, Gunter B, Russell C, Withey R, Polydoratou P. Consumer trust in health information on the web. Aslib Proc 2004;56(6):373-382. [doi: 10.1108/00012530410570417]
66. Hong T. Contributing factors to the use of health-related websites. J Health Commun 2006 Mar;11(2):149-165. [doi: 10.1080/10810730500526679] [Medline: 16537285]
67. Lim S, Lee S, Kim D. An empirical study of intention of usage of health information on the Internet: comparison by gender. J Korea Soc IT Serv 2011 Sep 30;10(3):77-94. [doi: 10.9716/KITS.2011.10.3.077]
68. Eastin M. 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]
69. Stella N. Credibility and consistency earn users' trust. Internet Healthc Strateg 2003 Feb;5(2):8-9. [Medline: 12674074]
70. Rains SA. Perceptions of traditional information sources and use of the world wide web to seek health information: findings from the health information national trends survey. J Health Commun 2007;12(7):667-680. [doi: 10.1080/10810730701619992] [Medline: 17934943]
71. Lim SH, Kim D. The role of trust in the use of health infomediaries among university students. Inform Health Soc Care 2012 Mar;37(2):92-105. [doi: 10.3109/17538157.2011.647933] [Medline: 22360740]
72. Escoffery C, Miner KR, Adame DD, Butler S, McCormick L, Mendell E. Internet use for health information among college students. J Am Coll Health 2005;53(4):183-188. [doi: 10.3200/JACH.53.4.183-188] [Medline: 15663067]
73. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the Internet. Med Inform Internet Med 2006 Mar;31(1):45-52. [doi: 10.1080/14639230600552601] [Medline: 16754366]
74. Bates BR, Romina SM, Ahmed R. The effect of improved readability scores on consumers' perceptions of the quality of health information on the internet. J Cancer Educ 2007;22(1):15-20. [doi: 10.1080/08858190701348067] [Medline: 17570803]
75. Lemire M, Paré G, Sicotte C, Harvey C. Determinants of Internet use as a preferred source of information on personal health. Int J Med Inform 2008 Nov;77(11):723-734. [doi: 10.1016/j.ijmedinf.2008.03.002] [Medline: 18434246]
76. Lee H, Yang T, Tsai C. To explore the social psychological factors influencing outcomes of health websites. Ad Sci Lett 2012;13(1):774-779 [FREE Full text]
77. Sheng X, Simpson PM. Health care information seeking and seniors: determinants of Internet use. Health Mark Q 2015 Mar;32(1):96-112. [doi: 10.1080/07359683.2015.1000758] [Medline: 25751321]
78. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
79. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
80. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
81. Chung JE. Patient-provider discussion of online health information: results from the 2007 Health Information National Trends Survey (HINTS). J Health Commun 2013;18(6):627-648. [doi: 10.1080/10810730.2012.743628] [Medline: 23590202]
82. Stern MJ, Cotten SR, Drentea P. The separate spheres of online health: gender, parenting, and online health information searching in the information age. J Fam Issues 2012 Oct;33(10):1324-1350. [doi: 10.1177/0192513X11425459]
83. Miller LM, 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]
84. Chaudhuri S, Le T, White C, Thompson H, Demiris G. Examining health information-seeking behaviors of older adults. Comput Inform Nurs 2013 Nov;31(11):547-553 [FREE Full text] [doi: 10.1097/01.NCN.0000432131.92020.42] [Medline: 23974574]
85. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun 2010;15(Suppl 3):279-293. [doi: 10.1080/10810730.2010.522700] [Medline: 21154099]
86. Hou J, Shim M. The role of provider-patient communication and trust in online sources in Internet use for health-related activities. J Health Commun 2010;15(Suppl 3):186-199. [doi: 10.1080/10810730.2010.522691] [Medline: 21154093]
87. Ye Y. Correlates of consumer trust in online health information: findings from the Health Information National Trends Survey. J Health Commun 2011 Jan;16(1):34-49. [doi: 10.1080/10810730.2010.529491] [Medline: 21086209]
88. 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]
89. Ye Y. A path analysis on correlates of consumer trust in online health information: evidence from the health information national trends survey. J Health Commun 2010;15(Suppl 3):200-215. [doi: 10.1080/10810730.2010.522687] [Medline: 21154094]
90. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]
91. Kwon JH, Kye S, Park EY, Oh KH, Park K. What predicts the trust of online health information? Epidemiol Health 2015;37:e2015030 [FREE Full text] [doi: 10.4178/epih/e2015030] [Medline: 26212505]

92. Kim H, Park S, Bozeman I. Online health information search and evaluation: observations and semi-structured interviews with college students and maternal health experts. Health Info Libr J 2011 Sep;28(3):188-199 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00948.x] [Medline: 21831218]
93. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
94. McMillan SJ, Macias W. Strengthening the safety net for online seniors: factors influencing differences in health information seeking among older internet users. J Health Commun 2008 Dec;13(8):778-792. [doi: 10.1080/10810730802487448] [Medline: 19051113]
95. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
96. Robertson-Lang L, Major S, Hemming H. An exploration of search patterns and credibility issues among older adults seeking online health information. Can J Aging 2011 Dec;30(4):631-645. [doi: 10.1017/S071498081100050X] [Medline: 22085455]
97. Feufel MA, Stahl SF. What do web-use skill differences imply for online health information searches? J Med Internet Res 2012;14(3):e87 [FREE Full text] [doi: 10.2196/jmir.2051] [Medline: 22695686]
98. Ha S, Lee Y. Determinants of consumer-driven healthcare: self-confidence in information search, health literacy, and trust in information sources. Int J Pharm Healthc Mark 2011;5(1):8-24. [doi: 10.1108/17506121111121550]
99. Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003;5(3):e21 [FREE Full text] [doi: 10.2196/jmir.5.3.e21] [Medline: 14517112]
100. Porter CE, Donthu N, Baker A. Gender differences in trust formation in virtual communities. J Mark Theory Pract 2012 Jan 25;20(1):39-58. [doi: 10.2753/MTP1069-6679200103]
101. Riedl R, Hubert M, Kenning P. Are there neural gender differences in online trust? An fMRI study on the perceived trustworthiness of eBay offers. MIS Quart 2010;34(2):397-428.
102. LaValley SA, Kiviniemi MT, Gage-Bouchard EA. Where people look for online health information. Health Info Libr J 2016 May 21;34(2):146-155. [doi: 10.1111/hir.12143] [Medline: 27207817]
103. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Comm Res 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
104. Sillence E, Briggs P. Ubiquitous computing: trust issues for a “healthy” society. Soc Sci Comput Rev 2008 Feb 01;26(1):6-12. [doi: 10.1177/0894439307307680]
105. Weiser M. The computer for the 21st century. SIGMOBILE Mob Comput Commun Rev 1999 Jul 01;3(3):3-11. [doi: 10.1145/329124.329126]
106. Singal H, Shruti K. Mitigating information trust: taking the edge off health websites. Int J Techn 2016;7(1):16-33. [doi: 10.4018/IJT.2016010102]
107. Soederberg Miller LM, 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]

Abbreviations

HCP: health care professional
WHI: Web-based health information

Edited by A Keepanasseril; submitted 24.02.17; peer-reviewed by E Sillence; comments to author 12.04.17; revised version received 21.04.17; accepted 24.04.17; published 19.06.17.

Please cite as:
Sbaffi L, Rowley J
Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research
J Med Internet Res 2017;19(6):e218
URL: http://www.jmir.org/2017/6/e218/
doi:10.2196/jmir.7579
PMID:28630033

Original Paper

Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review

Su Golder[1], BSc (Hons), MSc, PhD; Shahd Ahmed[1], MD; Gill Norman[2], BSc (Hons), PhD; Andrew Booth[3], BA (Hons), MA, PhD

[1]Department of Health Sciences, University of York, York, United Kingdom

[2]School of Nursing, Midwifery & Social Work, University of Manchester, Manchester, United Kingdom

[3]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Su Golder, BSc (Hons), MSc, PhD
Department of Health Sciences
University of York
Heslington
York, YO10 5DD
United Kingdom
Phone: 44 1904 321904
Email: su.golder@york.ac.uk

Abstract

Background: Although primarily used for social networking and often used for social support and dissemination, data on social media platforms are increasingly being used to facilitate research. However, the ethical challenges in conducting social media research remain of great concern. Although much debated in the literature, it is the views of the public that are most pertinent to inform future practice.

Objective: The aim of our study was to ascertain attitudes on the ethical considerations of using social media as a data source for research as expressed by social media users and researchers.

Methods: A systematic review was conducted, wherein 16 databases and 2 Internet search engines were searched in addition to handsearching, reference checking, citation searching, and contacting authors and experts. Studies that conducted any qualitative methods to collect data on attitudes on the ethical implications of research using social media were included. Quality assessment was conducted using the quality of reporting tool (QuaRT) and findings analyzed using inductive thematic synthesis.

Results: In total, 17 studies met the inclusion criteria. Attitudes varied from overly positive with people expressing the views about the essential nature of such research for the public good, to very concerned with views that social media research should not happen. Underlying reasons for this variation related to issues such as the purpose and quality of the research, the researcher affiliation, and the potential harms. The methods used to conduct the research were also important. Many respondents were positive about social media research while adding caveats such as the need for informed consent or use restricted to public platforms only.

Conclusions: Many conflicting issues contribute to the complexity of good ethical practice in social media research. However, this should not deter researchers from conducting social media research. Each Internet research project requires an individual assessment of its own ethical issues. Guidelines on ethical conduct should be based on current evidence and standardized to avoid discrepancies between, and duplication across, different institutions, taking into consideration different jurisdictions.

(J Med Internet Res 2017;19(6):e195) doi:10.2196/jmir.7082

KEYWORDS

review literature as topic; social media; ethics; research design; qualitative research

Introduction

Background

Social media are any Web-based computer-mediated tools to cocreate, share or exchange information, ideas, pictures or videos in virtual communities and networks (such as message boards, social networks, patient forums, Twitter, blogs, and Facebook) [1]. The availability of social media opens up new avenues for researchers to easily collect data, especially from sources that may have previously been difficult to access. This has led to a massive surge in social media analytics (whereby posts or chats are analyzed via qualitative methods or aggregate numerical data collection). The order of magnitude of data and the speed with which it is made available (approaching real time) make social media a potential tool to revolutionize health research [2]. Health-related social media analytics has taken many forms, including drug or product surveillance, monitoring disease or health patterns, or views or experiences of patients. Examples are pharmacovigilance (such as the discovery of adverse events), monitoring outbreaks of the flu epidemic, illicit drug usage or suicide patterns, and views on vaccinations and health service quality [3-29].

However, these new research avenues are not without ethical challenges [30-34]. In common with other research, potentially difficult considerations surround the purpose and value of the research, benefits and harm to participants, as well as privacy, informed consent, and confidentiality. However, Internet research is very different from traditional research and as such brings about many different ethical challenges. Whereas procedures are well established for obtaining ethical approval for traditional research, how far these can be transferred directly to Internet-mediated research is difficult to decipher. Whereas the ethical issues of social media research have been much debated [34-37], the attitudes of social media users (either posters or lurkers) and researchers have rarely been sought [38-42]. Researchers currently seek guidance from a wide variety of sources, such as individual institutions, research supervisors, subject specialist guidance [43], and increasing guidelines proposed specifically for research using social media [39,44-49].

This systematic review summarizes the existing research that has evaluated attitudes on the ethical considerations of research using social media. This will help to contribute to, and consolidate, current research practice as well as to clarify those ethical issues most pertinent to the public and researchers. This, in turn, will help guideline developers to formulate evidence-based guidelines for researchers conducting research using social media.

Objective

This study aimed to systematically review the research evidence that has evaluated attitudes of social media users, researchers, and other stakeholders on the ethical considerations of using social media as a data source for research.

Methods

Inclusion Criteria

Due to the anticipated dearth of studies specific to health-related research and the potential for generalizability of other research, any research area was considered. To reflect the qualitative or mixed methods nature of the research, we adopted SPIDER (sample, phenomenon of interest, design, evaluation, research type) for defining the inclusion criteria [50]:

S- Sample: Any sample of people (such as social media posters or lurkers, researchers, academics or other stakeholders). No minimum sample size was implemented.

P- Phenomenon of interest: Attitudes held on the ethical implications of conducting research using social media analytics.

D- Design: Any qualitative data collection methods (eg, surveys, questionnaires, interviews, observations, or focus groups) independent of the analysis conducted. Discussion papers were excluded.

E- Evaluation: Any information on attitudes to the ethical implications of research using social media. Such information may be the primary or secondary focus of the study.

R- Research type: Qualitative (such as interviews or focus groups), quantitative (such as surveys or questionnaires with fixed responses only) or mixed methods (such as research which collates a combination of fixed and open-ended responses).

No date, language or publication type restrictions were applied to the inclusion criteria. However, financial and logistical restraints did not enable translation from all languages.

Exclusion Criteria

Papers that are non–research based such as discussion papers were excluded. Such papers have been summarized elsewhere [35].

Research on the ethics of individual “look-ups” were excluded, for example, employers seeking information on employees or prospective employees, parents viewing their children’s posts, and health professionals seeking information on patients (or vice versa). Research on individual privacy or security issues such as fraud, cyberbullying, grooming, and child protection were also excluded.

Search Methods

A total of 16 databases and 2 Internet search engines were searched in addition to handsearching journals and conferences (Multimedia Appendix 1). Databases were carefully selected to reflect the multidisciplinary nature of the review. Other methods included reference checking all included articles and any related systematic reviews, citation searching of key papers on Google Scholar and Web of Science, and contacting authors and experts.

Search Strategies

The search strategies contained 3 facets—social media, ethics, and qualitative research. A date restriction of 1996 onwards was placed on the searches as blogging first began in 1997. No language limits were placed on the searches. The Embase search strategy is contained in Multimedia Appendix 2 and was translated as appropriate for each database.

Selection of Studies

The results of the searches were entered into an EndNote library and duplicates were removed. The titles and abstracts were screened by 2 researchers independently (SG and GN). The full text of any potentially relevant articles was assessed for eligibility by 2 researchers independently (SG and GN or SA). Disagreements were resolved by consensus based discussion and, if necessary, a third reviewer.

Data Extraction

Information was collected on the research question, the numbers of respondents, the characteristics of the participants (such as age and gender), the methods used to ascertain attitudes (such as interviews and survey), sampling methods or survey distribution (such as email and snowballing), questions or methods used to ascertain ethical considerations, and key findings. Data were extracted independently by 2 reviewers (SG and SA). Any disagreements were resolved by discussion or a third reviewer where necessary.

Assessment of Methodological Quality

Two reviewers (SG, SA) conducted independent quality assessment using the methodological assessment tool—quality of reporting tool (QuaRT). This involves checking the reporting quality of the articles using 4 elements: (1) the question and study design, (2) recruitment and selection, (3) methods of data collection, and (4) analysis. Studies were categorized as “adequately reported” when a “yes” had been assigned against 2 or more criteria or “inadequately reported” where a study was assigned a single “yes” response, or no yes responses [51]. Any disagreements were resolved by discussion or a third reviewer where necessary.

Analysis

Although the quality of the reporting of the included papers was assessed, no quality threshold was implemented. All studies which provided insight or contributed to the analysis were included.

We chose an inductive theme analysis with descriptive analysis rather than a framework approach. We acknowledge that a framework approach is an equally valid approach. However, we did not identify any framework completely compatible with our intended purpose. For instance, the most relevant framework identified was restricted to Twitter [52]. In addition, framework analysis has a risk of suppressing “interpretative creativity” and thus reducing some of the “vividness of insight” [53].

Inductive thematic synthesis aims to identify salient themes via coding of the data without the use of a preexisting coding frame, or any preconceptions held by the analysts [54] . We undertook coding in QSR NVivo Pro 11 by assigning text on a line-by-line basis to nodes developed by one author (SG) and then checked by a second author (SA). An aggregation approach to the synthesis of the data was applied with data from each study extracted and grouped together to form themes with supporting quotations. Finally, interrelationships between themes were assessed and organized into a structure to produce synthesized findings.

Reporting

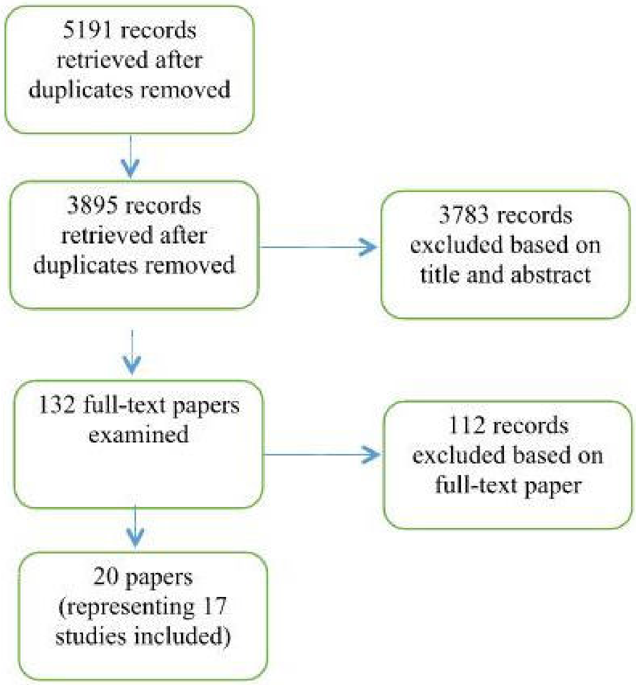
The enhancing transparency in reporting the synthesis of qualitative research (ENTREQ) statement of 21 items was used to report the stages of this review [55].

Results

Included Studies

Of the 3340 records (4636 before duplicates removed) identified by the original searches in February 2016 and a further 555 unique records by update searches in July or August 2016, 132 full-text papers were retrieved of which 112 were excluded. Excluded studies were mostly concerned with personal privacy or security (such as bank details), or with individual “look-ups” (such as seeking information on a particular person; Multimedia Appendix 3). Overall, 17 studies (from 20 publications) met the inclusion criteria (Figure 1).

Figure 1. Flow diagram.



Characteristics of Included Studies

Of the 17 studies, 12 explored the ethical concerns of social media users [38,40,56-67] and 5 asked researchers (mainly academics) [42,43,68-71] (Table 1 and Multimedia Appendix 4). The earliest study was published in 2001 and the latest in 2016, with 12 studies published after 2012.

Numbers of study participants were reported in 16 of the 17 studies. The smallest study involved 26 respondents and the largest study 2260. The total number of participants was over 5453.

In 7 studies the participants were from a single country: the United Kingdom (4), the United States (2), or Australia (1). The other studies either explicitly indicated an international participant coverage (3 studies) or implied an international coverage of participants (7 studies)—most of which used social media for recruitment or study observation.

8 studies used surveys of participants' attitudes, whereas others carried out more in-depth research and analysis using interviews (6 studies), or focus groups (4 studies). A mixture of open and closed questions was used. Three studies used observational or experimental techniques; with an original design devised by the authors (4 studies used more than one technique).

The majority of the studies did not indicate the demographic details of the respondents; however, in those that did males and females were well represented. Some studies specifically targeted young people [38,59,65].

5 studies were specifically concerned with health-related research [38,57,64,66,69], whereas the remaining studies tended to be more general.

Quality Assessment

Only 3 studies (from 4 reports) were assessed as "inadequately reported" [42,56,67,71] (Multimedia Appendix 5). The contribution to the synthesis of these 3 "inadequately reported" studies was assessed as being limited. These 3 studies did not impact on the presence of concepts within the synthesis and had only a marginal effect on the detail within the concepts.

Analysis

Whereas some included studies reported summary data from surveys or questionnaires, the majority of studies presented author interpretations supported by verbatim extracts from participants.

Emerging Themes

Some responses spanned multiple themes and attempted to categorize overall reactions to social media research use. Many closely related themes underpin the diverse attitudes exhibited by the study respondents. The framework we adopted from our emerging themes was: the 2 actors (researchers and social media users), managing their relationship (consent), and framing the context (responsibilities of the social media site (Tables 2 and Multimedia Appendix 6). Some studies assessed multiple themes in a more integrated and quantifiable way and we consider these first under general reactions.

Within the arena of "researchers," we identified themes of the perceived benefit of the research, the affiliation or type of researcher, the validity of the methods used to conduct the research, and the risks to the researchers themselves. Those themes related to the "social media users" were concerned with the risks involved with particular concern for vulnerable groups. Linked to these risks were the intended purpose of the social media poster and their ability to self-regulate through personal censorship or privacy settings. The next theme related to "consent" and the importance of and difficulties of informed consent and research disclosure. The last theme was regarding "responsibilities"—either via the social media site (including the platform and site administrators) or legal requirements (Tables 2 and Multimedia Appendix 6).

General Reactions

Three studies attempted to quantify general reactions to social media research. Moreno found that over half (56.1%, 74/132) of university students "strongly like" or "somewhat like" the concept of using Facebook for research by university researchers, nearly a third (28.8%, 38/132) were "neutral" and only 15.2% (20/132) were "unsure or uneasy" or had an "overt concern" [38].

Williams found that 37.2% (N=564) of social media users were "not at all concerned" with their social media information being used by university researchers, whereas 46.4% (N=564) were "slightly concerned," 11.2% (N=564) "quite concerned," and 5.2% (N=564) "very concerned" [67]. Evans found that 60% (N=1250) of social media users felt that "sharing individuals' social media data with third parties, such as the government or companies, for research purposes" should not happen and 32% (N=1250) felt that "sharing overall numbers of social media data with third parties, such as the government or companies, for research purposes (but not linked to individuals)" should not happen [59].

Respondents in Evans may have been less positive because they were given use of examples relating to government and companies rather than university researchers [38,59,67]. The more positive attitudes in Moreno may be linked to surveying younger people [38,59,67].

Researchers

Perceived Benefit of Research

Research was considered more acceptable if "it's going to a good cause" [65] and was "morally right" or for "general good" (such as social benefit or to help others) [38,40,56-59,62-65]. Some respondents were more specific—stating that social media research could give a voice to patients and other groups, uncover true prevailing issues, and improve patient care [57,58,65];

> *I have no reservations about your mining information from forums...it will provide much information about the human side of illness and how individuals singly and collectively approach and cope through sharing. Dare I say its importance cannot be understated.* [Diabetic forum user 57]

> *I kind of think it's cool when it's stuff that's like the flu, because then that's how they know to get the vaccines to a place.* [Twitter user 64]

Others described the general benefit of social media research as a precondition to acceptance [38,40,57,65] or considered if the benefits outweighed the risks [40,64]. Others felt positive provided a caveat or set of conditions had been met, such as informed consent (see following sections).

Whereas some stipulated the good that should come from social media research, others stated the research they would not like to see, such as research with a "bad intention" [38,65], for commercial gain [40,57,62,63] or to drive an agenda [40].

The strong feeling for the "public good" meant that some felt service to the greater social good was more important than individual privacy concerns [64]. However, others considered that social media users' desire for privacy should take precedence over researchers' goals [66].

Type of Researcher

Linked to the purpose of the research was the affiliation of the researcher. The type of organization or company influenced whether or not respondents viewed research as "good quality" and user concerns about use of social media information [40,42,58,59,64,65,67,69]. Generally, respondents were less concerned about use of social media information by university researchers, than by students [58], the police, government organizations [67], commercial organizations [40,59,64,67] or journalists [58]. But no difference was reported between health organizations and researchers [69].

Social media users who preferred not-for-profit researchers (such as academics) to commercial organizations did so because the former were felt to be more "productive," more "ethical," and "not exploitative." Furthermore, users did not like their social media posts being used to generate a profit for others [40,58,64].

Table 1. Brief summary of characteristics of included studies.

Source		Number of respondents	Characteristics of the participants	Survey distribution	Specific questions or methods used to ascertain ethical considerations	Key findings identified in terms of the ethical considerations raised
Researchers as respondents	Alim 2014 [43]	64	International and interdisciplinary researchers and academics	Emailed questionnaire	*Open and closed questions*	Researchers aware of ethical issues but require clarity in informed consent and public and private data.
	Bakardjieva 2001 [56]	NR[a]	Mailing list discussants	Mailing list discussion	Discussions provoked by post	An ethical approach to Web-based research is practically achievable.
	Carter 2015 [68]	30	Academic staff from UK university	Emailed Web-based survey	Respondents asked to agree, disagree or neither to 12 statements.	Recognize importance of avoiding deception and gaining consent, but acknowledge problems. Most disagreed that studying public social media data is same as studying documented text and disagreed that individuals wouldn’t be identified if anonymous.
	Denecke 2014 [69]	45	International Medical Informatics Association (IMIA) social media working group members	Mailing list members	3 questions asked	Different social media platforms should be managed in different ways in terms of confidentiality and privacy. Individuals should be deidentified and cited only indirectly.
	McKee 2009 [70]	30	International and interdisciplinary researchers from corporate research centers and academia.	Contacted known researchers	Used open-ended interviews	Researchers strived to follow “do no harm” principle. Agreement that there is no blanket approach to Internet research ethics.
	Woodfield 2013 [71]/Salmons 2013 [42]	465	International and interdisciplinary researchers	Program of Web-based and offline activities including conference and Twitter chat	Open discussions in which ethics consistently raised	Discussions focused on informed consent, confidentiality or anonymity, role and safety of the researcher, and research setting or social media platform. Concern of a lack of agreement.
Social media users as respondents	Beninger 2014 [40]	34	Adult male and females	British social attitudes 29 (BSA 29) survey and external recruitment agency	Focus groups and interviews using vignettes	Conditional acceptance of using social media for research.
	Bond 2013 [57]	26	Male and females	Diabetes forums	Web-based semistructured asynchronous (email) interviews	Agreed forum posts in the public domain and aggregated information could be used by researchers.
	Chen 2004 [58]	47	Mailing list owners and moderators or long standing members on sensitive and controversial topics	Emailed mailing lists and newsgroups	Survey questionnaire	Animosity toward researchers. Research should be conditional on research identify disclosure, informed consent, and feedback.

Source		Number of respondents	Characteristics of the participants	Survey distribution	Specific questions or methods used to ascertain ethical considerations	Key findings identified in terms of the ethical considerations raised
	Evans 2015 [59]	1265+	Aged 13 to 75 years and broadly reflective of UK population	NR	Web-based survey, workshops or interviews included example research projects	60% do not support use of social media data for research. Terms and conditions not sufficient for consent and need option to opt out. Biggest factor in likelihood to approve research is whether data is public.
	Hudson 2004 [60] and Hudson 2005 [61]	Up to 2260	Chatroom users on ICQ chat—range of topics including geographical region or language, age-orientated, romance or friends, adult or sexuality, technical, trivia avoided sensitive discussions such as “breast cancer survivors”	Monitored chatrooms	Recorded if “kicked-out” of chatroom and messages on why or comments to researchers	Kicked out of 63.3% of chatrooms when message posted about research. Reasons were— prohibition of spamming, opposition to being studied, general requests to leave, and insults.
	Michaelidou 2016a [62] and Michaelidou 2016b [63]	405	Adult male and females	NR	Focus groups and Web-based survey	10-item scale of transparency, legality, approval, privacy concerns, permission, vulnerability, reward, consumer responsibility, protection, and terms.
	Mikal 2016 [64]	26	Male and females with average age 26 years. Over half with depression history	Advertised on list serves, discussion boards, and local Internet community websites	Focus group s *emistructured* interviews	Relatively positive view provided data are anonymous and aggregated to protect identities.
	Monks 2015 [65]	48	13-14 years old Australian school students	Sample of convenience drawn from students participating in a leadership workshop	Focus groups with 3 main open questions	Some concerns about privacy but open to the use of social media for research if opportunity to provide consent and assured of confidentiality and anonymity.
	Moreno 2012 [38]	132	Male and female adolescents aged 18-19 years within US university	Used Facebook to identify students	At end of interview about health, asked “We identified participants for this study by looking at publicly available Facebook profiles. Do you have any thoughts about that?”	Endorsement by 26 respondents. 48 were “fine” with it, 38 were neutral or no specific comments, 12 were uneasy, 8 had overt concerns.
	Petersen 2013 [66]	27	Members of Medical Webmasters (MWM), an open, unmoderated list and Patient Advocates in Research (PAIR), a closed, unmoderated list for cancer patient advocates	Posted on 2 electronic lists	Survey	Two themes emerged. Respondents believed journalists should seek permission from list members and/or webmasters and viewed members’ desire for privacy as taking precedence over researchers’ goals.

Source		Number of respondents	Characteristics of the participants	Survey distribution	Specific questions or methods used to ascertain ethical considerations	Key findings identified in terms of the ethical considerations raised
	Williams 2015 [67]	564	Social media users	Web-based survey	Survey with open and closed questions	37% are not at all concerned about their social media information being used by university researchers, whereas 46% are slightly concerned, 11% are quite concerned, and 5% are very concerned.

[a]NR: not reported.

Table 2. Summary of coding framework.

Concept headings	Concepts derived for coding	Definitions
Researchers	Perceived benefit of research	Overall outcome or intention of the research to do "good."
	Type of researcher	The affiliation of the researcher (such as university or commercial company). This is associated with the perceived benefit of the research.
	Validity of the research methods	High or low validity of the methodology used, including risk of bias.
	Risks to researchers	Any risks that the researcher is exposed to.
Social media users	Risks to social media users	Any risks that the social media users are exposed to either individually or as a group.
	Vulnerable groups	Groups could be determined as vulnerable by either their individual characteristics or the topic discussed.
	Original purpose of posts	The intent of the poster at the time the message was placed.
	Privacy and self-regulation	The public versus private nature of social media and the need for anonymity or confidentiality. Connected to this issue is self-regulation whereby individuals control content.
Consent	Informed consent	Permission for posts to be used in a study.
	Researcher disclosure	Researchers being transparent and honest about their intent. This can be either up-front or at a later stage.
Social media site responsibilities	Terms of service	Also known as "terms of use" or "terms and conditions," these are the rules agreed to in order to use social media sites.
	Site administrators	Site administrators, list administrators or list moderators are often in charge of maintaining a discussion or mailing list.
	Legality	Refers to legal issues, regulation or government oversight and includes issues of copyright.
	Type of platform	The type of social media platform, for example closed or open, personal or professional. This is connected to issues of public versus private space, and the original purpose of the postings.

Validity of the Research Methods (High and Low)

Whereas users' perceptions of the validity of social media research was partially influenced by the researcher affiliation, they also discussed its methodological rigor on its own merits. Attitudes were divided as to whether social media research could be viewed as high or low validity research, particularly compared with more traditional research methods. Those users who viewed social media research as high quality cited it as a means of quick access to vast amounts of timely information and large samples to mitigate the effect of false information or extreme views and improve research accuracy [40]. Anonymity of posts was also thought to encourage open and honest opinions and discussions particularly about sensitive issues or nonconventional or "politically incorrect" views. Research using social media was also seen to avoid biases inherent in having to answer questions in the presence of others, such as in a survey [40].

In contrast, other social media users were concerned about the low validity of the research in terms of quality of the data, representativeness, and poor methodological approach [40,58,64,65]. Concerns were raised about inaccurate or false data or accounts with people severely limiting what they post on the Web (see self-regulation), and behaving differently off and on the Web [40,58 64,65];

> *I've never once posted anything negative. So if you took that data, it would not be accurate, because of course I have had bad days or sad days.* [Twitter user [64]]

Social media posters were not considered representative of the general population and using social media would lead to "only the loudest voices heard" [40,64].

Some social media users were skeptical not only about the data posted on social media but also about the accuracy of the methods of using social media, and biased research;

> *You can't even get targeted advertising right, what makes you think public health accuracy is going to be any better?* [Twitter user [64]]
>
> *Many researchers in the fields covered by this list do research solely to "prove" that our illness are faked or psychological. Most of us do not care to operate with people like that. Any truly unbiased research is fine.* [Chemical-injury mailing list owner [58]]

Risks to Researchers

Insults or being "flamed" were the most common threat posed to researchers [58,60,61]. Researchers spoke from experience of the need for care that they do not become victims to trolls and to separate their "researcher" persona from their "personal" persona and thus protect the boundaries between their professional and personal lives [60,61,71].

The potential for more extreme harm was cited in countries with governments which control Internet access and communications [70]. For example, an associate professor carrying out a study in Central Asia stated that local researchers could be risking their lives by conducting social media research;

> *I can't get anyone to work with me right now because the Uzbek government just passed a law that anyone accused of giving sensitive information to foreigners will be accused of treason, and the law doesn't define what is sensitive information. Now the penalty for treason in Uzbekistan is death, I believe.* [Associate professor, University of Washington [70]]

Social Media Users

Risks to Social Media Users

Social media users were also worried about the risks of judgment or ridicule or unsolicited attention on the Web and, more seriously, "abuse" or bullying [40,43,64,65,70]. Other possible harm included exploitation from organizations or use by the police or courts for purposes of prosecution in divorce cases, child custody cases or lawsuits [40,58]. Other social media users felt very uneasy about social media research or felt it was "creepy" or "scary" [38,58,60,61,64,65]. Respondents commonly held the perception that posters are being "exploited" or "used," with researchers using social media posters to "get someone to do their work for them" was widely held [58,65].

Risks were associated with data being taken out of context, used inappropriately or the poster being identifiable. Users were concerned at the potential to distort the context in which something was said or that findings would be used to defend or promote something that was not intended (see purpose and validity of the research) [40,64,65]. Some respondents, although happy for researchers to use verbatim quotes, felt that "if it actually involves taking your comments and interpreting it, then it's a very different thing" [40]. Use of verbatim quotes, rather than some form of interpretation [65], was one solution to taking comments out of context. However, this bought about issues of anonymity and privacy.

Researchers were generally aware of the risks to social media users (even with anonymized data) and considered these risks in their studies [43,70]. In addition, researchers were worried that the risks were not taken seriously by international review boards (IRBs) [43,70]. Deidentification of social media posts was seen as crucial to minimize negative consequences [65]. However, using verbatim quotes could compromise individual anonymity (see privacy).

Whereas users were well aware of the risks, they exhibited a feeling of apathy [38,40,56,64] with risks just being something to be accepted and the only way to stop it happening being to stop using social media [38,40,56,64];

> *With some of the stuff I write, I am uncomfortable thinking it is going to be accessible for a long time but this is after all the Internet and it's hardly private...The alternative, (that is) total privacy is to sit here in my house alone and not communicate. I'd give it about three weeks before total insanity set in.* [Discussion list member [56]]

Risks extended beyond individuals to social media groups. It was considered important to maintain social media as a protected space where members may speak openly without concern that their words will be shared outside the group [65,66]. Researchers were seen by some as an intrusion which can destroy the dynamics and enjoyment of using social media and curtail freedom of expression [58,65,66,70]. This was reiterated by some users who reluctantly self-regulated their posts [65]. Some users even felt that the damage to communication and community within these forums could lead to people not participating or sites closing [58,70].

Vulnerable Groups

Certain groups such as children and teenagers [42,43,62,63,65,71], individuals suffering from mental health issues [64], and even the deceased [66] were perceived as vulnerable and thus required extra emphasis on respect and caution to counteract this vulnerability [65,66]. The legal context and government practice of the country from which the post originates may also affect the potential for harm. For example, homosexuality is illegal, or at least taboo, in certain regions. Thus, individuals could be exposed to severe harm if their sexual orientation were publicized [70]. Risks to professional reputations and careers were also raised for those, such as school teachers or health professionals, with responsibility for potentially vulnerable or impressionable individuals [40].

Whereas some social media users thought ethical principles of research should be upheld regardless of the topic of the research [40,65], others thought that topics of a sensitive or personal nature needed more consideration [58-61,65,70]. The sensitive nature of some discussion groups was cited as good reason to prohibit or discourage researchers;

> *Our code of conduct explicitly prohibits information gathering from SPALS (subsequent pregnancy after a loss support) for other than immediate personal use.…Privacy and confidentiality are also concerns. We don't want to attract the "research-paparazzi."* [List owner [58]]

Original Purpose of Posts

When social media users post on social media, they may have no expectation that this would ever be used for research [40,56,57,65,71]. The intended audience may be limited to friends and family and possibly "friends-of-friends" [65,71]. Considerations of the wider implications and access by third parties are not likely to be at the forefront of many social media users' minds [40,56,57,65].

Even if people are aware of the public nature of social media, people may still "get carried away with themselves when they are writing (on social media platforms)" [40] and then "once it is on there, to try and get rid of it, it's too late or it's too hard" [40]. Thus, there is a need to consider carefully the impact of reporting of verbatim social media data no matter how open or public a site is considered to be [71];

> *I was also very irritated with people who used that argument that we should not ask for informed consent because it is easy to get into the groups. It is the participants' purpose for being in the group that is important in a way and their feelings about what kind of space this is.* [Researcher, Norwegian University of Science and Technology [70]]

Privacy and Self-Regulation

Whether social media should be seen as a public or private space occasioned contention as well as confusion [40,42,43,56-66,68-71] and was the principal factor in the likelihood of approval for social media research projects [59]. Some social media users believe that "there is no such thing as privacy online" [40], so once information is posted it is available to the public and thus can be accessed and used for research purposes [38,40,57].

Despite the commonly held view that social media are public, some privacy was still expected and this raises caution for Internet research [38,56,57,62-64]. Some users still expressed discomfort about it being used—although appreciated how contradictory this may appear [57,63];

> *I write a blog about my experience of diabetes and would feel very aggrieved if I found any of it quoted in a medical research paper without having been asked. I realise this is slightly contrary (since I am posting and effectively actively encouraging readership) but nevertheless it would feel like "theft" of my content.* [Diabetic forum user [57]]

An expectation of privacy resulted, in part, from a lack of understanding of the extent of the public nature of social media exchanges [56]. Navigating the privacy settings of social media isn't always simple or straightforward. Some social media users did not know what is public or private on social media or what their settings were [38,64,65]. Some users did not understand the permanent nature of social media data, how extensive data reach can be, and the big data computational tools that can be used to analyze posts [64]. Users also felt it was easy to forget or not think about this while posting (see risks) [57].

Confidentiality and anonymity were thought to help protect privacy [42,43,70,71]. Social media users emphasized the importance of anonymity [40,59,64,65,67,69] with approximately three-quarters preferring to remain anonymous [59,67]. Users were also more likely to accept research if it looked at deauthored data or used overall aggregated numbers [40,57-59,64,65,69].

Some respondents (both users and researchers) identified the challenge of using quotations and maintaining anonymity [43,57,68,70] given that quotes, even if anonymized, can be traced back to their origins using a Google search [57,70]. These views were confirmed in a small Web-based survey of academics from a single UK university where only 10 out of 30 researchers agreed it was "very unlikely that individuals will be able to be identified if social media datasets are anonymised" [68].

One solution to using direct quotes was to cite only indirectly or to paraphrase quotes [69]. This, however, could have implications for those social media users' who were concerned about their posts being taken out of context (see risks).

Respondents who disagreed with the need for anonymity believed that users are responsible for managing their identity as people "can always be anonymous if you want to be" [40]. For example, users could use a username unrelated to their real name, utilize privacy settings, and select what to share on the Web [40,43,62-65,68]. The idea that there is no such thing as privacy was reiterated (52) and as such self-regulation was key [64];

> *I think our generation is gravitating towards (the idea that) privacy is not to be expected anymore. You have to create it yourself. You have to enable it yourself, because it just doesn't exist anymore.* [Twitter user [64]]

This idea that self-regulation should be relied upon was reflected in the results of the survey of UK academics where 17 out of 30 agreed "it is the responsibility of individuals to rethink how they use social media if they are unwilling for their online public behavior to be studied by researchers" [68].

These views were inextricable from views on informed consent as social media users who actively "self-regulated" did not think researchers needed to gain consent [40].

Consent

Informed Consent

Generally social media users were divided as to whether they agreed or disagreed that social media research required informed consent from posters [67]. Users and researchers who did not feel informed consent was necessary tended to feel that informed consent was implied by the public nature of social media

[40,43,57,64]. Others felt that anonymization removed the need for informed consent [40,43,57,64,65,69].

Other social media users expressed reservations feeling that researchers should obtain permission for use [57,58,62,63,65,66]. They linked this to the original intention of the post [57], data ownership, or to difficulties of anonymizing direct quotes [57,65]. Attitudes appear to be changing as users learn from experience and social media develops;

> *I have allowed this in the past, but I feel that they should get permission first.* [Mailing list owner [58]]

Gaining consent was seen as part of common decency and not solely to ensure good ethical practice [40,57,65]. Researchers were felt to have a moral responsibility toward Web-based content [40,57,65] and to protect citizens from violations linked to social media research [62,63].

Some researchers tended to assume that proceeding without informed consent was acceptable because social media are public [38,40,43,57,68].

> *Only publicly visible data was extracted so we thought that, because the data was publicly available, no ethics applied.* [Researcher [43]]

This was reiterated in the survey of 30 UK academics where 10 felt that there is no need for informed consent if social media were publically accessible [68].

However, they also raised the issue of whether posts represent a “human subject” or text [58,70].

> *If I think of Perry’s comments as the letter for the editors, I don’t have to get any informed consent, but if I think of it as a personal conversation, I have to get informed consent.* [Doctoral student, York University in Canada [70]]

This issue was not discussed among social media users but when asked, 22 out of 30 UK academics disagreed that “studying public data on social media is essentially same as studying documented text” [68].

Some users saw the necessity for informed consent as depending upon both the content and type of analysis. Many thought that sensitive, personal posts or posts with a “sexual, political or religious” focus required informed consent [40,58] and were more accepting of the use of aggregate data, generalizations or observational overviews than case studies or the use of quotes (this was interconnected with anonymity) [56-58,64,69,70].

Both social media users [40,65] and researchers [43,68,70] felt that it was difficult to implement informed consent. Challenges related to the large amounts of data involved, the impossibility of getting informed consent from all the users, and difficulties in how and whom to ask [43,70]. The impracticalities of detecting minors were also highlighted [43,65]. However, where respondents favored consent, they did not think that logistical burdens offered a justification for not seeking permission [40,65].

Research Disclosure

Many social media users felt that the collection, access, and use of social networking data should be transparent [40,56-58,60-66]. Some felt that authors of postings should know how their comments might be used “up-front” at the time of producing them [40,42,56] with an option to opt-out of research (or even better, an opt-in) [58-61];

> *Any researcher that joins a mailing list should identify themselves as such as soon as they have joined-opt better yet before they have joined and ask permission of the list owner. As a person I have a right to know I am being experimented on or studied.* [List owner [58]]

Many researchers agreed that being explicit and transparent as possible about one’s role as a researcher was the best possible action [68,70] whereas some were uncertain about when it is appropriate to collect data without disclosing their identities [71]. Deception of social media users was generally seen as unacceptable [68].

Respondents drew a distinction between naturally occurring social media data and data stimulated by a researcher’s intervention. It was more important for a researcher to make themselves and their intentions known in advance when participating in forums [71].

Social Media Site Responsibilities

Terms of Service

Most researchers factor in consideration of “terms of service” for the social media platforms from which they extract data into their research planning [43]. Whereas the vast majority of social media users were aware of “terms of service,” neither researchers nor most social media users agreed that this is sufficient for informed consent [59,64,67,68]. Terms of service were considered too long, dense, and confusing [40,64]; “constantly changing” [40]; and unread by most members of social media sites [40,64,65,67,69,71]. A few social media users felt that the public openness and accessibility of the platform of social media (such as Twitter or LinkedIn) implies that third parties may use the data [40]. Thus, researchers should not rely upon the terms of service.

Site Administrators

There was an absence of consensus over the role of site administrators. Some social media users thought that researchers need to gain permission from the list owner [58,66,69], in addition to the users’ permission;

> *No individual or entity should be using it (private forum) for research without explicit permission from both the people who writes the message as well as the people/group who runs the mailing list.* [List owner [58]]

Others were vehemently opposed to list owners giving permission on behalf of members;

> *I think it would be a complete betrayal if (admin) were to give permission on behalf of the members.* [Diabetic forum user [57]]

Legality

Some social media users thought government oversight or regulation should ensure the ethical use of data and protect the rights of users [64]. Some suggested a law against collecting information about social media users [62,63] and not gaining consent was compared with "hacking" [40]. Others suggested that governmental oversight was unnecessary leading to fears of Orwellian monitoring [64].

The public nature of social media platforms and their content raises the issue of data ownership [40,57,62,63,66,70]. Some social media users thought that users automatically surrender their right to ownership by posting and that as they are public they are "uncopyrighted so they are 'free' for anyone to use" [56] or the social media platform owns all data on the site [40,57,64].

However, other social media users and researchers thought that users own the intellectual property of content they post and that posts should be treated in line with copyright laws [40,43,56,58,59,66,69,70]. This requires that proper referencing or acknowledgment is in place [40,57,58,70];

> *If someone decided to republish my post in another forum or document, I would expect my comments to be kept in context and credited to me.* [Diabetic forum user [57]]

However, including a "handle" or the Web-based username in a reference was perceived as problematic by some social media users who valued anonymity over credit [40,57,58,70].

Type of Platform

The type of social media website, such as open and closed groups or sites with privacy settings, also influenced whether consent was considered implicit [40,64,65,69,71];

> *I feel if a forum is viewable to the public, IE you don't have to be a member to view any of the forum threads, then you or anyone else can use any of the information you find on any forum.* [Diabetic forum user [57]]

Twitter and Web-based open forums were seen as inherently public forums [40,64,65]. In contrast, Facebook has explicit privacy control settings and was therefore viewed differently [40,64];

> *It all comes down to the fact that we know that we're using Twitter and it's public. I think I might honestly feel differently about that if it were Facebook, because I do feel like there is some degree of privacy in Facebook.* [Twitter user [64]]

iMessage and other messaging functions were viewed as private spaces with closed conversations [58,65];

> *A mailing list unlike a sidewalk has a membership list and only members are part of that list… the mailing list retains an identity as a PRIVATE forum. With that in mind, no individual or entity should be using it for research without explicit permission..* [List owner [58]]

Social media websites with a fun, social purpose are likely to contain much more "personal" content and were therefore viewed differently from websites with a professional aim such as LinkedIn [40].

Users were also more concerned about researchers accessing and using photos than written content because text could have been written by anyone—whereas it is more difficult to falsify photos [40].

Discussion

Principal Findings

This systematic review demonstrates the need to understand a complex array of interrelated and challenging factors with respect to ethical considerations in social media research. Diverse important issues and concerns remain to be addressed, but consensus proves difficult to achieve. Respondent views varied considerably from complete "animosity" to being "overly positive." In between were attitudes of conditional acceptance (for instance social media research only to be undertaken if informed consent is gained or the site is public) and complete apathy. In addition, even views from the same respondent were conflicting with many being well aware of their contradictory views. The comprehensiveness of this review enables us to develop a broader view of the contrasting issues rather than being restricted to findings from a single study or single population that is not as transferable.

The differing attitudes can partially be explained by the heterogeneity of respondents, their different understandings of social media research, and different methodologies used in each included study. In particular, Twitter users and the younger generation tended to be more accepting of social media research. Research using social media is a relatively new phenomenon. Some respondents had little understanding of social media research particularly in the earlier studies or were unaware of the range of examples. Social media users were more likely than researchers to evaluate the perceived benefit of the research and the validity of the research when considering the acceptance of social media research. Researchers, on the other hand, were also more likely than social media users to be concerned about risks to researchers themselves.

There was general agreement on several issues. Whereas there was strong support for social media research that is for the "good" of society (and thus greater acceptance of university researchers)—this should not offer researchers carte blanche or an opportunity to ignore ethical principles. There are different definitions of "beneficial" research and some people could still be at risk of harm. Respondents acknowledged possible risks with these being more complex for such groups as, children, adolescents or vulnerable adults [38,72-75] or for discussions on sensitive topics. Risks to social media group dynamics and freedom of expression have been highlighted in the literature [36,76] and this review demonstrates the strength of feeling about the importance of keeping safe and supportive environments for people to post on. The potential benefits, therefore, should still be weighed up against any potential harm. However, often researchers considered general principles such as "do no harm" as difficult to follow [70].

Respondents were much more likely to support the use of numerical aggregate data (such as overall statistics) than qualitative research involving quotes or interpretation of quotes. This view was almost unanimous [40]. Respondents agreed overwhelmingly that the terms of service of social media platforms are infrequently read and should not be relied upon. This finding may be related to newsworthy cases where terms of service were relied upon. Over time much has been learnt through trial and error. As a consequence, platforms such as Twitter and Facebook have changed their terms of service with much tighter control over external research [76,77]. However, this review does not suggest that users feel this is the best way forward.

Respondents also thought that the conduct of social media research should not just obey the principles of the law (although there was little agreement as to what the law is) but also follow ethical principles and a moral obligation.

Many issues encountered little agreement and conflicting attitudes. The boundaries between public availability and privacy were particularly complex. The public nature of social media has been used to support current practice, whereby a minority of studies apply for institutional ethics board or other approval and most studies do not mention ethical approval [43,73,78-80]. Social media research, therefore, often involves a lack of informed or valid consent with group members often unaware that they are being monitored [76,80,81]. Whereas some social media users seemed happy with this lack of consent, quoting the public nature of social media and the potential to self-regulate, others strongly opposed this on the grounds of privacy, original intent of the post, risks to users, and ownership of the data. These arguments are similarly rehearsed in the literature [34,43,78,82-86].

Rooted within the concepts of risk to users are the issues of privacy and traceability or anonymity of the poster [34,38,39,41,48,71-73,77,81,87-93]. Whereas some respondents were happy for use of anonymous posts, others wanted to be cited (as is the case for published works). Both users and researchers displayed a lack of understanding of the difficulty in upholding anonymity [80]. In 2006, deanonymization of social media users was carried out by journalists of the New York Times [94], and in 2011 data aggregated and allegedly anonymized by researchers from Harvard were deanonymized [83,95]. Covert participation inside Web-based communities has also occasioned controversy [96].

Respondents lacked clarity with regard to informed consent, when it should be implemented, and how [42,71]. However, users agreed that it is unacceptable to use names and direct quotes without consent and that it is practically difficult to obtain informed consent. The issue of informed consent was highlighted in 2012 when Facebook manipulated users' posts in an experiment to influence people's mood. This caused much anger and dismay among users who were unaware they were part of any experiment [97,98]. The included studies in this review, however, did not examine the ethics of manipulation of sites and this needs separate consideration.

Researchers may access content posted on the Web without interacting with the individuals who wrote the posts or even considering them "human subjects." The "human subject" or "published author" debate, however, was little discussed in the included studies, yet, has important implications for ethical approval and is often debated in the literature [70].

The absence of an overarching consensus with regard to social media research ethics is apparent. Many complex ethical dilemmas persist [73]. Whereas researchers have found it useful to read and understand ethical considerations faced by other researchers, they find it challenging to translate approaches used in another research context to their own particular research [70]. Each research project, therefore, requires individual consideration of its ethical issues. Just as a blanket approach to the ethics of traditional research (such as surveys or focus groups) would prove unsuitable so, too, we should resist a "one rule fits all" to social media research. Researchers should consider the type of research (such as aggregate, qualitative); the nature of their topic (whether sensitive or trivial); as well as issues of anonymization, confidentiality, informed consent, privacy, and the benefits and risks involved. The ability to undertake Internet research may depend on the level of trust and confidence the public have in the research being undertaken when people choose to contribute to studies or make posts publically available [77]. Koene [77] considers a possible revolt with a call for a public backlash and boycott of Internet-mediated research. This review does not suggest that social media users feel as strongly as Koene [77] suggests, but we should still be mindful of these issues.

Limitations

Although the heterogeneity of included studies contributes strength to this review, this also makes it challenging to arrive at definitive conclusions. In particular, studies differed in their methodological approaches and by their presentation of results.

Social media are constantly evolving and have changed considerably since the first of the included studies was undertaken in 2001. It would be interesting to uncover how people's perception of the ethical issues in social media research have changed; however, unfortunately the sparsity and heterogeneity of studies made it challenging to reveal any time trends.

Whereas additional non-database searches (such as citation searching and reference checking) were used in an attempt to overcome the difficulties in searching for this type of research, it is likely that some relevant studies remain unidentified. The review was also limited to papers in English or for which a translation could easily be obtained.

Conclusions

It remains unlikely that a consensus on the ethical considerations on using social media research will ever be reached. Each Internet research project requires an individual assessment of its ethical issues and selection of the most appropriate methodological approach.

Whereas the issues raised in this review suggest that ethical considerations in using social media for research are complex and require thoughtful consideration, this should not deter researchers from conducting social media research.

Contributions from social media offer a more immediate time window than experiences documented in formally published qualitative research. This is because social media data can be analyzed immediately whereas published qualitative findings may take at least two years from data collection, through peer review and the wider editorial process, to publication. To a large degree such contributions may be unfiltered and unfettered and less subject to the influence of the researcher. They are less constrained by the temporal and spatial limitations encountered when planning and conducting qualitative research. However, social media contributions are public offerings largely written in the knowledge that they could be read by a wider audience. They offer a perspective that is often stripped, or at the very least, lacking a grounding in context and that may challenge representation and interpretation. Thus, we cannot afford to miss the considerable potential of social media research and its unique contribution to knowledge.

Guidelines for ethical conduct should be based on the available best practices and standardized to avoid discrepancies and duplication from one institution to another. This methodological review is offered to initiate ongoing discussion within the research community of how such guidelines might be formulated. It highlights the importance of properly conducted social media research of benefit to the public. It also highlights the need to consider informed consent and privacy and researchers should not rely solely on regulation but have a moral and ethical duty to consider social media users and the main purpose of social media groups.

Acknowledgments

This report is an independent research arising from a postdoctoral research fellowship, Su Golder (PDF-2014-07-041), supported by the National Institute for Health Research (NIHR). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

The authors would like to thank Professor Yoon Loke, University of East Anglia, and Dr Darren Read, University of York for valuable discussion or comments on the paper. We would also like to thank Amy Hicks, MD, University of York for checking the interpretation of the results and reviewing the report.

Authors' Contributions

SG developed the study protocol, carried out the scientific literature search, screened the EndNote libraries, extracted and analyzed data, carried out the quality assessment, interpreted the results, and drafted the report. SH carried out the update searches, extracted the data, carried out the quality assessment, and reviewed the report. GN screened the search results, acted as third reviewer for disagreements on data extraction and quality assessment, and reviewed the report. AB gave substantive comments on all stages of the review from protocol to report.

Conflicts of Interest

SG had support from the NIHR for the submitted work; SG, SH, and GN had no financial relationships with any organizations that might have an interest in the submitted work, and they had no other relationships or activities that could appear to have influenced the submitted work.

Multimedia Appendix 1

Sources searched.

[PDF File (Adobe PDF File), 23KB - jmir_v19i6e195_app1.pdf]

Multimedia Appendix 2

Embase search strategy.

[PDF File (Adobe PDF File), 28KB - jmir_v19i6e195_app2.pdf]

Multimedia Appendix 3

Excluded studies.

[PDF File (Adobe PDF File), 86KB - jmir_v19i6e195_app3.pdf]

Multimedia Appendix 4

Characteristics of included studies.

[PDF File (Adobe PDF File), 67KB - jmir_v19i6e195_app4.pdf]

Multimedia Appendix 5

Quality of reporting in included studies.

[PDF File (Adobe PDF File), 32KB - jmir_v19i6e195_app5.pdf]

Multimedia Appendix 6

Results matrix for studies by emerging themes.

[PDF File (Adobe PDF File), 43KB - jmir_v19i6e195_app6.pdf]

References

1. Obar JA, Wildman SS. SSRN Journal. Social media definition and the governance challenge: an introduction to the special issue URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2647377 [accessed 2017-04-25] [WebCite Cache ID 6pz0vskJB]
2. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):e13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
3. Toussi M, De MC, Nadarajah S, Breckenridge A. Tracking of adverse drug reactions in social media: current status of requirements, best practices, methods and tools. Pharmacoepidemiol Drug Saf 2014;23:206-207. [Medline: 25483472]
4. Sampathkumar H, Chen X, Luo B. Mining adverse drug reactions from online healthcare forums using hidden Markov model. BMC Med Inform Decis Mak 2014;14:91 [FREE Full text] [doi: 10.1186/1472-6947-14-91] [Medline: 25341686]
5. Pages A, Bondon-Guitton E, Montastruc J, Bagheri H. Adverse drug reactions related to oral chemotherapy drugs: patients' internet narratives vs Pharmacovigilance Database. Fundam Clin Pharmacol 2014;28:52.
6. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol 2015 Oct;80(4):878-888. [doi: 10.1111/bcp.12746] [Medline: 26271492]
7. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on twitter through automated learning of naïve language. PLoS One 2013;8(12):e82489 [FREE Full text] [doi: 10.1371/journal.pone.0082489] [Medline: 24324799]
8. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
9. Witteman HO, Zikmund-Fisher BJ. The defining characteristics of Web 2.0 and their potential influence in the online vaccination debate. Vaccine 2012 May 28;30(25):3734-3740. [doi: 10.1016/j.vaccine.2011.12.039] [Medline: 22178516]
10. Firestone SM, Christley RM, Ward MP, Dhand NK. Adding the spatial dimension to the social network analysis of an epidemic: investigation of the 2007 outbreak of equine influenza in Australia. Prev Vet Med 2012 Sep 15;106(2):123-135. [doi: 10.1016/j.prevetmed.2012.01.020] [Medline: 22365721]
11. 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]
12. Collier N, Son NT, Nguyen NM. OMG U got flu? analysis of shared health messages for bio-surveillance. J Biomed Semantics 2011;2 Suppl 5:S9 [FREE Full text] [doi: 10.1186/2041-1480-2-S5-S9] [Medline: 22166368]
13. Corley CD, Cook DJ, Mikler AR, Singh KP. Using web and social media for influenza surveillance. In: Arabnia HR , editor. Advances in computational biology. Berlin: Springer; 2010:559-564.
14. Martin V, Zhou X, Marshall E, Jia B, Fusheng G, FrancoDixon MA, et al. Risk-based surveillance for avian influenza control along poultry market chains in South China: the value of social network analysis. Prev Vet Med 2011 Dec 01;102(3):196-205. [doi: 10.1016/j.prevetmed.2011.07.007] [Medline: 21925753]
15. Daniulaityte R, Carlson R, Falck R, Cameron D, Perera S, Chen L, et al. “I just wanted to tell you that loperamide WILL WORK”: a web-based study of extra-medical use of loperamide. Drug Alcohol Depend 2013 Jun 1;130(1-3):241-244 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.11.003] [Medline: 23201175]
16. MacKridge A, Sampere K. Online discussion of pharmaceutical cognitive enhancement: a qualitative analysis. Int J Pharm Pract 2012;20:30-31.
17. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
18. Goh T, Huang Y. Monitoring youth depression risk in Web 2.0. VINE 2009 Aug 28;39(3):192-202. [doi: 10.1108/03055720911003969]
19. Cox A, Anton C. Varenicline: evaluation of the utility of spontaneous consumer reports of suspected adverse effects filed on internet sites. Drug Saf 2009;32(10):903-904.
20. Young AM, DiClemente RJ, Halgin DS, Sterk CE, Havens JR. Drug users' willingness to encourage social, sexual, and drug network members to receive an HIV vaccine: a social network analysis. AIDS Behav 2014 Sep;18(9):1753-1763 [FREE Full text] [doi: 10.1007/s10461-014-0797-9] [Medline: 24849621]

21. Penţa MA, Băban A. Dangerous agent or saviour? HPV vaccine representations on online discussion forums in Romania. Int J Behav Med 2014 Feb;21(1):20-28. [doi: 10.1007/s12529-013-9340-z] [Medline: 24057409]
22. Coloma P, Becker B, Van ME, Kors J. Use of social media in pharmacovigilance: testing the waters. Pharmacoepidemiol Drug Saf 2014;23:290.
23. Larson HJ, Smith DM, Paterson P, Cumming M, Eckersberger E, Freifeld CC, et al. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. Lancet Infect Dis 2013 Jul;13(7):606-613. [doi: 10.1016/S1473-3099(13)70108-7] [Medline: 23676442]
24. Denecke K, Krieck M, Otrusina L, Smrz P, Dolog P, Nejdl W, et al. How to exploit twitter for public health monitoring? Methods Inf Med 2013;52(4):326-339. [doi: 10.3414/ME12-02-0010] [Medline: 23877537]
25. Nicholson MS, Leask J. Lessons from an online debate about measles-mumps-rubella (MMR) immunization. Vaccine 2012 May 28;30(25):3806-3812. [doi: 10.1016/j.vaccine.2011.10.072] [Medline: 22063388]
26. Seeman N, Ing A, Rizo C. Assessing and responding in real time to online anti-vaccine sentiment during a flu pandemic. Healthc Q 2010;13:8-15. [Medline: 20959725]
27. Keelan J, Pavri V, Balakrishnan R, Wilson K. An analysis of the Human Papilloma Virus vaccine debate on MySpace blogs. Vaccine 2010 Feb 10;28(6):1535-1540. [doi: 10.1016/j.vaccine.2009.11.060] [Medline: 20003922]
28. Skea ZC, Entwistle VA, Watt I, Russell E. 'Avoiding harm to others' considerations in relation to parental measles, mumps and rubella (MMR) vaccination discussions - an analysis of an online chat forum. Soc Sci Med 2008 Nov;67(9):1382-1390. [doi: 10.1016/j.socscimed.2008.07.006] [Medline: 18703263]
29. Huesch MD, Currid-Halkett E, Doctor JN. Public hospital quality report awareness: evidence from National and Californian Internet searches and social media mentions, 2012. BMJ Open 2014 Mar 11;4(3):e004417 [FREE Full text] [doi: 10.1136/bmjopen-2013-004417] [Medline: 24618223]
30. Swirsky ES, Hoop JG, Labott S. Using social media in research: new ethics for a new meme? Am J Bioeth 2014;14(10):60-61 [FREE Full text] [doi: 10.1080/15265161.2014.948302] [Medline: 25229594]
31. Flicker S, Haans D, Skinner H. Ethical dilemmas in research on Internet communities. Qual Health Res 2004 Jan;14(1):124-134. [doi: 10.1177/1049732303259842] [Medline: 14725180]
32. Elgesem D. What is special about the ethical issues in online research? Ethics Inf Technol 2002;4:195-203. [Medline: 15977360]
33. Heilferty CM. Ethical considerations in the study of online illness narratives: a qualitative review. J Adv Nurs 2011 May;67(5):945-953. [doi: 10.1111/j.1365-2648.2010.05563.x] [Medline: 21231955]
34. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [Medline: 11701577]
35. Herron M, Sinclair M, Kernohan W, Stockdale D. Ethical issues in undertaking internet research of user-generated content: a review of the literature. Evid Based Mid 2011;9(1):9-15.
36. Whitehead LC. Methodological and ethical issues in Internet-mediated research in the field of health: an integrated review of the literature. Soc Sci Med 2007 Aug;65(4):782-791. [doi: 10.1016/j.socscimed.2007.03.005] [Medline: 17512105]
37. Capurro R, Pingel C. Ethical issues of online communication research. Ethics Inf Technol 2002;4:189-194.
38. Moreno MA, Grant A, Kacvinsky L, Moreno P, Fleming M. Older adolescents' views regarding participation in facebook research. J Adolesc Health 2012 Nov;51(5):439-444 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.02.001] [Medline: 23084164]
39. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):708-713 [FREE Full text] [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
40. Beninger K, Fry A, Jago N, Lepps H, Nass L, Silvester H. NatCen. 2014. Research using social media; users' views URL: http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf [accessed 2017-04-25] [WebCite Cache ID 6pyzSYOf6]
41. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
42. Salmons J, Woodfield K. SRA. London: NatCen; 2013. Social media, social science & research ethics URL: http://the-sra.org.uk/wp-content/uploads/salmons_woodfield.pdf [accessed 2017-04-25] [WebCite Cache ID 6pyzUiA2j]
43. Alim S. An initial exploration of ethical research practices regarding automated data extraction from online social media user profiles. First Monday 2014;19(7):105-127. [doi: 10.5210/fm.v19i7.5382]
44. Markham A, Buchanan E. AOIR. Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0) URL: http://aoir.org/reports/ethics2.pdf [accessed 2017-04-25] [WebCite Cache ID 6pyzXJFZ0]
45. Im E, Chee W. Practical guidelines for qualitative research using online forums. Comput Inform Nurs 2012 Nov;30(11):604-611 [FREE Full text] [doi: 10.1097/NXN.0b013e318266cade] [Medline: 22918135]
46. British Psychological Society. BPS. Leicester; 2013. Ethics guidelines for internet-mediated research URL: http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf [accessed 2017-04-25] [WebCite Cache ID 6pyzZbMN2]

47. Liberty University Institutional Review Board. Liberty. 2012. Guidelines for internet research with human subjects URL: https://www.liberty.edu/media/9997/Internet_Research.pdf [accessed 2017-04-25] [WebCite Cache ID 6pyzak1yy]
48. Townsend L, Wallace C. Dotrural.: The University of Aberdeen; 2017. Social media research: a guide to ethics URL: http://www.dotrural.ac.uk/socialmediaresearchethics.pdf [accessed 2017-05-03] [WebCite Cache ID 6qBEyf6S3]
49. The Norwegian National Research Ethics Committees. Etikkom. 2014. Ethical guidelines for internet research URL: https://www.etikkom.no/en/ethical-guidelines-for-research/ethical-guidelines-for-internet-research/ [accessed 2017-04-25] [WebCite Cache ID 6pyzdR0Xz]
50. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
51. Carroll C, Booth A, Lloyd-Jones M. Should we exclude inadequately reported studies from qualitative systematic reviews? an evaluation of sensitivity analyses in two case study reviews. Qual Health Res 2012 Oct;22(10):1425-1434. [doi: 10.1177/1049732312452937] [Medline: 22865107]
52. Conway M. Ethical issues in using twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. J Med Internet Res 2014;16(12):e290 [FREE Full text] [doi: 10.2196/jmir.3617] [Medline: 25533619]
53. Dixon-Woods M. Using framework-based synthesis for conducting reviews of qualitative studies. BMC Med 2011 Apr 14;9:39 [FREE Full text] [doi: 10.1186/1741-7015-9-39] [Medline: 21492447]
54. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
55. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
56. Bakardjieva M, Feenberg A. Involving the virtual subject. Ethics Inf Technol 2001;2:233-240.
57. Bond CS, Ahmed OH, Hind M, Thomas B, Hewitt-Taylor J. The conceptual and practical ethical dilemmas of using health discussion board posts as research data. J Med Internet Res 2013;15(6):e112 [FREE Full text] [doi: 10.2196/jmir.2435] [Medline: 23748147]
58. Chen S, Hall G, Johns M. Research paparazzi in cyberspace: the voices of the researched. In: Johns MF, Chen SS, Hall GJ, editors. Online Social Research: Methods, Issues, and Ethics. New York: Peter Lang; 2004:157-175.
59. Evans H, Ginnis S, Barlett J. Ipsos. 2015. #SocialEthics a guide to embedding ethics in social media research URL: https://www.ipsos-mori.com/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf [accessed 2017-04-25] [WebCite Cache ID 6pyzmUaiL]
60. Hudson J, Bruckman A. 'Go Away': participant objections to being studied and the ethics of chatroom research. Inform Soc 2004;20:127-139.
61. Hudson J, Bruckman A. Using empirical data to reason about internet research ethics. 2005 Presented at: ECSCW : Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work; 2005; Paris, France p. 287-306. [doi: 10.1007/1-4020-4023-7_15]
62. Michaelidou N. Lboro. 2016. Social media as a means of collecting information URL: http://www.lboro.ac.uk/departments/sbe/inspire-08/ethicality-social-media/ [accessed 2017-04-25] [WebCite Cache ID 6pzYu07tb]
63. Michaelidou N, Moraes C, Micevski M. A scale for measuring consumers? ethical perceptions of social media research. In: Let's Get Engaged! Crossing the Threshold of Marketing's Engagement Era. Cham: Springer; 2016:97-100.
64. Mikal J, Hurst S, Conway M. Ethical issues in using twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
65. Monks H, Cardoso P, Papageorgiou A, Carolan C, Costello L, Thomas L. Young people's views regarding participation in mental health and wellbeing research through social media. Int J Emot Educ 2015;7(1):4-19.
66. Petersen C. Permission, privacy are key concerns in internet research. BMJ 2003;323:1103.
67. Williams M, editor. Socialdatalab. 2015. Towards an ethical framework for using social media data in social research URL: http://socialdatalab.net/wp-content/uploads/2016/08/EthicsSM-SRA-Workshop.pdf [accessed 2017-04-25] [WebCite Cache ID 6pzZ9S8zP]
68. Carter CJ, Koene A, Perez E, Statache R, Adolphs S, O'Malley C. Understanding academic attitudes towards the ethical challenges posed by social media research. 2015 Presented at: ACM SIGCAS Computers and Society - Special Issue on Ethicomp; September 2015; New York p. 202-210 URL: http://dl.acm.org/citation.cfm?id=2874268
69. Denecke K. Ethical aspects of using medical social media in healthcare applications. Stud Health Technol Inform 2014;198:55-62. [Medline: 24825685]
70. McKee H, Porter J. The ethics of internet research: a rhetoical, case-based process. New York: Peter Lang Publishing; 2009.
71. Woodfield K, Morrell G, Metzler K, Blank G, Salmons J, Finnegan J. NCRM. Southampton: National Centre for Research Methods; 2013. Blurring the boundaries? new social media, new social research: developing a network to explore the issues faced by researchers negotiating the new research landscape of online social media platforms: a methodological review paper URL: http://eprints.ncrm.ac.uk/3168/1/blurring_boundaries.pdf [accessed 2017-04-25] [WebCite Cache ID 6pzZlVwdw]

72. Brownlow C, O'Dell L. Ethical issues for qualitative research in on-line communities. Disabil Soc 2002;17(6):685-694.
73. Henderson M, Johnson N, Auld G. Silences of ethical practice: dilemmas for researchers using social media. Educ Res Eval 2013;19(6):546-560.
74. Trevisan F, Reilly P. Ethical dilemmas in researching sensitive issues online: lessons from the study of British disability dissent networks. Inform Commun Soc 2013;17(9):1131-1146.
75. Battles H. Exploring ethical and methodological issues in internet-based research with adolescents. Int J Qual Methods 2010;9(1):27-39.
76. King SA. Researching internet communities: proposed ethical guidelines for the reporting of results. Inform Soc 1996;12(2):119-128.
77. Koene A, Perez E, Carter C, Statache R, Adolphs S, O?Malley C. Research ethics and public trust, preconditions for continued growth of internet mediated research. 2015 Presented at: ICISSP 2015-1st International Conference on Information Systems Security and Privacy; 2015; Anglers, France URL: http://casma.wp.horizon.ac.uk/wp-content/uploads/2015/02/ICISSP_2015_60_CaSMa.pdf
78. Gustafson DL, Woodworth CF. Methodological and ethical issues in research using social media: a metamethod of Human Papillomavirus vaccine studies. BMC Med Res Methodol 2014;14:127 [FREE Full text] [doi: 10.1186/1471-2288-14-127] [Medline: 25468265]
79. Zimmer M, Proferes N. A topology of Twitter research: disciplines, methods, and ethics. Aslib Journal of Information Management 2014;66(3):250-261.
80. Dawson P. Our anonymous online research participants are not always anonymous: is this a problem? Br J Educ Technol 2014;45(3):428-437.
81. Barnes S. Issues of attributionidentification in online social research. In: Johns MD, Chen SS, Hall GJ, editors. Online Social Research: Methods, Issues, and Ethics. New York: Peter Lang; 2004:203-222.
82. Schrum L. Ethical research in the information age: Beginning the dialog. Comput Human Behav 1997 May;13(2):117-125. [doi: 10.1016/s0747-5632(97)80001-5]
83. Zimmer M. “But the data is already public”: on the ethics of research in Facebook. Ethics Inform Technol 2010;12(4):313-325.
84. Stevens G, O'Donnell VL, Williams L. Public domain or private data? developing an ethical approach to social media research in an inter-disciplinary project. Educ Res Eval 2015;21(2):154-167.
85. Taylor HA, Kuwana E, Wilfond BS. Ethical implications of social media in health care research. Am J Bioeth 2014;14(10):58-59. [doi: 10.1080/15265161.2014.947820] [Medline: 25229593]
86. Boyd DC. Critical questions for big data. Information, Communication & Society 2012;15(5):662-679.
87. DeCamp M. Ethical issues when using social media for health outside professional relationships. Int Rev Psychiatry 2015 Apr;27(2):97-105. [doi: 10.3109/09540261.2014.1001726] [Medline: 25738215]
88. Eastham LA. Research using blogs for data: public documents or private musings? Res Nurs Health 2011 Aug;34(4):353-361. [doi: 10.1002/nur.20443] [Medline: 21626517]
89. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
90. Hopewell-Kelly N, Baillie J, Sivell S, Harrop E, Bowyer A, Taylor S, et al. Palliative care research centre's move into social media: constructing a framework for ethical research, a consensus paper. BMJ Support Palliat Care 2016 Jan 28:28. [doi: 10.1136/bmjspcare-2015-000889] [Medline: 26823291]
91. Information Commissioner's Office. 2014. Big data and data protection URL: https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf [accessed 2017-04-25] [WebCite Cache ID 6pzakfnqX]
92. Roberts L. Ethical issues in conducting qualitative research in online communities. Qual Res Psychol 2015;12(3):314-325.
93. Convery I, Cox D. A review of research ethics in internet-based research. Practitioner Research in Higher Education 2012;6(1):1 [FREE Full text]
94. Barbaro M, Zeller T. A face is exposed for AOL searcher no. 4417749. New York Times 2006:9.
95. Parry M. Harvard researchers accused of breaching students' privacy. The Chronicle of Higher Education 2011 Jul:10.
96. Brotsky SR, Giles D. Inside the “pro-ana” community: a covert online participant observation. Eat Disord 2007;15(2):93-109. [doi: 10.1080/10640260701190600] [Medline: 17454069]
97. Booth R. The Guardian. 2014. Facebook reveals news feed experiment to control emotions URL: https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds [accessed 2017-04-25] [WebCite Cache ID 6pz0eBj8o]
98. Kleinsman J, Buckley S. Facebook study: a little bit unethical but worth it? J Bioeth Inq 2015 Jun;12(2):179-182. [doi: 10.1007/s11673-015-9621-0] [Medline: 25740760]

Edited by G Eysenbach; submitted 01.12.16; peer-reviewed by C Bond, K Gardner, M Conway; comments to author 23.02.17; revised version received 13.03.17; accepted 30.03.17; published 06.06.17.

Please cite as:
Golder S, Ahmed S, Norman G, Booth A
Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review
J Med Internet Res 2017;19(6):e195
URL: http://www.jmir.org/2017/6/e195/
doi:10.2196/jmir.7082
PMID:28588006

Original Paper

A Medical Student-Delivered Smoking Prevention Program, Education Against Tobacco, for Secondary Schools in Germany: Randomized Controlled Trial

Titus Josef Brinker[1,2,3,4,5], MD; Andreas Dawid Owczarek[1], MD; Werner Seeger[1], PhD, MD; David Alexander Groneberg[2], MD, PhD; Christian Martin Brieske[3,4,5]; Philipp Jansen[3,4,5], MD; Joachim Klode[3,4,5], MD; Ingo Stoffels[3,4,5], MD; Dirk Schadendorf[3,4,5], MD, PhD; Benjamin Izar[6], Ph.D., MD; Fabian Norbert Fries[7], MD; Felix Johannes Hofmann[1]

[1]Universities of Giessen and Marburg Lung Center; Member of the German Center for Lung Research (DZL), Department of Internal Medicine, Justus-Liebig-University of Giessen, Giessen, Germany

[2]Institute of Occupational Medicine, Social Medicine and Environmental Medicine, Goethe-University, Frankfurt am Main, Germany

[3]Essen University Hospital, Department of Dermatology, Venerology and Allergology, University of Duisburg-Essen, Essen, Germany

[4]West German Cancer Center, University Duisburg-Essen, Essen, Germany

[5]German Cancer Consortium (DKTK), University of Heidelberg, Heidelberg, Germany

[6]Dana-Farber Cancer Institute, Department of Medical Oncology, Boston, MA, United States

[7]Saarland University Medical Center and Saarland University Faculty of Medicine, Department of Ophthalmology, University of Homburg, Homburg, Germany

Corresponding Author:
Titus Josef Brinker, MD
Universities of Giessen and Marburg Lung Center; Member of the German Center for Lung Research (DZL)
Department of Internal Medicine
Justus-Liebig-University of Giessen
Klinikstr. 33
Giessen, 35392
Germany
Phone: 49 15175084347
Email: titus.brinker@gmail.com

Abstract

Background: More than 8.5 million Germans suffer from chronic diseases attributable to smoking. Education Against Tobacco (EAT) is a multinational network of medical students who volunteer for school-based prevention in the classroom setting, amongst other activities. EAT has been implemented in 28 medical schools in Germany and is present in 13 additional countries around the globe. A recent quasi-experimental study showed significant short-term smoking cessation effects on 11-to-15-year-old adolescents.

Objective: The aim of this study was to provide the first randomized long-term evaluation of the optimized 2014 EAT curriculum involving a photoaging software for its effectiveness in reducing the smoking prevalence among 11-to-15-year-old pupils in German secondary schools.

Methods: A randomized controlled trial was undertaken with 1504 adolescents from 9 German secondary schools, aged 11-15 years in grades 6-8, of which 718 (47.74%) were identifiable for the prospective sample at the 12-month follow-up. The experimental study design included measurements at baseline (t1), 6 months (t2), and 12 months postintervention (t3), via questionnaire. The study groups consisted of 40 randomized classes that received the standardized EAT intervention (two medical student-led interactive modules taking 120 minutes total) and 34 control classes within the same schools (no intervention). The primary endpoint was the difference in smoking prevalence from t1 to t3 in the control group versus the difference from t1 to t3 in the intervention group. The differences in smoking behavior (smoking onset, quitting) between the two groups, as well as gender-specific effects, were studied as secondary outcomes.

Results: None of the effects were significant due to a high loss-to-follow-up effect (52.26%, 786/1504). From baseline to the two follow-up time points, the prevalence of smoking increased from 3.1% to 5.2% to 7.2% in the control group and from 3.0%

to 5.4% to 5.8% in the intervention group (number needed to treat [NNT]=68). Notable differences were observed between the groups for the female gender (4.2% to 9.5% for control vs 4.0% to 5.2% for intervention; NNT=24 for females vs NNT=207 for males), low educational background (7.3% to 12% for control vs 6.1% to 8.7% for intervention; NNT=30), and migrational background (students who claimed that at least one parent was not born in Germany) at the 12-month follow-up. The intervention appears to prevent smoking onset (NNT=63) but does not appear to initiate quitting.

Conclusions: The intervention appears to prevent smoking, especially in females and students with a low educational background.

(J Med Internet Res 2017;19(6):e199) doi:10.2196/jmir.7906

KEYWORDS

medical students; tobacco prevention; secondary schools; smoking cessation; adolescents; school-based prevention

Introduction

Most smokers start smoking during their early adolescence with the idea that smoking is glamorous; the problems related to vascular disease, lung cancer, chronic pulmonary disease and cataracts are too far in the future to fathom [1]. After multiple failed quit attempts, however, many smokers end up with tobacco-related diseases that physicians are unable to cure [2]. Despite the fact that effectiveness of inpatient smoking cessation was demonstrated in major trials [3] and was implemented in guidelines of almost all medical specialties [4], research has shown that physicians in Germany lack both the motivation (eg, role incongruence as a major barrier [5,6]) and education to deliver smoking cessation advice [5-8], especially before the onset of chronic disease [6]. The issue of undertreatment of tobacco use by physicians is known on a global scale [9,10]. It is estimated that global mortality attributable to tobacco will double from 5 million (2010) to 10 million per year in the coming decades [2].

Education Against Tobacco (EAT) is a multinational network of medical students that aims to provide science-based tobacco prevention to a large number of adolescents, and also sensitize prospective physicians to the importance of inpatient smoking cessation [11,12]. The network currently involves approximately 80 medical schools in 14 countries, with 1500 medical students educating more than 40,000 secondary school students in the classroom setting per year, while using and optimizing apps and strategies [13-16]. Since its foundation in Germany in 2012, the network has also instructed science-based smoking cessation curricula in 13 medical schools in Germany (of the 28 medical schools in Germany that take part in EAT) that are currently prospectively evaluating their effectiveness in preparing medical students for inpatient smoking cessation [7]. The two free science-based quit apps of EAT (*Smokerface* and *Smokerstop*) are downloaded more than a 1000 times per day and have been translated into most commonly spoken languages [14,15].

The 2016 HBSC international report by the WHO revealed that 13% of German boys and 15% of German girls at 15 years of age smoke cigarettes at least once a week [17]. Despite the decline in adolescent smoking over the last two decades, prevalence in Germany is still high and strong socioeconomic differences exist [17-19].

Current Knowledge of School-Based Tobacco Prevention

Most school-based tobacco prevention curricula are ineffective and the evaluation of new curricula is mandatory [20]. A recently published evaluation of a short student and student-parent program from Germany did not show significant effectiveness among 7th grade students (7.6% and 7% prevalence in intervention groups vs 10.1% control group) at a 24-month follow-up. However, this result was mostly due to a too low sample size: only 47 schools were randomized due to an underestimated intracluster correlation coefficient [21,22]. The largest tobacco prevention program for secondary schools in Germany, the Smoke-Free Class Competition, has demonstrated limited effectiveness in increasing knowledge and making students initiate a quit attempt, but was not able to prevent smoking onset [23-25].

Physician-based programs relying on fear-inducing statements show no overall long-term effectiveness in reducing smoking prevalence [26-29], while limited new evidence suggests that asking questions about health consequences (rather than making statements) might be more effective to motivate current smokers to make a quit attempt [30].

A physician-based multimodal program in Berlin was evaluated in a quasi-experimental study, and showed significant short-term effects in preventing smoking onset, which might be a promising alternative to the traditional fear-based approaches of physician programs [31]. Outside of schools, a systematic review of inpatient physician-based smoking prevention and cessation programs for adolescents revealed that behavioral interventions show overall effectiveness in primary care [32].

Previous Research on Education Against Tobacco

The effectiveness of an old version of the EAT curriculum on reducing smoking prevalence among German adolescents has only been investigated with a quasi-experimental design that contained potential sources of bias [11,12]. However, the study showed a significant association for reducing the smoking prevalence of secondary school students in Germany at 6-month follow-up by motivating them to make a quit attempt (n=1474 students). After this first evaluation, the curriculum was optimized for students with a lower educational level by using cognitive interviewing (we asked the students after the interventions what they found most convincing). The curriculum received more age-appropriate content, was optimized to be more interactive and gain-framed [33], and was equipped with a photoaging software [11,14].

Introduction of the Education Against Tobacco App Smokerface

Photoaging desktop programs, in which an image is altered to predict future appearance, were effective in motivating 14-to-18-year-old females to quit smoking and increased the rate of quit attempts in 18-to-30-year-old young adults of both genders by 21% [34,35]. The broad availability of smartphones and adolescents' interest in their appearance [36] were harnessed to create the free 3D-photoaging smartphone app *Smokerface* [15] which animates the users' selfies and reacts to touch (Multimedia Appendix 1). This app is downloaded 200 times per day and the current version of the app has a rating of 4.2/5 stars in the Playstore (Android, USA). This app was implemented via a poster-campaign in German secondary schools and is currently being evaluated in a large multicentered trial [14,37]. In 2014, the EAT curriculum was only available as a software program that was run on notebook computers, which captured participant's faces via webcams.

The aim of this study was to provide the first randomized long-term evaluation of the optimized 2014 EAT curriculum involving a photoaging software for its effectiveness in reducing the smoking prevalence among 11-to-15-year-old pupils in German secondary schools.

Methods

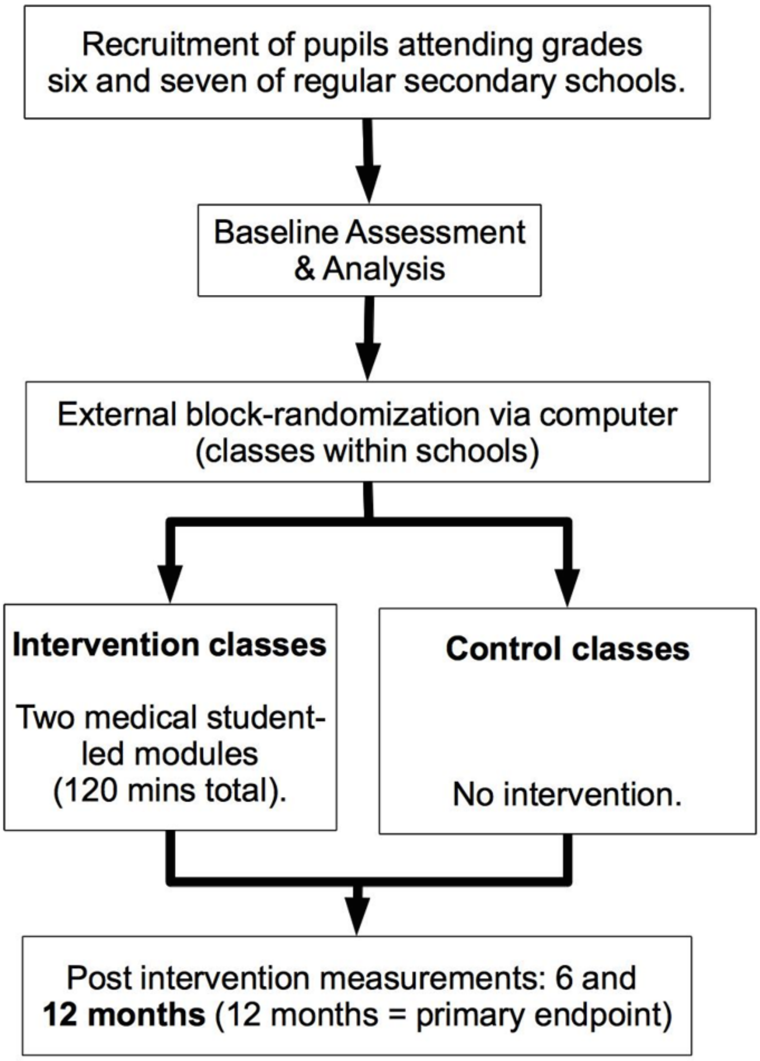
Design

The survey was designed as a randomized controlled trial with three measurements (baseline, 6 months, and 12 months postintervention) [12]. The study period was October 2014 until March 2016. Participants in the two study groups (intervention and control) were questioned up to 2 weeks in advance of the intervention (t1), and 6 months (t2) and 12 months (t3) thereafter (Figure 1).

Randomization

In accordance with the recommendations of the Cochrane Analysis [20], randomization was externally and centrally performed by a statistician from the University of Frankfurt, Germany on the class level within each school via block randomization. Due to the fact that the statistician performed the randomization one school at a time, there was a slight imbalance between groups (40 intervention and 34 control classes).

Figure 1. Study design.



Participants

A total of 1504 eligible secondary school students from 74 classes (from 9 eligible schools) entered baseline data. All participants fulfilled the inclusion criteria. Students aged 10-15 years attending grades 6-8 of a secondary general, intermediate, grammar, or comprehensive school were eligible [12]. Baseline data were collected from October 2014 to March 2015. Follow-up data were collected from April 2015 to March 2016. A total of 718 participants provided data at both time points (t1 and t3) that was used for primary endpoint analyses. The loss to follow-up effect was 52.26% (786/1504) due to problems with the identifier code (see Discussion section).

Attrition Analysis

The participants who dropped out at follow-up (t3) were analyzed with logistic regression analysis and showed no systematic bias regarding the interaction between: study group and smoking status (P=0.520), study group and gender (P=0.131), study group and age (P=0.427), or study group and school type (grammar school vs lower educational school type; P=0.440). However, there was a systematic bias regarding gender as a main effect (odds ratio [OR]=0.511, 95% CI 0.412-0.633, P<0.001) meaning that more boys dropped out than girls. There was a systematic difference between groups regarding school type. The interaction between study group (0=control and 1=intervention) and school type (0=no grammar school and 1=grammar school) was significant (OR=0.618, 95% CI 0.410-0.933, P=0.022). In the intervention group, the dropout was lower in grammar schools, whereas the dropouts were at similar levels in the control group (intervention group: grammar 37.9% [165/435], no grammar 55.0% [213/387]; control group: grammar 45.9% [174/379], no grammar 49.8% [151/303]). Dropouts include cases that could not be matched due to difficulties with the identifier code.

Intervention

The intervention was the 2014 EAT curriculum, which consisted of two interactive 60-minute modules delivered by medical students from the University of Giessen. The medical students did not volunteer but had the duty to perform a school visit based on their participation in the EAT course. The app version of *Smokerface* was available at the time but was not advertised or mentioned to the students, to avoid contamination of the control group. The medical students received standardized training that lasted 45 minutes. Students were asked in advance to read through the classroom curriculum and then met with an experienced medical student who went through all key procedures of the intervention.

The first part of the intervention was presented by two-to-six medical students and a patient with a tobacco-related disease to all pupils in a large room within the school. The presentation consisted of a PowerPoint (Microsoft; Redmond, WA, USA) presentation which aimed at addressing age-appropriate topics in an interactive manner to help the students reframe a positive nonsmoking image, which included: physical performance (with the example of famous German soccer players), saved money, addiction versus freedom, examples of tobacco advertising and how the students would advertise tobacco products if given the opportunity, attractiveness, and photoaged images. At the end of each presentation, a patient with a tobacco-related disease (chronic obstructive pulmonary disease) was introduced and then interviewed by the medical students and pupils.

The second part of the intervention consisted of a classroom seminar and a photoaging intervention for every individual pupil. Two medical students tutored one school class consisting of approximately 25 students in an interactive manner. While one medical student presented the curriculum, groups of approximately five pupils were sent to another room for the morphing procedure with an older version of the self-developed *Smokerface* software, which was in 2D (Figure 2) [15]. Topics included: skin damage, harm to the rest of the body, drop in physical performance, growth during adolescence [38], freedom and independence, advertising, and cost. The full 2014 classroom curriculum can be accessed online in English [13].

Photoaging Intervention

Four Macbook Air (Apple, California) computers were purchased to run the photoaging intervention in four classes simultaneously, and to make sure that every single pupil got their face photoaged. Wireless Local Area Network (W-LAN) sticks were also purchased, as many schools did not have W-LAN at the time but the app had to communicate with servers to perform the morphing process (the current version of the app runs offline). Every student received the photoaging intervention; groups of five pupils were sent out to an external room where one medical student per class was taking pictures with the webcam and then showed them the predicted result of smoking on their own faces (Figure 3). The pupils chose whether they kept their image strictly for themselves or shared it with friends. Each student got a brief explanation of the skin changes by the medical student, which was reinforced in the classroom.

Data Collection

Data were collected via a published questionnaire that was used in the previous investigation of the same age group, which was optimized by cognitive interviewing [11,39]. All items were based on three established studies declared to be high quality by the recent Cochrane Analysis and were either used in their original form or adapted to the specific circumstances of the recent study [20,40-42]. In addition to sociodemographic data (age, gender, school type), the questionnaire captured the smoking status of the school students concerning e-cigarette, water-pipe, and cigarette consumption. The only alteration to the protocol was that data entry was not manually performed; a scanning software was used to make the process less time consuming. However, for this approach to work the identifier code was changed from letters and numbers to numbers only, which resulted in a large fraction of students that were not identifiable at follow-up.

Outcomes

The primary endpoint was the difference in smoking prevalence from t1 to t3 in the control group versus the difference from t1 to t3 in the intervention group. The differences in smoking behavior (smoking onset, quit attempts) between the two groups, as well as gender-specific effects, were studied as secondary outcomes.

Figure 2. Effects of the first version of the Smokerface software used in the study; left image: original picture; middle: normal aging for 15 years; right: smoking a pack a day for 15 years.

Figure 3. "Facemorphing room" hosted by three medical students in a school with three classes.

Statistical Analyses

To examine baseline differences χ^2tests, t-tests and Mann-Whitney U Tests were used (see Multimedia Appendix 2 for details). The effects of predictors (gender, culture, and social characteristics) on smoking behavior after 12 months (t3) were calculated by robust panel logistic regression analysis. The significance level was 5% for t-tests (double-sided) and 95% for CIs (double-sided). Statistical analyses were performed using SPSS Statistics Version 23 (IBM; Armong, USA) and STATA 14 (StataCorp; Texas, USA). The group allocation in the sample was based on the class level. To account for this clustering, statistically robust panel logistic regression was used (xtlogit procedure with vce [cluster] option). This procedure was also used to calculate the difference from t1 to t3 of the smoking prevalence in the control group versus the difference from t1 to t3 in the intervention group (the primary endpoint) by using STATA 14. The number needed to treat (NNT) calculation was performed without controlling for baseline differences, as these data were comparable due to randomization. The NNT was calculated for the total effect (preventing smoking onset and initiating quit attempts) and for quit attempts and smoking onset individually. Any NNT values higher than 200 were not reported due to lack of relevance.

Legal Approval

In accordance with Good Epidemiologic Practice guidelines, an ethics waiver and all legal permissions were obtained from the responsible institutions before data collection started. Written informed consent was obtained from all students and their parents. Teachers had to be present during the intervention and school personnel in charge of possible adverse health events were present in each school visited. No adverse events occurred.

Table 1. Cigarette smoking prevalence at baseline (t1), 6 months (t2), and 12 months (t3).

	Time point	Intervention and Control Group	Intervention Group	Control Group	Number Needed to Treat
Total	t1	22/718 (3.1%)	12/400 (3.0%)	10/318 (3.1%)	
	t2	42/788 (5.3%)	23/425 (5.4%)	19/363 (5.2%)	
	t3	46/718 (6.4%)	23/400 (5.8%)	23/318 (7.2%)	68
Grammar schools	t1	4/446 (0.9%)	3/252 (1.2%)	1/194 (0.5%)	
	t2	8/473 (1.7%)	4/251 (1.6%)	4/222 (1.8%)	
	t3	17/406 (4.2%)	9/227 (4.0%)	8/179 (4.5%)	199
Lower educational background	t1	18/272 (6.6%)	9/148 (6.1%)	9/124 (7.3%)	
	t2	34/315 (10.8%)	19/174 (10.9%)	15/141 (10.6%)	
	t3	28/275 (10.2%)	13/150 (8.7%)	15/125 (12%)	30
Male	t1	6/280 (2.1%)	3/146 (2.1%)	3/134 (2.2%)	
	t2	9/325 (2.8%)	4/168 (2.4%)	5/157 (3.2%)	
	t3	12/264 (4.5%)	6/139 (4.3%)	6/125 (4.8%)	207
Female	t1	16/392 (4.1%)	9/226 (4.0%)	7/166 (4.2%)	
	t2	28/392 (7.1%)	16/216 (7.4%)	12/176 (6.8%)	
	t3	27/387 (7.0%)	12/229 (5.2%)	15/158 (9.5%)	24
Migrational background: no	t1	15/480 (3.1%)	7/264 (2.7%)	8/216 (3.7%)	
	t2	24/507 (4.7%)	13/259 (5.0%)	11/248 (4.4%)	
	t3	31/476 (6.5%)	15/263 (5.7%)	16/213 (7.5%)	56
Migrational background: yes	t1	6/211 (2.8%)	4/123 (3.3%)	2/88 (2.3%)	
	t2	13/231 (5.6%)	9/136 (6.6%)	4/95 (4.2%)	
	t3	14/197 (7.1%)	7/114 (6.1%)	7/83 (8.4%)	44

Results

Baseline Data

The mean age of the 1504 eligible participants at baseline (Multimedia Appendix 2) was 12.4 years (range 11-15 years) and 49.17% (681/1385) were female. Of the participants, 54.12% (814/1504) attended grammar schools and the remaining 45.88% (690/1504) attended comprehensive schools (which were classified in the survey as *lower education level*). The survey identified 3.89% (55/1414) of participants as cigarette smokers at baseline. There were no significant differences concerning the number of cigarette smokers in each group (*P*=0.797; Multimedia Appendix 2).

Follow-Up Data

Analyses of the data were based on the originally assigned groups (see Table 1). There were 400 pupils in the intervention group and 318 pupils in the control group who had participated in the survey at both relevant time points that could be identified (baseline sample=1504; prospective sample=718 pupils; loss to follow-up=786 pupils).

None of the effects were significant due to a high loss-to-follow-up effect (52.26%, 786/1504) including the primary endpoint. From baseline to the two follow-up time points, the smoking prevalence increased from 3.1% to 5.2% to 7.2% in the control group and from 3% to 5.4% to 5.8% in the intervention group (NNT=68) with notable nonsystematic effects for the female gender (4.2% to 9.5% for control vs 4% to 5.2% for intervention; NNT=24 for females vs NNT=207 for males), low educational background (7.3% to 12% for control vs 6.1% to 8.7% for intervention; NNT=30), and migrational background (students who claimed that at least one parent was not born in Germany) at 12-month follow-up. The intervention appears to prevent smoking onset (NNT=63) but does not appear to initiate quit attempts. Details on smoking prevalence among subgroups can be found in Table 1 (including NNT).

Primary Endpoint

There was no significant effect for the defined primary endpoint (OR=0.74; 95% CI 0.21-2.56; *P*=0.63) calculated with the prospective sample of 718 participants (Table 2). The percentage of students who claimed to be smokers increased from 3.1% (t1) to 7.2% (t3) in the control group, but the increase was less dramatic in the intervention group (3% [t1] to 5.4% [t3]).

Secondary Outcomes

At the 12-month follow-up, 17 of 318 control group students (5.3%) had started smoking but only 15 of 400 intervention group students had started smoking (3.8%; Table 3). No secondary preventive effects (quit attempts) were noted.

Table 2. Primary endpoint calculated by robust panel logistic regression (xtlogit procedure with vce [cluster] option). Difference in smoking prevalence from t1 to t3 of the smoking prevalence in the control group versus the difference from t1 to t3 in the intervention group (see Methods section).

	Odds Ratio	Standard Error	*P*-value	Lower-CI	Upper-CI
Complete sample, strongly balanced (n=718)	0.74	0.47	0.630	0.21	2.56
Only nongrammar schools (n=272)	0.72	0.45	0.606	0.21	2.47
Only female students (n=417)	0.44	0.40	0.368	0.08	2.60
Only students with migration background (n=206)	0.84	0.74	0.848	0.15	4.76

Table 3. Nominal and percentage effects of the intervention on the smoking status (secondary outcomes, from t1 to t3).

	Prospective smoking status (t1-t3)			
	Remains nonsmoker	Commences smoking	Ceases smoking	Remains smoker
Control Group, n (%)	291 (91.5)	17 (5.3)	4 (1.3)	6 (1.9)
Intervention Group, n (%)	373 (93.2)	15 (3.8)	4 (1.0)	9 (2.0)
Total, N	664	32	8	14

Discussion

This is the first long-term evaluation of a photoaging intervention to prevent smoking and the first completed randomized trial on medical-student-delivered school-based tobacco prevention to date [16]. The present study suggests that photoaging is effective at preventing smoking onset, especially in female students (NNT=24) and students with a low educational (NNT=30) or migrational backgrounds (NNT=44).

Interpretation

Available cross-sectional data reveals that photoaging interventions are effective in motivating 14-18-year-old female smokers to make a quit attempt [35], so it was hypothesized that secondary preventive effects would be present in the sample, which was not the case. However, the intervention showed a smaller NNT for females versus males in preventing smoking onset (NNT=24 for females vs 207 for males) and the uptake of smoking for females in the intervention group was lower after 6-month and 12-month follow-ups compared to males exposed to the intervention. These data reflect the findings of a recent study by Baudson et al that was conducted with 2950 German adolescents of both genders aged 10-19 years, which showed that self-concept of appearance is the strongest predictor for self-esteem, and that this is especially true for girls and adolescents from schools with a low educational level [36]. It is notable that our data shows an NNT of 30 for nongrammar schools but an NNT of 199 for grammar schools.

However, a recent study demonstrated that the theory of planned behavior needs to be taken into account when implementing photoaging in school settings with adolescents [14]. It is possible that the intervention could not show an effect on males due to a lack of group effects and elements that increase catamnesis (such as posters [37]), as other data also indicate that photoaging interventions are effective for both genders [14,34,43].

The implementation of cost-effective measures to prevent smoking in adolescents and, moreover, the sensitization of prospective physicians to tobacco-attributable diseases, tobacco prevention, and improved communication of these issues in medicine, is addressed by the program [5,9,44,45].

Limitations

While the groups were successfully randomized and provided comparable baseline data, an unusually high loss-to-follow-up effect led to no significant results to report. However, the descriptive data collected at three different time points and effect size estimates, such as number needed to treat (which is widely accepted as an indicator for clinical relevance), allowed for data interpretation [46]. The photoaging software became available as an app during the study period and was advertised via television. Thus, a small portion of the control group might have been exposed to the photoaging intervention. In addition to this variable, intervention and control classes were in the same schools, which made it possible for the curriculum content to be exchanged. Furthermore, in some cases teachers would not adhere to the handouts pertaining to nondisclosure towards the control classes and disclose information on curriculum content. A follow-up study should use schools, but not classes, as a cluster.

Our study relied on self-reports obtained from adolescents via questionnaires that has been critically reviewed by Gorber et al in 2009, in which cotinine saliva testing was suggested [47]. However, Gorber et al did not take into account e-cigarettes, which are a limitation to cotinine saliva testing (because they also may contain nicotine). Additionally, saliva testing was prohibited by the ministry of cultural affairs in Germany, and because this is a randomized study, the influence of social desirability bias should have the same influence in both groups.

Dissemination of the Intervention

Approximately 5 years after EAT was founded (January 2012), the program has more participating mentors (1500 medical students) and interactively educates more secondary school students per year (40,000) than any other known school-based physician-delivered or medical-student-delivered tobacco prevention program in Germany or, to our knowledge, worldwide. This program is currently present in over 80 medical schools in 14 countries. The apps used by the medical students in the classroom are freely available around the globe and have been translated into the six most commonly spoken languages worldwide.

Conclusions

In conclusion, the 2014 EAT curriculum focusing on photoaging aspects of smoking appears to be most effective in females or students with a low educational background, but appears to lack effectiveness in grammar school students and male adolescents. Further research and long-term evaluation in sufficiently powered trials, as well as new ways of implementation, are needed to further evaluate and optimize our program.

Acknowledgements

We would like to thank the volunteer medical students Hannes Tabert, Svea Holtz, Stefan Henkel, Laura Schwab, Thorben Sämann, Andreas Owczarek, Felix Neumann, Anika Wolf, Dominik Penka, and Lorena Steinbach for their strong engagement within the EAT program, which made the relatively large sample size possible. This study was funded by the German Center for Lung Research (DZL).

Authors' Contributions

TJB conducted the study, invented, designed and organized the intervention, wrote the manuscript, contributed to the design of the study, coordinated and conducted data entry, and performed the statistical analyses. DAG and WS contributed to the design of the study, monitored data entry, and proofread the manuscript. CMB, JK, PJ, IS, BI, FNF, and DS contributed to the design of the study and analyses of data, and proofread the manuscript. FJH and ADO conducted data collection, data management and entry, contributed to the design of the study, and proofread the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video 1 (touch effect of the Smokerface App).

[MP4 File (MP4 Video), 556KB - jmir_v19i6e199_app1.mp4]

Multimedia Appendix 2

Descriptive data at baseline.

[PDF File (Adobe PDF File), 38KB - jmir_v19i6e199_app2.pdf]

Multimedia Appendix 3

CONSORT EHealth Checklist.

[PDF File (Adobe PDF File), 541KB - jmir_v19i6e199_app3.pdf]

References

1. Petrescu D, Vasiljevic M, Pepper J, Ribisl K, Marteau TM. What is the impact of e-cigarette adverts on children's perceptions of tobacco smoking? An experimental study. Tob Control 2016 Sep 05:5-8 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-052940] [Medline: 27601455]
2. Jha P, Peto R. Global effects of smoking, of quitting, and of taxing tobacco. N Engl J Med 2014 Jan 02;370(1):60-68. [doi: 10.1056/NEJMra1308383] [Medline: 24382066]
3. Mons U, Müezzinler A, Gellert C, Schöttker B, Abnet C, Bobak M, et al. Impact of smoking and smoking cessation on cardiovascular events and mortality among older adults: meta-analysis of individual participant data from prospective cohort studies of the CHANCES consortium. BMJ 2015 Apr 20;350:h1551 [FREE Full text] [Medline: 25896935]
4. Batra A, Petersen K, Hoch E, Mann K, Kröger C, Schweizer C, et al. Psychotherapy and pharmacotherapy for harmful tobacco use and tobacco dependency. Nervenarzt 2016 Jan;87(1):35-45. [doi: 10.1007/s00115-015-0037-1] [Medline: 26666768]
5. Raupach T, Falk J, Vangeli E, Schiekirka S, Rustler C, Grassi M, et al. Structured smoking cessation training for health professionals on cardiology wards: a prospective study. Eur J Prev Cardiol 2014 Jul;21(7):915-922. [doi: 10.1177/2047487312462803] [Medline: 23008136]
6. Balmford J, Leifert J, Jaehne A. “Tobacco dependence treatment makes no sense because”…: rebuttal of commonly-heard arguments against providing tobacco dependence treatment in the hospital setting. BMC Public Health 2014 Nov 19;14:1182 [FREE Full text] [doi: 10.1186/1471-2458-14-1182] [Medline: 25410166]

7. Strobel L, Schneider N, Krampe H, Beißbarth T, Pukrop T, Anders S, et al. German medical students lack knowledge of how to treat smoking and problem drinking. Addiction 2012 Oct;107(10):1878-1882. [doi: 10.1111/j.1360-0443.2012.03907.x] [Medline: 22551065]
8. Raupach T, Merker J, Hasenfuss G, Andreas S, Pipe A. Knowledge gaps about smoking cessation in hospitalized patients and their doctors. Eur J Cardiovasc Prev Rehabil 2011 Apr;18(2):334-341. [doi: 10.1177/1741826710389370] [Medline: 21450679]
9. Bernstein S, Yu S, Post L, Dziura J, Rigotti NA. Undertreatment of tobacco use relative to other chronic conditions. Am J Public Health 2013 Aug;103(8):e59-e65. [doi: 10.2105/AJPH.2012.301112] [Medline: 23763395]
10. Richter K, Ellerbeck EF. It's time to change the default for tobacco treatment. Addiction 2015 Mar;110(3):381-386. [doi: 10.1111/add.12734] [Medline: 25323093]
11. Brinker T, Stamm-Balderjahn S, Seeger W, Klingelhöfer D, Groneberg DA. Education Against Tobacco (EAT): a quasi-experimental prospective evaluation of a multinational medical-student-delivered smoking prevention programme for secondary schools in Germany. BMJ Open 2015 Sep 18;5(9):e008093 [FREE Full text] [doi: 10.1136/bmjopen-2015-008093] [Medline: 26384722]
12. Brinker TJ, Stamm-Balderjahn S, Seeger W, Groneberg DA. Education Against Tobacco (EAT): a quasi-experimental prospective evaluation of a programme for preventing smoking in secondary schools delivered by medical students: a study protocol. BMJ Open 2014 Jul 24;4(7):e004909 [FREE Full text] [doi: 10.1136/bmjopen-2014-004909] [Medline: 25059969]
13. Education Against Tobacco Website. 2014. 2014 EAT curriculum URL: http://educationtobacco.org/material/AGT_English.docx [accessed 2017-05-28] [WebCite Cache ID 6qnCzQFYC]
14. Brinker TJ, Seeger W, Buslaff F. Photoaging mobile apps in school-based tobacco prevention: the mirroring approach. J Med Internet Res 2016 Jun 28;18(6):e183 [FREE Full text] [doi: 10.2196/jmir.6016] [Medline: 27352819]
15. Brinker TJ, Seeger W. Photoaging mobile apps: a novel opportunity for smoking cessation? J Med Internet Res 2015 Jul 27;17(7):e186 [FREE Full text] [doi: 10.2196/jmir.4792] [Medline: 26215210]
16. De Freitas XL, Bernardes-Souza B, Lisboa O, Seeger W, Groneberg D, Tran T, et al. A medical student-delivered smoking prevention program, Education Against Tobacco, for secondary schools in Brazil: study protocol for a randomized trial. JMIR Res Protoc 2017 Jan 30;6(1):e16 [FREE Full text] [doi: 10.2196/resprot.7134] [Medline: 28137703]
17. Inchley J, Currie D. Growing up unequal: gender socioeconomic differences in young people's health and well-being. Health Behaviour in School-aged Children (HBSC) study: international report from the 2013/2014 survey. World Health Organization: health policy for children and adolescents 2016;2016(7):18-21.
18. Kuntz B, Lampert T. Smoking and passive smoke exposure among adolescents in Germany. Dtsch Arztebl Int 2016 Jan 22;113(3):23-30 [FREE Full text] [doi: 10.3238/arztebl.2016.0023] [Medline: 26857509]
19. Kuntz B, Lampert T. Educational differences in smoking among adolescents in Germany: what is the role of parental and adolescent education levels and intergenerational educational mobility? Int J Environ Res Public Health 2013 Jul 19;10(7):3015-3032 [FREE Full text] [doi: 10.3390/ijerph10073015] [Medline: 23877770]
20. Thomas R, McLellan J, Perera R. Effectiveness of school-based smoking prevention curricula: systematic review and meta-analysis. BMJ Open 2015 Mar 10;5(3):e006976 [FREE Full text] [doi: 10.1136/bmjopen-2014-006976] [Medline: 25757946]
21. Krist L, Lotz F, Bürger C, Ströbele-Benschop N, Roll S, Rieckmann N, et al. Long-term effectiveness of a combined student-parent and a student-only smoking prevention intervention among 7th grade school children in Berlin, Germany. Addiction 2016 Dec;111(12):2219-2229. [doi: 10.1111/add.13537] [Medline: 27447693]
22. Müller-Riemenschneider F, Krist L, Bürger C, Ströbele-Benschop N, Roll S, Rieckmann N, et al. Berlin evaluates school tobacco prevention - BEST prevention: study design and methodology. BMC Public Health 2014 Aug 23;14:871 [FREE Full text] [doi: 10.1186/1471-2458-14-871] [Medline: 25150368]
23. Isensee B, Morgenstern M, Stoolmiller M, Maruska K, Sargent J, Hanewinkel R. Effects of Smokefree Class Competition 1 year after the end of intervention: a cluster randomised controlled trial. J Epidemiol Community Health 2012 Apr;66(4):334-341. [doi: 10.1136/jech.2009.107490] [Medline: 21071561]
24. Stucki S, Kuntsche E, Archimi A, Kuntsche S. Does smoking within an individual's peer group affect intervention effectiveness? An evaluation of the Smoke-Free Class Competition among Swiss adolescents. Prev Med 2014 Aug;65:52-57. [doi: 10.1016/j.ypmed.2014.04.018] [Medline: 24786759]
25. Johnston V, Liberato S, Thomas D. Incentives for preventing smoking in children and adolescents. Cochrane Database Syst Rev 2012 Oct 17;10:CD008645. [doi: 10.1002/14651858.CD008645.pub2] [Medline: 23076949]
26. Kreuter M, Bauer C, Ehmann M, Kappes J, Drings P, Herth FJ. Efficacy and sustainability of a smoking prevention program for pupils--“ohnekippe”. Dtsch Med Wochenschr 2014 Jul;139(27):1403-1408. [doi: 10.1055/s-0034-1370129] [Medline: 24937077]
27. Thrul J, Bühler A, Herth FJ. Prevention of teenage smoking through negative information giving, a cluster randomized controlled trial. Drugs: Educ Prev Polic 2014;21(1):e.
28. Sack PM, Hampel J, Bröning S, Petersen K, Andritzky B, Andritzky B, et al. Was limitiert schulische Tabakprävention? Präv Gesundheitsf 2013 Apr 26;8(4):246-251. [doi: 10.1007/s11553-013-0388-z]

29. Kok G, Bartholomew LK, Parcel GS, Gottlieb NH, Fernández ME. Finding theory- and evidence-based alternatives to fear appeals: Intervention Mapping. Int J Psychol 2014 Apr;49(2):98-107 [FREE Full text] [doi: 10.1002/ijop.12001] [Medline: 24811880]
30. Müller BC, Ritter SM, Glock S, Dijksterhuis A, Engels RC, van Baaren RB. Smoking-related warning messages formulated as questions positively influence short-term smoking behaviour. J Health Psychol 2016 Jan;21(1):60-68. [doi: 10.1177/1359105314522083] [Medline: 24567301]
31. Stamm-Balderjahn S, Groneberg DA, Kusma B, Jagota A, Schönfeld N. Smoking prevention in school students: positive effects of a hospital-based intervention. Dtsch Arztebl Int 2012 Nov;109(44):746-752 [FREE Full text] [doi: 10.3238/arztebl.2012.0746] [Medline: 23189108]
32. Peirson L, Ali M, Kenny M, Raina P, Sherifali D. Interventions for prevention and treatment of tobacco smoking in school-aged children and adolescents: a systematic review and meta-analysis. Prev Med 2016 Apr;85:20-31. [doi: 10.1016/j.ypmed.2015.12.004] [Medline: 26743631]
33. Mays D, Niaura R, Evans W, Hammond D, Luta G, Tercyak KP. Cigarette packaging and health warnings: the impact of plain packaging and message framing on young smokers. Tob Control 2015 Mar;24(e1):e87-e92 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051234] [Medline: 24420310]
34. Burford O, Jiwa M, Carter O, Parsons R, Hendrie D. Internet-based photoaging within Australian pharmacies to promote smoking cessation: randomized controlled trial. J Med Internet Res 2013 Mar 26;15(3):e64 [FREE Full text] [doi: 10.2196/jmir.2337] [Medline: 23531984]
35. Weiss C, Hanebuth D, Coda P, Dratva J, Heintz M, Stutz EZ. Aging images as a motivational trigger for smoking cessation in young women. Int J Environ Res Public Health 2010 Sep;7(9):3499-3512 [FREE Full text] [doi: 10.3390/ijerph7093499] [Medline: 20948939]
36. Baudson TG, Weber K, Freund PA. More than only skin deep: appearance self-concept predicts most of secondary school students' self-esteem. Front Psychol 2016;7:1568 [FREE Full text] [doi: 10.3389/fpsyg.2016.01568] [Medline: 27803681]
37. Brinker TJ, Holzapfel J, Baudson TG, Sies K, Jakob L, Baumert HM, et al. Photoaging smartphone app promoting poster campaign to reduce smoking prevalence in secondary schools: the Smokerface Randomized Trial: design and baseline characteristics. BMJ Open 2016 Nov 07;6(11):e014288 [FREE Full text] [doi: 10.1136/bmjopen-2016-014288] [Medline: 27821601]
38. O'Loughlin J, Karp I, Henderson M, Gray-Donald K. Does cigarette use influence adiposity or height in adolescence? Ann Epidemiol 2008 May;18(5):395-402. [doi: 10.1016/j.annepidem.2007.12.010] [Medline: 18346909]
39. Hinds J, Loukas A, Chow S, Pasch K, Harrell M, Perry C, et al. Using cognitive interviewing to better assess young adult e-cigarette use. Nicotine Tob Res 2016 Oct;18(10):1998-2005. [doi: 10.1093/ntr/ntw096] [Medline: 27029822]
40. Peterson A, Kealey K, Mann S, Marek P, Sarason IG. Hutchinson Smoking Prevention Project: long-term randomized trial in school-based tobacco use prevention--results on smoking. J Natl Cancer Inst 2000 Dec 20;92(24):1979-1991. [Medline: 11121460]
41. Van De Ven MO, Engels RC, Otten R, Van Den Eijnden RJ. A longitudinal test of the theory of planned behavior predicting smoking onset among asthmatic and non-asthmatic adolescents. J Behav Med 2007 Oct;30(5):435-445. [doi: 10.1007/s10865-007-9119-2] [Medline: 17605099]
42. Singh T, Arrazola RA, Corey CG, Husten CG, Neff LJ, Homa DM, et al. Tobacco use among middle and high school students--United States, 2011-2015. MMWR Morb Mortal Wkly Rep 2016 Apr 15;65(14):361-367 [FREE Full text] [doi: 10.15585/mmwr.mm6514a1] [Medline: 27077789]
43. Lo PL, Chang P, Taylor MF. Young Australian adults' reactions to viewing personalised UV photoaged photographs. Australas Med J 2014;7(11):454-461 [FREE Full text] [doi: 10.4066/AMJ.2014.2253] [Medline: 25550717]
44. Anders S, Strobel L, Krampe H, Raupach T. Do final-year medical students know enough about the treatment of alcohol use disorders and smoking? Dtsch Med Wochenschr 2013 Jan;138(1-2):23-27. [doi: 10.1055/s-0032-1327367] [Medline: 23250691]
45. Paling J. Strategies to help patients understand risks. BMJ 2003 Sep 27;327(7417):745-748 [FREE Full text] [doi: 10.1136/bmj.327.7417.745] [Medline: 14512489]
46. McGough J, Faraone SV. Estimating the size of treatment effects: moving beyond p values. Psychiatry (Edgmont) 2009 Oct;6(10):21-29 [FREE Full text] [Medline: 20011465]
47. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res 2009 Jan;11(1):12-24. [doi: 10.1093/ntr/ntn010] [Medline: 19246437]

Abbreviations

EAT: Education Against Tobacco
NNT: number needed to treat
OR: odds ratio
W-LAN: Wireless Local Area Network

Edited by G Eysenbach; submitted 23.04.17; peer-reviewed by B Kuntz, O Burford; comments to author 11.05.17; revised version received 15.05.17; accepted 23.05.17; published 06.06.17.

Please cite as:
Brinker TJ, Owczarek AD, Seeger W, Groneberg DA, Brieske CM, Jansen P, Klode J, Stoffels I, Schadendorf D, Izar B, Fries FN, Hofmann FJ
A Medical Student-Delivered Smoking Prevention Program, Education Against Tobacco, for Secondary Schools in Germany: Randomized Controlled Trial
J Med Internet Res 2017;19(6):e199
URL: http://www.jmir.org/2017/6/e199/
doi:10.2196/jmir.7906
PMID:28588007

Original Paper

Integrated Detection and Prediction of Influenza Activity for Real-Time Surveillance: Algorithm Design

Armin Spreco[1*], PhD; Olle Eriksson[2*], PhD; Örjan Dahlström[3*], PhD; Benjamin John Cowling[4*], PhD; Toomas Timpka[1,5*], MD, PhD

[1]Faculty of Health Sciences, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

[2]Department of Computer and Information Science, Linköping University, Linköping, Sweden

[3]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[4]WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, The University of Hong Kong, Hong Kong, China (Hong Kong)

[5]Region Östergötland, Center for Health Services Development, Linköping, Sweden

[*]all authors contributed equally

Corresponding Author:
Armin Spreco, PhD
Faculty of Health Sciences
Department of Medical and Health Sciences
Linköping University
IMH:s kansli
Sandbäcksgatan 7
Linköping, 581 83
Sweden
Phone: 46 737543032
Fax: 46 101031865
Email: armin.spreco@liu.se

Abstract

Background: Influenza is a viral respiratory disease capable of causing epidemics that represent a threat to communities worldwide. The rapidly growing availability of electronic "big data" from diagnostic and prediagnostic sources in health care and public health settings permits advance of a new generation of methods for local detection and prediction of winter influenza seasons and influenza pandemics.

Objective: The aim of this study was to present a method for integrated detection and prediction of influenza virus activity in local settings using electronically available surveillance data and to evaluate its performance by retrospective application on authentic data from a Swedish county.

Methods: An integrated detection and prediction method was formally defined based on a design rationale for influenza detection and prediction methods adapted for local surveillance. The novel method was retrospectively applied on data from the winter influenza season 2008-09 in a Swedish county (population 445,000). Outcome data represented individuals who met a clinical case definition for influenza (based on International Classification of Diseases version 10 [ICD-10] codes) from an electronic health data repository. Information from calls to a telenursing service in the county was used as syndromic data source.

Results: The novel integrated detection and prediction method is based on nonmechanistic statistical models and is designed for integration in local health information systems. The method is divided into separate modules for detection and prediction of local influenza virus activity. The function of the detection module is to alert for an upcoming period of increased load of influenza cases on local health care (using influenza-diagnosis data), whereas the function of the prediction module is to predict the timing of the activity peak (using syndromic data) and its intensity (using influenza-diagnosis data). For detection modeling, exponential regression was used based on the assumption that the beginning of a winter influenza season has an exponential growth of infected individuals. For prediction modeling, linear regression was applied on 7-day periods at the time in order to find the peak timing, whereas a derivate of a normal distribution density function was used to find the peak intensity. We found that the integrated detection and prediction method detected the 2008-09 winter influenza season on its starting day (optimal timeliness 0 days), whereas the predicted peak was estimated to occur 7 days ahead of the factual peak and the predicted peak intensity was estimated to be 26% lower than the factual intensity (6.3 compared with 8.5 influenza-diagnosis cases/100,000).

Conclusions: Our detection and prediction method is one of the first integrated methods specifically designed for local application on influenza data electronically available for surveillance. The performance of the method in a retrospective study indicates that further prospective evaluations of the methods are justified.

(J Med Internet Res 2017;19(6):e211) doi:10.2196/jmir.7101

KEYWORDS

human influenza; algorithms; epidemiological surveillance; public health surveillance; evaluation research; epidemiological methods

Introduction

In light of the rapidly growing availability of “big data” from both diagnostic and prediagnostic (syndromic) sources in health care and public health settings, a new generation of epidemiological and statistical methods is needed for reliable analyses and modeling [1]. This need of new methods adapted to extensive but heterogeneous datasets extends to algorithms for detection and prediction of winter influenza seasons and influenza pandemics. Each year epidemics of influenza occur in communities worldwide and cause extensive morbidity and mortality [2,3]. Preparing for and responding appropriately to winter influenza seasons and pandemics is a critical function of public health. However, a concern regarding forecasting algorithms is that reports of methods used for analyses of extensive datasets originating from different sources do not always meet basic standards. The reports fail with regard to the requirement that evaluators should be able to assess the design and performance of the methods when building the next generation of algorithms [4]. However, regardless of the transparency problems in reporting, the potential of big data analyses in infectious disease control is widely recognized. For instance, autoregressive models for influenza forecasts have shown satisfactory performances when applied on large populations [5]. This implies that the area where the knowledge need presently is most immediate is the detection and prediction of influenza activity at local levels [6]. Such granular views, in turn, can provide input into large-scale models and accurate prediction of influenza spread in wide geographical areas.

Several weaknesses of infectious disease surveillance and prediction systems described in previous decades [7] have still not been addressed in methods design. Responding to this situation, the Centers for Disease Control and Prevention (CDC) in the United States initiated the “predict the influenza season challenge” [8], which encouraged researchers to forecast features of winter influenza season progression that are useful to policy makers and to take advantage of big data resources. For this competition, existing methods for modeling influenza were grouped into three categories:

(1) Compartmental models: These are based on mechanistic assumptions about how the influenza virus is transmitted and use these assumptions to estimate the number of individuals in various states related to a disease [9]. For instance, the Susceptible-Infectious-Recovered (SIR) model approximates the dynamics between groups susceptible to influenza, infected with the virus, and recovered from infection over time [10]. Assumptions typical for this category of models include that any pair of individuals in a defined group are equally likely to interact socially and that genetic subgroups within influenza strains behave identically.

(2) Agent-based models: These are more complex types of mechanistic models that typically use synthetic populations based on census data and build complex schemes of social interaction and disease progress in simulated individuals and communities [11,12]. These models can incorporate more detailed assumptions about transmission dynamics but can be computationally intensive.

(3) Nonmechanistic statistical models: These are phenomenological approaches, that is, they aim to model patterns and trends in the data without necessarily considering the underlying mechanisms. Typical approaches of this type are linear autoregression, which estimate influenza activity using a linear function based on recorded past activity. More complex methods in this category include generalized linear models, Box-Jenkins analysis [13], seasonal autoregressive integrated moving-average models [14], and generalized autoregressive moving- average models [15].

More alternative surveillance methods include, for instance, prediction markets [16] that combine expert predictions using a stock market-like system, and the method of analogues (k nearest neighbors) [17] that makes predictions of future influenza activity levels using similar patterns from the past without assuming a strict model. However, it is disconcerting that none of the above categories of forecasting methods evaluated in the CDC challenge generated satisfying results for all four aspects of influenza epidemics (ie, start week, peak week, peak percentage, and duration) [18].

The aim of this study was to present a novel method for integrated detection and prediction of influenza activity using data electronically available for real-time surveillance in local settings in the Western hemisphere and to evaluate its performance by retrospective application on authentic data from a Swedish county. By local settings in the Western hemisphere is meant communities with specified populations in Europe and North America. Winter influenza seasons and pandemics can be expected to spread to these settings, but the dissemination of the actual types and strains of influenza virus does not likely originate from there. In the presentation of the integrated detection and prediction method, the term epidemic is used as a summary label for both winter influenza seasons and pandemics.

Methods

Study Design

The novel method was formally defined based on a design rationale for integrated detection and prediction methods specifically adapted for application in local influenza surveillance. An overview of the method design is exhibited in the Results section, followed by detailed descriptions of the detection and prediction modules. The results of a retrospective performance study evaluation based on authentic data from a Swedish county are also presented. The study design was approved by the regional research ethics board in Linköping (dnr. 2012/104-31).

Design Rationale

The rationale for design of a novel integrated detection and prediction method is that the aim of local influenza surveillance is early detection and prediction of infected individuals requiring clinical attention, with the purpose of timely allocation of scarce health care resources. Precious time is lost before laboratory data are available for algorithmic processing and test samples are not taken from all patients. Syndromic data are used for peak timing prediction because it is challenging to only use unidimensional gold standard data to predict the peak timing.

Both the detection and prediction functions are to comply with requisite quality and accuracy criteria for technologies to be used in health care and public health practice [19]. A theoretical assumption underpinning the design of the detection module is that the number of influenza cases grows exponentially in the beginning of periods with increased activity. Another assumption is that an alerting threshold can be determined using historical data from previous winter influenza seasons. For prediction of the peak timing, evidence of a strong association between the gold standard and syndromic data sources used for the surveillance is assumed to be available. For prediction of the peak intensity, the peak timing must have been determined and the influenza-diagnosis case rates must follow a bell-shaped function of time around the peak.

Definitions

Influenza detection is defined as indicating the initiation of an epidemic in the community, that is, a prolonged period of elevated incidence rates (exceeding a given limit) of influenza cases, as defined by the rate of individuals clinically diagnosed with influenza in a population under surveillance. Influenza prediction denotes foretelling the peak timing and the peak intensity of an epidemic in the community. For detection, weekday effects and optimal alerting thresholds with reference to influenza-diagnosis data are retrospectively established in the method calibration. For prediction, both the weekday effects and the grouping of variables in the syndromic data with the largest correlation strength and longest lead time to influenza-diagnosis data are established.

The influenza case-rate level when a local influenza epidemic factually takes off was set to 6.3 influenza-diagnosis cases/100,000 during a floating 7-day period. This limit was determined by inspecting the epidemic curves of previous local influenza epidemics in the learning dataset. A similar definition (6.4 influenza-diagnosis cases/week/100,000) was determined for the winter influenza season in 2008-09 in a recent comparison of influenza intensity levels in Europe [20]. The definition of when an epidemic ends was set to the interepidemic (period between two epidemics) influenza-diagnosis level for the specific setting where the method is applied. This was done because the detection algorithm requires that the influenza activity is at an interepidemic level before the algorithm can start its search.

Retrospective Performance Study

For a retrospective performance evaluation of the integrated detection and prediction method, outcome cases were represented by individuals clinically diagnosed with influenza during the 2008-09 winter influenza season in a Swedish county (population 445,000). The thresholds used in epidemic detection were determined using data from a learning dataset containing the 2008-09 winter influenza season. The metrics used to evaluate the detection of influenza epidemics were timeliness, sensitivity, and specificity. Timeliness was defined as the time difference (in days) between the actual start of the epidemic and the start indicated by the model. Specificity was calculated from when the detection algorithm is started (ie, when previous epidemic has come to an end) and until the beginning of the current epidemic per the standard definition (6.3 influenza-diagnosis cases/100,000 during a floating 7-day period). This means that the period length for specificity calculations varies with the interepidemic period. Sensitivity was calculated from the beginning of the current epidemic (according to the same definition) and 45 days into the epidemic. The optimal alerting threshold was decided by calculating sensitivity and specificity and studying them on a receiver operating characteristic (ROC) curve, giving specificity priority over sensitivity because a high level of false alarms is undesirable in public health practice.

To evaluate the prediction of the peak timing, timeliness (defined as time between the predicted day of the influenza-diagnosis peak (highest number of daily cases) and the day of the peak in the observed smoothed series (using moving average of influenza-diagnosis data) was used as metric. To evaluate the prediction of the peak intensity, the absolute and relative differences between the predicted peak intensity expressed as the number of influenza-diagnosis cases at the predicted day of the peak and the observed peak intensity were used as metrics. The reason for not comparing the predicted peak intensity with the actual peak intensity (ie, without smoothing data first) was to reduce the impact from possible outliers.

Data Sources

Influenza cases were identified using the International Classification of Diseases version 10 (ICD-10) codes for influenza (J10.0, J10.1, J10.8, J11.0, J11.1, J11.8) [21] from the local electronic health data repository. For individuals having received an influenza-diagnosis at both primary and secondary levels of care, the diagnosis code recorded at the first contact was used for the analyses. If the codes were recorded at the same day, only the secondary-level diagnosis code was used. Correspondingly, information collected from the calls to a telenursing service in the county was used as syndromic data. Influenza-related telenursing call cases were identified by the chief complaint codes associated with influenza symptoms (dyspnea, fever [child and adult], cough [child and adult], sore throat, lethargy, syncope, dizziness, and headache [child and adult]) from the fixed-field terminology register. In accordance with Swedish legislation (SFS 2008:355), personal identifiers were removed from the records. In this study, only the chief complaints of fever in a child and adult were used because a previous study showed that this combination of complaints was most strongly associated with influenza diagnoses [22].

Results

Method Design Overview

The integrated detection and prediction method is based on nonmechanistic statistical models, that is, patterns and trends in the data are modeled without necessarily considering underlying mechanisms. It is designed for integration in local health information systems. Accordingly, the underpinning structure is defined at four levels, ranging from data sources to performance validation (Figure 1). The method is divided into separate modules for detection and prediction of influenza activity, respectively. The function of the detection module is to alert for an upcoming period of increased load of influenza-diagnosis cases on local health care services, whereas the function of the prediction module is to predict the timing of the activity peak and its intensity. Early detection of increased influenza activity and prediction of peak intensity are based on streams of the gold standard data, whereas prediction of peak timing is based on syndromic data. In this setting, patients clinically diagnosed with influenza were used as gold standard.

An overview of the main statistical assumptions and equations for each component is displayed in Figure 2.

Figure 1. Structure of the integrated detection and prediction method displayed design patterns.

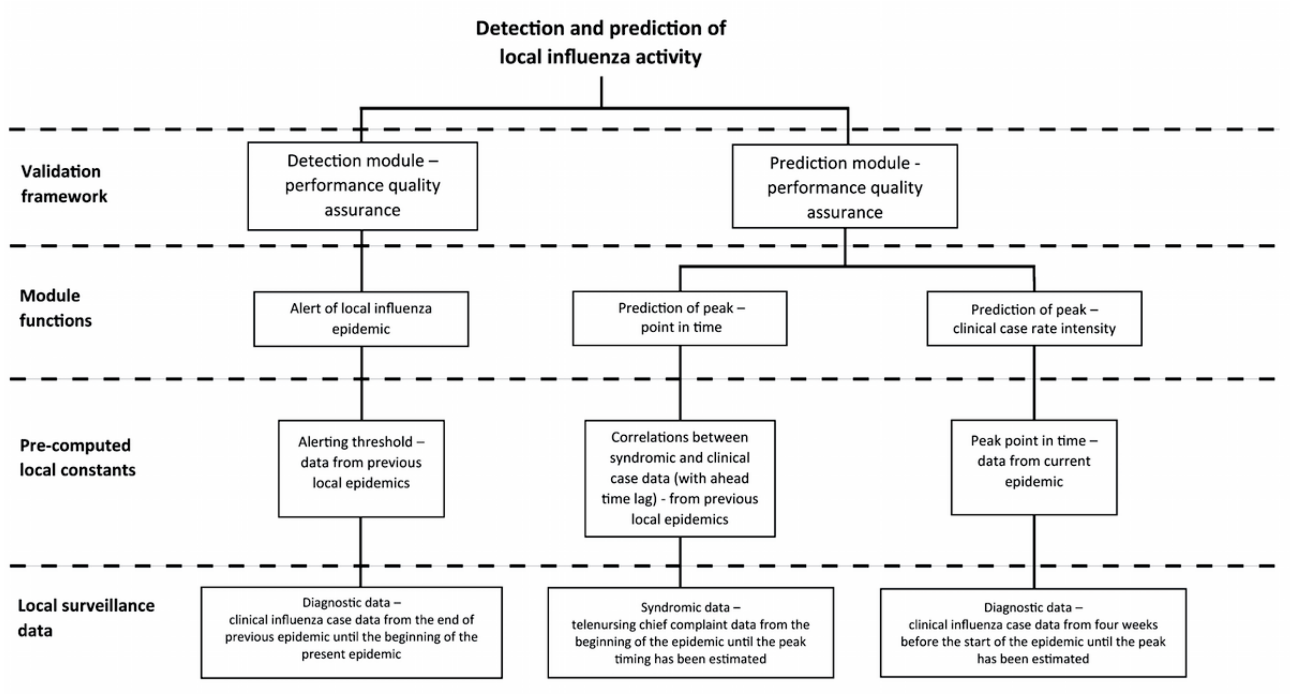
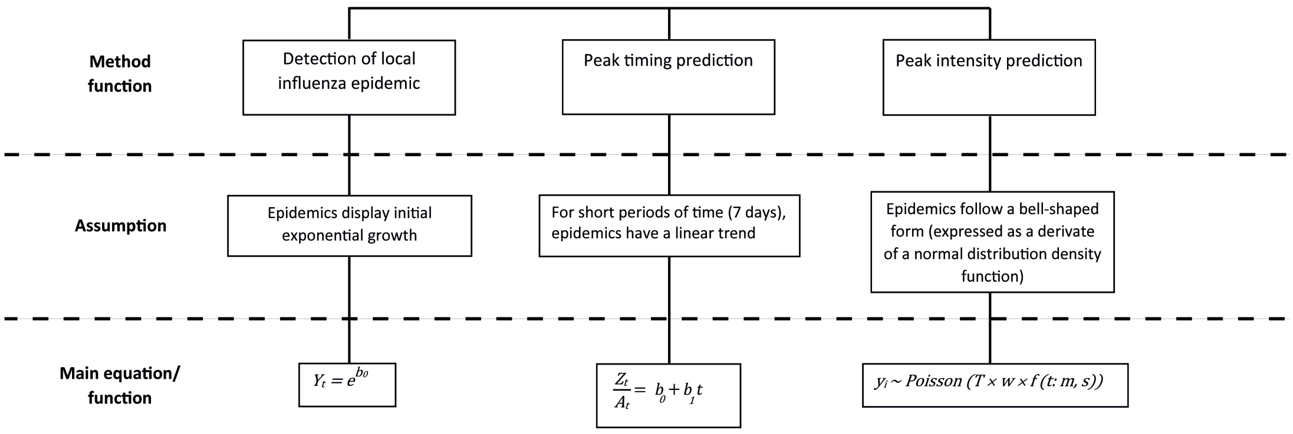


Figure 2. An overview of the main mathematical equations or functions used for each component.



Detection Module

Exponential regression (1) is used for detection modeling, based on the observation that the beginning of an influenza epidemic is assumed to have an exponential growth of infected individuals:

(1) $X_t = e^{a + b_1 t}$

with t representing the time, a_0 representing the level, and b_1 representing the trend. The expected number of visits at local health care services, $E[Y_t]$ is the product of X and the probability p for an infected individual to visit the local health care service. This expectation is also exponential in time:

(2) $E[Y_t] = e^{a + b_1 t} p = e^{a + \ln(p) + b_1 t} = e^{b + b_1 t}$

Where b_0 now combines the current level of number of infected and probability of visiting the local health care service without any possibility to separate them. As daily data are used in the analysis, weekday effects, A_w, are also calculated and used as an offset variable in the exponential regression analysis. The weekday effects are calculated as follows: let A_{Monday} be the average number of events on Mondays during previous epidemics and denote the values for other weekdays by $A_{Tuesday}$, $A_{Wednesday}$, and so on. Let $A_{Total} = (A_{Monday} + \ldots + A_{Sunday})/7$. The multiplicative weekday effect for Mondays is A_{Monday}/A_{Total} and so on. The weekday effects are included in the model:

(3) $E[Y_t] = e^{b + b_1 t + \ln(Aw)}$

If X is large, p is small, and the infected individuals act independently, then Y is approximately Poisson distributed:

(4) $Y_t \sim Poisson\ (e^{b + b_1 t + \ln(Aw)})$

Furthermore, the time is shifted, that is, the most recent day is considered as $t=0$, the second most recent day is considered as $t=-1$, and so on. For every new day, the time axis is moved one step so that the new “most recent day” is considered as $t=0$. For each day an exponential regression analysis (1) is run and a fitted value $\hat{y}$ is calculated by inserting $t=0$ in equation (3) giving

(5) $Y_t = e^b$

as an estimate of the current level of visits which is smoothed for random variation and adjusted for weekday effects. This is repeated for each day by moving the time axis one day at a time so that the most recent point in time of the series is considered $t=0$. Doing this, one value is obtained for every day representing the level for that day. Finally, the lower 95% confidence limit is calculated to represent the level of influenza activity, which is then compared with a predetermined threshold. If the level is above the threshold, an alarm is raised, which means that the epidemic has started; and if the level is below the threshold, no alarm is raised.

Detection starts when the previous epidemic has ended (the interepidemic period level for the community where the detection component is applied), and runs during the inter-epidemic period until an increase in diagnosed influenza cases is detected. When the increase is confirmed, the algorithm is paused and restarted when the epidemic has ended.

The detection algorithm is adjusted in exceptional situations, that is, if an epidemic “simmers” before it begins. The risk of simmering is extensive for a pandemic or an exceptionally mild winter influenza season. In the first case, if there is a fear of a pandemic outbreak among the population, individuals are more likely to contact medical services for influenza symptoms, leading to an increased baseline which increases the risk for false alarms. Also, if a winter influenza season is exceptionally mild, individuals contacting medical services for influenza-like symptoms in the winter will sporadically be misdiagnosed with influenza before the actual circulation of the influenza virus, leading to an increased baseline and thus, an increased risk for false alarms. The alerting threshold determined in the learning set is therefore doubled in these particular cases. It was contended that a strong indication of preepidemic simmering is when it takes extended time between when the influenza incidence increases above a baseline level and when the start of the epidemic occurs (according to the standard definition 6.3 influenza-diagnosis cases/100,000 during a floating 7-day period). The definition for when the influenza incidence has increased above the baseline level is set to 3.2 influenza-diagnosis cases/100,000 during a floating 7-day period (ie, half of the start-of-epidemic definition). An epidemic is then defined to simmer if the time-period separating these 2 dates is longer than three times the average length of the period during previous local influenza epidemics. In other words, the alerting threshold is only doubled due to simmering if the incidence has increased over the baseline level but not exceeded the start-of-epidemic level during this observation period.

Prediction Module

The prediction process is divided into two components. In the first component, syndromic data are used to predict the peak timing, and in the second component, influenza-diagnosis data are used to estimate the peak intensity.

Peak Timing Prediction

In the first component, the aim is to predict the peak timing using linear regression. Including weekday effects A_w and smoothed for random variation, the model for the number of cases in syndromic data is expressed as

(6) $Z_t = (b_0 + b_1 t) \times A_w$,

with b_0 representing the level and b_1 representing the trend. Since the weekday effects A_w are known, a model smoothed for weekday effects and random variation can be expressed as:

(7) $Z_t / A_w = b_0 + b_1 t$

For each 7-day period, a linear regression (7) is run and parameter estimates b_0 and b_1 are fitted. The idea is to estimate the trend in syndromic data for every 7-day period (the first period being days 1-7 and the second being days 2-8), from the beginning of an epidemic and until the peak is found. Although it is unlikely that an epidemic curve increases and decreases linearly, the assumption can be made that the trend during a short period of 7 days has almost a linear increase or decrease.

The search for the peak starts when the detection algorithm signals that an epidemic has taken off and continues until the peak is detected. To identify the peak timing, two conditions are set. As per the first condition, it is essential to ensure that the epidemic has a sufficiently sharp upward trend. The trend is therefore defined as sufficiently sharp when significantly positive ($P<.05$) trends b_1 have occurred either during two consecutive or during three different 7-day periods. When one of these events has occurred, the second condition is applied. According to this condition, when the first significantly negative trend (b_1) during a 7-day period has occurred, it is assumed that the peak has been reached on the first day of this period. However, there is a possibility that this 7-day period “overlaps” with a previous 7-day period, which includes a significantly positive trend. In that case, the first 7-day period with a significantly negative trend is ignored and the peak is instead assumed to appear during the second 7-day period with a significantly negative trend. The search is aborted if the peak is not found when the epidemic has already descended in the local setting where the algorithm is applied.

When the peak is found in the syndromic data, the 14 days preceding influenza-diagnosis data [22] is utilized to find the peak (in influenza-diagnosis data). In other words, if the peak in the syndromic data appears on day 0, the influenza-diagnosis peak is assumed to appear on day 14. However, it is possible that the peak in the syndromic data occurs on a day during the weekend but highly unlikely that the peak in influenza-diagnosis data occurs on one of these days as, for instance, primary care centers are closed during weekends in Sweden. Instead, it is reasonable to assume that the influenza-diagnosis peak occurs at the beginning of the week because individuals who suffer influenza symptoms during the weekend visit primary care centers when they reopen on Monday or possibly Tuesday. Adjustments are therefore made by moving the influenza-diagnosis peak to the following Monday if it is expected to occur on a Friday, Saturday, or Sunday according to syndromic data and to the previous Tuesday if the peak is expected to take place on a Wednesday or Thursday. If the peak is expected to occur on a Monday or Tuesday, no adjustments are made. In other words, in the first case the syndromic data precedes influenza-diagnosis data between 15 and 17 days, in the second case between 12 and 13 days, and in the third case 14 days.

Depending on what day of the week the peak in the syndromic data is expected to take place, the prediction of the influenza-diagnosis peak is made between 6 and 11 days before it is expected to occur, as the syndromic peak can be determined first after 6 days has passed of the syndromic data series.

Peak Intensity Prediction

In the second component of the prediction module, the aim is to predict only the peak intensity. Based on empirical assessments of previous epidemics, an epidemic adjusted for weekday effects is assumed to show a bell-shaped form from the beginning to the end, and can therefore be expressed using a derivate of a normal distribution density function. The intensity function must also include weekday effects and total number of events during the whole epidemic. Use of bell-shaped functions was systematically introduced in epidemiology by Brownlee in the early 20th century [23], and such functions have since then been applied in several contexts, for example, to predict the course of acquired immune deficiency syndrome (AIDS) in the United States [24]. Assuming that the peak timing is known (estimated in the first prediction component) and that an epidemic follows the bell-shaped function around the peak, the intensity function can be used to predict the peak intensity at time m.

Assume that day number $t=1, 2, 3,..., t_i$; the observed number of influenza-diagnosis cases is $y= y_1, y_2, y_3,..., y_i$, and that

$$(8)\ y \sim Poisson\ (T \times w \times f\,(t\!: m, s))$$

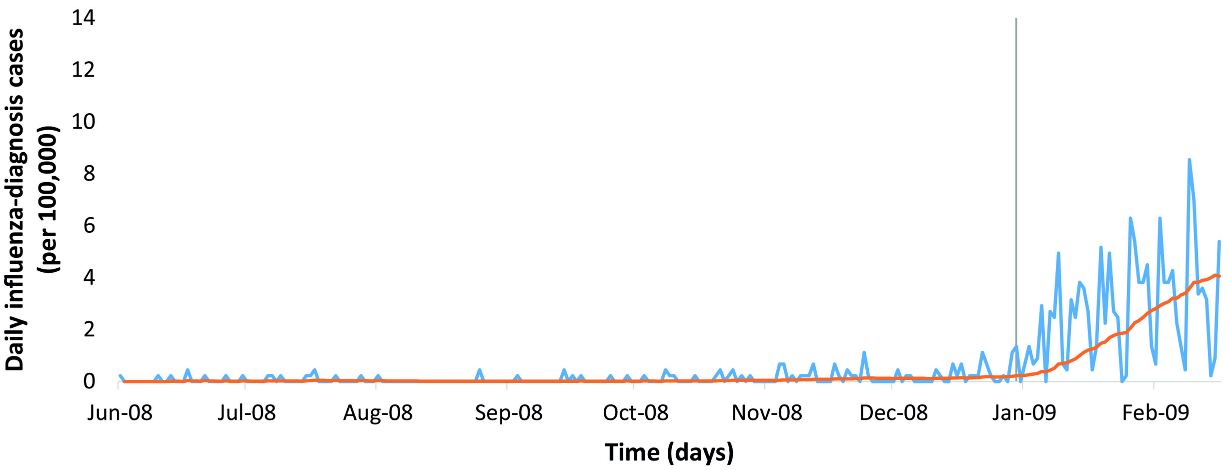
where T is the total number of health care visits of the whole epidemic, w is the weekday effects, f is the normal distribution density function, t is the day number, m is the center of the epidemic (which coincides with t for the peak), and s is the spread in time. Since t, w, and m are known, only the parameters T and s are estimated using y in such way so that the likelihood is maximized. However, in order to do that, first appropriate starting values for these parameters need to be selected. Finally, using the known parameter m and the estimated parameters T and s, the peak intensity at time m is calculated by replacing t with m in equation (8).

It is important that the start of the series seems appropriate because the second prediction component assumes that the level is zero or at an interepidemic level at the start and it is not optimal that there are single or occasional spikes at the beginning of the series. For that reason, the start of the series should be a couple of weeks before an epidemic is detected.

Evaluation of Detection Module

The optimal threshold for the lower confidence limit of the expected number of influenza-diagnosis cases was computed to 0.21/day/100,000 for the detection algorithm. The detection sensitivity and specificity (calculation based on the interepidemic period 211 days) were both 1.00 and the timeliness 0 (Figure 3). This means that the detection module according to the definition (6.3 influenza-diagnosis cases/100,000 during a floating 7-day period) detected the 2008-09 winter influenza season on the day it actually started.

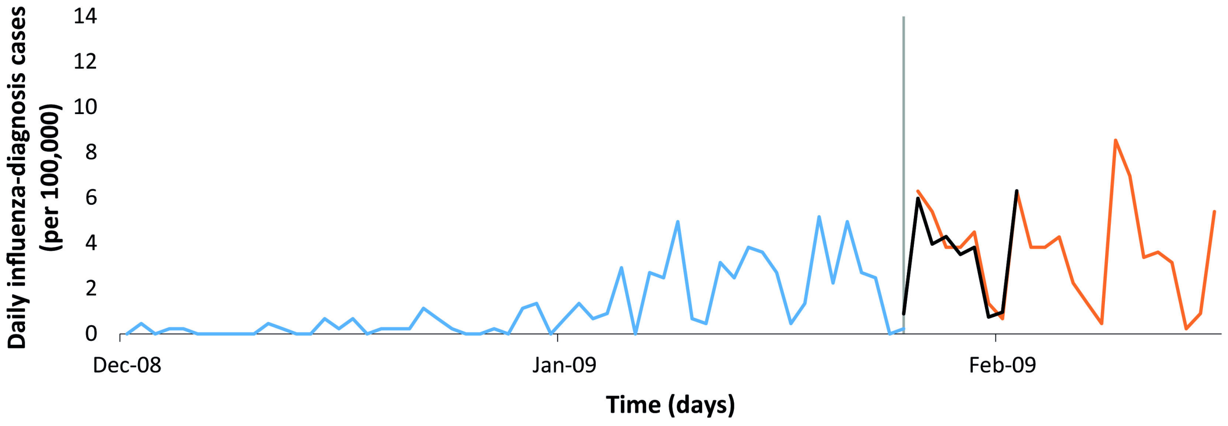
Figure 3. The detection algorithm applied on winter influenza season 2008-09 (A[H3N2]). The blue line represents the number of influenza-diagnosis cases/day/100,000, the gray bar marks the start of the winter influenza season according to the definition (6.3 influenza-diagnosis cases/100,000 during a floating 7-day period), and the orange line denotes the lower limit estimated using the detection algorithm.



Evaluation of Prediction Module

The prediction module performance was satisfying both with regard to the peak timing and peak intensity. The peak timing was estimated 8 days in advance and occurred 7 days before the factual peak occurred. The predicted peak intensity at the predicted day of the peak was estimated to 6.3 influenza-diagnosis cases/100,000 (Figure 4) compared with the factual 8.5 influenza-diagnosis cases/100,000 at the day of the actual peak, that is, the absolute difference between the predicted and the actual incidence was 26%.

Figure 4. The prediction method applied on winter influenza season 2008-09 (A[H3N2]). The blue line represents the number of known actual influenza-diagnosis cases/day/100,000 at the time when the prediction is performed, the orange line represents the number of “unknown” actual influenza-diagnosis cases/day/100,000 at the time when the prediction is performed from the first unknown day and until the peak has passed, the gray bar marks the end of the known and the beginning of the “unknown” actual influenza-diagnosis cases/day/100,000, and the black line denotes the predicted values (using the peak intensity prediction) from the first “unknown” day and until the predicted peak occurs.



Discussion

Principal Findings

The aim of this study was to present an integrated influenza detection and prediction method that uses data electronically available in local public health information systems for real-time surveillance. In the performance evaluation based on retrospective data, the method detected the winter influenza season of 2008-09 on the day it actually occurred, whereas the prediction module showed satisfying performance both with regard to the peak activity timing and its intensity.

Comparison With Prior Work

Many important policy decisions in the response to increased influenza activity are made at the local level, for example, planning of resources at intensive care units and deciding social distancing measures such as school closures. The design of the presented integrated detection and prediction method can be compared with current state-of-the-art big data approaches to influenza forecasting [18]. One such approach for local application is a multi-linear, auto-regressive framework in which information is synthesized from a variety of data sources, ranging from Google Trends to electronic patient records (EPRs) [5,6,25-27]. In this study, clinical influenza-diagnosis cases were used for detection of a period of increased influenza activity and for prediction of the peak intensity, whereas syndromic telenursing data were used for prediction of the peak timing. We chose not to use data directly from EPRs due to integrity and legal issues. In addition to telenursing data, several other syndromic data sources were available that also displayed satisfactory associations with clinical influenza-diagnosis data, for example, Google Trends data, website use data, and local media coverage data [22]. Among the data sources available, telenursing and Google Trends data had the longest lead times and showed the strongest correlations with influenza-diagnosis data. However, since daily Google Trends data could not be guaranteed to be available on a constant routine basis, only telenursing data were used as syndromic data source in the present study. Moreover, we employ full-season learning periods for model updates, whereas continuous updating of the model during an epidemic is used in the multi-linear auto-regressive framework. We chose seasonal updates because a central design prerequisite was to highlight transparency of the method construction and thus facilitate implementation by other researchers. Adjustments of the algorithm processes are only made before pandemics and during winter influenza seasons to adjust detection levels for simmering low influenza activity. We gather that methods for continuous updating have to be dependent on the characteristics of local data and it is challenging to formally define a continuous updating framework in a decontextualized format sufficiently transparent to allow transfer of the framework to other settings with maintained performance levels.

Our approach also differs from the framework with the addition of a detection function. Combined detection and prediction methods are common in weather forecasting but not in infectious disease epidemiology. It is somewhat surprising that this is the case, as there are several studies that have focused on developing either influenza detection or influenza prediction algorithms, but seldom a combination of these [28]. However, one possible reason for this fact can be that researchers or research groups have attempted to develop integrated detection and prediction methods but failed to obtain satisfying results for one or several components and therefore chosen not to publish the findings. We suggest researchers in this field to publish methods even if the obtained results are not satisfying because other researchers may want to further develop and improve these methods.

Public Health Implications

The performance evaluation of the integrated detection and prediction method based on retrospective data showed promising results. The rationale for developing our influenza detection and prediction method was to inform the planning of local response measures and adjustments of health care capacity. During emerging epidemics of infectious diseases, it is vital to have up-to-date information on epidemic trends because hospitals and intensive care units have limited excess capacity [29]. In Sweden, for example, the hospital bed capacity is habitually over-extended already before winter influenza seasons with on average 103 patients occupying 100 regular hospital bed units [30]. It is therefore important that increased influenza activity is noticed early at the local level to make time for adjusting primary care and hospital resources already under pressure to the demand in the community (especially hospitalizations requiring intensive care). Syndromic surveillance methods here serve as complements to traditional surveillance by provision of earlier indications of influenza activity [31,32]. However, although the method showed promising performance, we contend that a retrospective evaluation of a single season is insufficient for drawing valid conclusions about its effectiveness. We find that a retrospective evaluation of numerous seasons still would be insufficient. Instead, prospective evaluations are warranted where historical data are only used to determine thresholds and other parameters, and the method is applied on forthcoming epidemics.

Strengths and Weaknesses

The method presented in this paper has both strengths and weaknesses that need to be taken into regard. An important strength is that the design rationale is documented in detail in order to allow the researchers to consider the arguments for different design decisions when building next generation of integrated detection and prediction methods. Another key strength is that analyses of an epidemic is divided into three separate components (beginning of epidemic, peak timing of epidemic, and peak intensity of epidemic), where statistical and mathematical assumptions for each of these components are made independently of each other. Also, different data sources are applied in each component. Concretely, to detect the beginning of an epidemic, exponential regression is applied on influenza-diagnosis data; to predict the peak timing, simple linear regression is applied on syndromic telenursing data; and to predict the peak intensity, the epidemic is assumed to follow a bell-shaped function of time around the peak and therefore a derivate of the normal distribution density function is applied on influenza-diagnosis data. An approach similar to this has rarely been reported in the field of influenza surveillance. One possible limitation of the method design is that the series of actual influenza-diagnosis data are smoothed and the peak of the smoothed series is used as the actual peak. However, as mentioned in the Methods section, the reason for this design choice was to reduce the risk of misleading influence from outliers.

One potential limitation concerns the use of sensitivity and specificity in the method. These metrics are, however, restricted to assess the accuracy of the alerting threshold. We have previously contended that it is important to determine the appropriate period in time which calculations of sensitivity and specificity are to be based upon [33]. This issue mainly concerns sensitivity because once an epidemic has started, it is known that the daily incidence will exceed the predetermined threshold for a certain period ahead. Expanding this period would generate higher sensitivity and thereby overestimate the method performance. Similarly, if the periods are set too short, the performance of the method may be underestimated. A short period of the specificity can also lead to a situation where hypothetic increases of the incidence level during interepidemic periods are ignored in the calculations, leading to a higher value of the specificity which can also be deceiving. Therefore, we chose to base the calculations of the sensitivity on the first 45 days of an epidemic and the specificity calculations on the period from when the previous epidemic has ended and until the beginning of the current epidemic (ie, for specificity the period length varied).

Another possible limitation concerns the second prediction component, where we chose to apply linear regression on 7-day periods for the search of positive and negative trends in order to find the peak timing in the syndromic data. The length of the period could have been extended with 1-2 days to get more reliable estimates of the trend. However, this alternative was weighted against the risk of predicting the influenza-diagnosis peak with fewer days in advance, and the advantage with earlier prediction of saving these days was preferred. Another limitation is that the prediction of the peak intensity is affected by the peak timing prediction, since a precise prediction of the peak timing increases the chance of an accurate prediction of the peak intensity. Concretely, if the timeliness for the prediction of the peak timing was 0 days instead of 7 days in our retrospective evaluation of the 2008-09 winter influenza season, the predicted peak intensity would have been estimated to 7.7 instead of 6.3 influenza-diagnosis cases/100,000 compared with the factual 8.5 influenza-diagnosis cases/100,000. In other words, the absolute difference between the predicted and the actual incidence would have been 10% instead of 26%. Finally, in the second prediction component, we assumed that an influenza epidemic takes a bell-shaped form from the beginning to the end, and therefore we employed a derivate of a normal distribution density function to find the peak intensity. The same assumption was used by Bregman and Langmuir [24] to predict the course of the AIDS epidemic in the United States but was later shown to be inaccurate [34]. However, in the case of the AIDS epidemic, the bell-shaped function was applied in a setting where the underlying premises radically differed from that at hand in the present study, that is, to predict the course of the AIDS epidemic which had been ongoing for several years in an ill-defined population. In contrast, in our study, the function is used only to find the peak intensity in an increase of influenza activity that lasts for only one season. In this context and for these purposes, we believe that the assumption of a bell-shaped curve is defensible.

Conclusions

During the recent decade, a multitude of algorithms for influenza detection or prediction have been reported [28,35-38]. Unlike meteorology where methods for integrated very-short and long-term predictions have been used in practice settings for several decades (see eg, [39-41]), surprisingly few such approaches have been reported for influenza surveillance. Our integrated detection and prediction method is one of the first designed for application on naturally occurring local influenza epidemics. The results of this study indicate that further prospective evaluations of the method are justified.

Acknowledgments

This study was supported by grants from the Swedish Civil Contingencies Agency (2010-2788) and the Swedish Research Council (2008-5252). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

References

1. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016 Jun 22;353:i3140 [FREE Full text]
2. Cowling BJ, Caini S, Chotpitayasunondh T, Djauzi S, Gatchalian SR, Huang QS, Global Influenza Initiative. Influenza in the Asia-Pacific region: findings and recommendations from the Global Influenza Initiative. Vaccine 2017 Feb 07;35(6):856-864. [doi: 10.1016/j.vaccine.2016.12.064] [Medline: 28081970]
3. Lafond KE, Nair H, Rasooly MH, Valente F, Booy R, Rahman M, Global Respiratory Hospitalizations—Influenza Proportion Positive (GRIPP) Working Group. Global role and burden of Influenza in pediatric respiratory hospitalizations, 1982-2012: a systematic analysis. PLoS Med 2016 Mar 24;13(3):e1001977 [FREE Full text] [doi: 10.1371/journal.pmed.1001977] [Medline: 27011229]
4. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
5. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. Proc Natl Acad Sci U S A 2010 Oct 12;107(41):17486-17490 [FREE Full text] [doi: 10.1073/pnas.1005962107] [Medline: 20876140]
6. Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M. Real-time influenza forecasts during the 2012-2013 season. Nat Commun 2013;4:2837. [doi: 10.1038/ncomms3837] [Medline: 24302074]
7. Laporte RE. How to improve monitoring and forecasting of disease patterns. BMJ 1993 Dec;307(6919):1573-1574 [FREE Full text] [Medline: 8292938]
8. GPO. Announcement of Requirements and Registration for the Predict the Influenza Season Challenge URL: https://www.gpo.gov/fdsys/granule/FR-2013-11-25/2013-28198 [accessed 2016-11-21] [WebCite Cache ID 6mBDujTFF]
9. Hethcote HW. The mathematics of infectious diseases. SIAM Rev 2000;42(4):599-653. [doi: 10.1137/S0036144500371907]

10. Ong JB, Chen MI, Cook AR, Lee HC, Lee VJ, Lin RT, et al. Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore. PLoS One 2010 Apr 14;5(4):e10036 [FREE Full text] [doi: 10.1371/journal.pone.0010036] [Medline: 20418945]
11. Colizza V, Barrat A, Barthelemy M, Valleron AJ, Vespignani A. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med 2007 Jan;4(1):e13 [FREE Full text] [doi: 10.1371/journal.pmed.0040013] [Medline: 17253899]
12. Timpka T, Eriksson H, Gursky EA, Nyce JM, Morin M, Jenvald J, et al. Population-based simulations of influenza pandemics: validity and significance for public health policy. Bull World Health Organ 2009 Apr;87(4):305-311 [FREE Full text] [Medline: 19551239]
13. Sakia RM. The Box-Cox transformation technique: a review. The Statistician 1992;41(2):169-178. [doi: 10.2307/2348250]
14. Shumway RH, Stoffer DS. Time Series Analysis and its Applications: With R Examples. 3rd edition. New York: Springer Science & Business Media; 2011.
15. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
16. Polgreen PM, Nelson FD, Neumann GR. Use of prediction markets to forecast infectious disease activity. Clin Infect Dis 2007 Jan 15;44(2):272-279. [doi: 10.1086/510427] [Medline: 17173231]
17. Viboud C, Boëlle PY, Carrat F, Valleron AJ, Flahault A. Prediction of the spread of influenza epidemics by the method of analogues. Am J Epidemiol 2003 Nov 15;158(10):996-1006. [Medline: 14607808]
18. Biggerstaff M, Alper D, Dredze M, Fox S, Fung IC, Hickmann KS, Influenza Forecasting Contest Working Group. Results from the centers for disease control and prevention's predict the 2013-2014 Influenza Season Challenge. BMC Infect Dis 2016 Jul 22;16:357 [FREE Full text] [doi: 10.1186/s12879-016-1669-x] [Medline: 27449080]
19. Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making--an introduction: report 1 of the ISPOR MCDA emerging good practices task force. Value Health 2016 Jan;19(1):1-13 [FREE Full text] [doi: 10.1016/j.jval.2015.12.003] [Medline: 26797229]
20. Vega T, Lozano JE, Meerhoff T, Snacken R, Beauté J, Jorgensen P, et al. Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method. Influenza Other Respir Viruses 2015 Sep;9(5):234-246 [FREE Full text] [doi: 10.1111/irv.12330] [Medline: 26031655]
21. International statistical classification of diseases and related health problems. 10th Revision. Volume 2. Instruction manual ed. Geneva: World Health Organization (WHO); 2010.
22. Timpka T, Spreco A, Dahlström Ö, Eriksson O, Gursky E, Ekberg J, et al. Performance of eHealth data sources in local influenza surveillance: a 5-year open cohort study. J Med Internet Res 2014 Apr 28;16(4):e116 [FREE Full text] [doi: 10.2196/jmir.3099] [Medline: 24776527]
23. Brownlee J. On the curve of the epidemic. Br Med J 1915 May 08;1(2836):799-800 [FREE Full text] [Medline: 20767626]
24. Bregman DJ, Langmuir AD. Farr's law applied to AIDS projections. JAMA 1990 Mar 16;263(11):1522-1525. [Medline: 2308183]
25. Shaman J, Kandula S. Improved discrimination of influenza forecast accuracy using consecutive predictions. PLoS Curr 2015 Oct 05;7:- [FREE Full text] [doi: 10.1371/currents.outbreaks.8a6a3df285af7ca973fab4b22e10911e] [Medline: 26512336]
26. Yang W, Karspeck A, Shaman J. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. PLoS Comput Biol 2014 Apr 24;10(4):e1003583 [FREE Full text] [doi: 10.1371/journal.pcbi.1003583] [Medline: 24762780]
27. Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics. Proc Natl Acad Sci U S A 2015 Mar 03;112(9):2723-2728 [FREE Full text] [doi: 10.1073/pnas.1415012112] [Medline: 25730851]
28. Spreco A, Timpka T. Algorithms for detecting and predicting influenza outbreaks: metanarrative review of prospective evaluations. BMJ Open 2016 May 06;6(5):e010683 [FREE Full text] [doi: 10.1136/bmjopen-2015-010683] [Medline: 27154479]
29. Donker T, van Boven M, van Ballegooijen WM, Van't Klooster TM, Wielders CC, Wallinga J. Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in the Netherlands. Eur J Epidemiol 2011 Mar;26(3):195-201 [FREE Full text] [doi: 10.1007/s10654-011-9566-5] [Medline: 21416274]
30. Ingen på sjukhus i onödan. [No one in hospital unnecessarily.]. Stockholm: Sveriges Kommuner och Landsting (SKL); 2016:1-61.
31. Friesema IH, Koppeschaar CE, Donker GA, Dijkstra F, van Noort SP, Smallenburg R, et al. Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. Vaccine 2009 Oct 23;27(45):6353-6357. [doi: 10.1016/j.vaccine.2009.05.042] [Medline: 19840672]
32. Ma T, Englund H, Bjelkmar P, Wallensten A, Hulth A. Syndromic surveillance of influenza activity in Sweden: an evaluation of three tools. Epidemiol Infect 2015 Aug;143(11):2390-2398 [FREE Full text] [doi: 10.1017/S0950268814003240] [Medline: 25471689]

33. Spreco A, Eriksson O, Dahlström Ö, Timpka T. Influenza detection and prediction algorithms: comparative accuracy trial in Östergötland county, Sweden, 2008-2012. Epidemiol Infect 2017 May 17:1-10. [doi: 10.1017/S0950268817001005] [Medline: 28511741]
34. Nishiura H. Lessons from previous predictions of HIV/AIDS in the United States and Japan: epidemiologic models and policy formulation. Epidemiol Perspect Innov 2007 Jun 13;4:3 [FREE Full text] [doi: 10.1186/1742-5573-4-3] [Medline: 17567907]
35. Buckeridge DL. Outbreak detection through automated surveillance: a review of the determinants of detection. J Biomed Inform 2007 Aug;40(4):370-379 [FREE Full text] [doi: 10.1016/j.jbi.2006.09.003] [Medline: 17095301]
36. Chretien JP, George D, Shaman J, Chitale RA, McKenzie FE. Influenza forecasting in human populations: a scoping review. PLoS One 2014 Apr 08;9(4):e94130 [FREE Full text] [doi: 10.1371/journal.pone.0094130] [Medline: 24714027]
37. Hiller KM, Stoneking L, Min A, Rhodes SM. Syndromic surveillance for influenza in the emergency department-a systematic review. PLoS One 2013 Sep 13;8(9):e73832 [FREE Full text] [doi: 10.1371/journal.pone.0073832] [Medline: 24058494]
38. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir Viruses 2014 May;8(3):309-316 [FREE Full text] [doi: 10.1111/irv.12226] [Medline: 24373466]
39. Browning KA, Collier CG. Nowcasting of precipitation systems. Rev Geophys 1989 Aug;27(3):345-370. [doi: 10.1029/RG027i003p00345]
40. Dixon M, Wiener G. TITAN: thunderstorm identification, tracking, analysis, and nowcasting - a radar-based methodology. JTECH 1993;10(6):785-797.
41. Golding BW. Nimrod: A system for generating automated very short range forecasts. Met. Apps 1998;5:1-16. [doi: 10.1017/S1350482798000577]

Abbreviations

AIDS: acquired immune deficiency syndrome
CDC: Centers for Disease Control and Prevention
EPRs: electronic patient records
ICD-10: International Classification of Diseases version 10
ROC: receiver operating characteristic
SIR: Susceptible-Infectious-Recovered

Edited by A Keepanasseril; submitted 05.12.16; peer-reviewed by S Pesälä, E Lau; comments to author 29.01.17; revised version received 15.03.17; accepted 19.04.17; published 15.06.17.

Please cite as:
Spreco A, Eriksson O, Dahlström Ö, Cowling BJ, Timpka T
Integrated Detection and Prediction of Influenza Activity for Real-Time Surveillance: Algorithm Design
J Med Internet Res 2017;19(6):e211
URL: http://www.jmir.org/2017/6/e211/
doi:10.2196/jmir.7101
PMID:28619700

Original Paper

“You Sort of Go Down a Rabbit Hole...You’re Just Going to Keep on Searching”: A Qualitative Study of Searching Online for Pregnancy-Related Information During Pregnancy

Julie Prescott[1], PhD; Lynn Mackie[1], PhD

[1] Education and Psychology, University of Bolton, Bolton, United Kingdom

Corresponding Author:
Julie Prescott, PhD
Education and Psychology
University of Bolton
Deane Road
Bolton, BL3 5AB
United Kingdom
Phone: 44 0120490 ext 3676
Fax: 44 01204903646
Email: j.prescott@bolton.ac.uk

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2017/7/e223/

Abstract

Background: The Web is becoming increasingly popular for gaining information on medical or health issues; with women in particular likely to search online for this type of information and support. Despite the increased use of the Web for health-related information, we need to question whether the Web and the ease of seeking health information that it provides leads to more (patient) empowerment. As well as being a time of joy and expectations, pregnancy can be a worrying time for women, especially first time mums-to-be, with unfamiliar experiences and symptoms and concerns for the baby as well as the self.

Objective: Our aim was to explore how and why pregnant women use the Web to gain information and support during pregnancy and what they consider a reliable source.

Methods: To meet the objectives of the study, a qualitative approach was required to gather information on the experiences of currently pregnant women who use the Web to gain information and support during their pregnancy. Sixteen pregnant women took part in a semistructured interview, either face-to-face or via telephone. The interviews took place from January to March 2016, all participants were from England, and the health professionals are all employed by the National Health Service (NHS). Qualitative analytical procedures were employed using inductive thematic analysis supported by NVivo software (QSR International).

Results: Pregnant women found reassurance from the experiences of others. This reassurance resulted in them feeling less alone, as well as enabling them to normalize any symptoms or experiences they were undergoing. The women understood that caution was needed at times while reading the stories of others, acknowledging the potential for extreme cases or worst case scenarios. This is particularly pertinent to the Web, as this wide range of stories may not be as easily accessible if stories where confined to those in a woman’s offline social circle. The interviews provide insights into how and why pregnant women search online for information and perhaps more so, support while pregnant.

Conclusions: Searching for health information and advice online during pregnancy is viewed as quick, easy, and accessible. The affordances of the Web have provided women the opportunity to go online as a first port of call. Knowing they were not alone and reading the experiences or symptoms of other pregnant women enabled women to normalize their experience and was ultimately reassuring for pregnant women.

(J Med Internet Res 2017;19(6):e194) doi:10.2196/jmir.6302

KEYWORDS

pregnancy; information seeking behavior; qualitative research

Introduction

The Web is becoming increasingly popular for gaining information on medical or health issues [1]. Indeed, according to a Pew Internet report, 80% of American Web users look for health information online [2], and 35% of American adults have gone online to diagnose a medical condition, referred to as "online diagnosers" [1]. Women are more likely to go online to figure out a possible medical diagnosis than men [1] and search online for health information [3]. However, despite the increased use of the Web for health-related information, we need to question whether the Web and the ease of seeking health information that it provides leads to more (patient) empowerment. Women are more likely to search for health information online [3] for themselves and other family members, and as previously mentioned they are also more likely to go online to try and diagnose a medical condition [1]. A large Australian study [4] found that women who experienced stigmatized conditions such as sexually transmitted infections or mental health were more likely to search online for health information [4]. Women also use the Web to seek information regarding pregnancy and childbirth in particular, as these are often worrying times for women [3,5]. Research suggests that health anxiety is not elevated during pregnancy, although health anxiety is higher for those women who have experienced complications during pregnancy [6].

As women utilize the Web for health information more so than men, and pregnancy and childbirth can be worrying times for women, it was deemed important to consider how women use the Web during pregnancy for pregnancy-related information and how they negotiate and navigate through the mass of data available to them online. The vast amount of information and personal stories available and easily accessible online suggest that women must negotiate what information they take on board and how they react to the "horror stories" due to numerous risks involved in pregnancy and childbirth [7].

Previous research has found that around 80% of Web users seek health information online [8,2] for reasons such as convenience, anonymity, specialized advice, and social support when feeling stressed or worried [9]. Previous research has focused on parents who seek health information online for their children [10]. Bernhardt and Felter [10] found that women as parents are among the highest information-seekers who use the Web to confirm beliefs or get a "second opinion." They also found that online sources need to be clear with advice in order to gain more trust from mothers, and more trusted advice comes from clinical professionals or parents (in online blogs). The parents in the study briefly discussed their pregnancy health information seeking behavior, stating that they used the Web for support and health advice [10].

In a more recent study, Lagan et al [5] aimed to understand how and why pregnant women used the Web for health information and how this impacted on their decision making during pregnancy. The findings revealed that women googled information and used the information found online to supplement offline information they had received from health professionals, and the Web played a significant part in health information-seeking and decision-making during pregnancy. From the findings, the authors were left with questions with regard to the knowledge of women to evaluate the quality of the information sourced. In a 2011 study [11], Lagan et al investigated the perspectives of midwives toward the online searching behavior of pregnant women, finding a general increase in the use of the Web among pregnant women, with midwives concerned about the accuracy of information found, corroborating the authors earlier findings.

It is viewed as important that people know where to seek information online; having this knowledge could empower people and health care providers should suggest suitable sites for pregnancy-related information [12]. This is a view shared by Lagan et al [5]. In their study of online pregnancy-related information for nutrition and physical exercise, Huberty et al [12] found that the information women sourced via the Web during their pregnancy increased their confidence in decision-making during pregnancy. However, it was acknowledged that there is a need to question the trustworthiness of information found online compared to offline sources such as doctors or family and friends. Research has found that information sought and found online is secondary in both priority and reliability toward offline sources [5]. People in the digital age want to be informed patients and online health information searching is not due to any dissatisfaction with doctor's care (or any other health care professional [HCP]) or information provided [5,13].

Supporting this view, Song et al [7] found that although women use the Web, they still rely on their doctors. Women utilized the Web to confirm normalcy and help them take control. They used the online information to confirm knowledge received from health professionals and to seek reassurance of what is normal, with a general desire not to be the only one experiencing symptoms; this was found to be especially pertinent for first time mothers. The study found that pregnant women were interested in keeping track of fetal development [7], supporting early research that fetal development is the most cited online search topic of interest for pregnant women [10].

The overall aim of the larger study was to gain an understanding of how pregnant women use the Web for health-related information during pregnancy, as well as to gain an understanding of how women can skillfully surf in order to reduce any potential anxiety caused through seeking health information online. This paper focuses on how and why pregnant women use the Web to gain information and support during pregnancy and what they consider a reliable source.

Methods

The larger study from which this paper is drawn took a mixed-methods approach, using both qualitative semistructured interviews with pregnant women and a Web-based questionnaire for currently pregnant women and new parents, both male and female, in order to understand their online health information-seeking behavior. The interviews and questionnaire elements of the study ran simultaneously in order for a larger, more diverse dataset from which to derive results. This paper focuses on the findings of the qualitative element of the study.

Sampling and Participants

Using a snow-ball sampling technique, a total of 16 semistructured interviews were conducted with pregnant women. n order to recruit pregnant women to the study, the study was advertised around the university, and participants were also asking for referrals of other currently pregnant women. Although small, the sample size was deemed adequate due to data saturation [14]. Data saturation was reached when there was enough information to replicate the study [15,16]; new information was not obtained [17] and further coding was not feasible [17], and future interviews would not yield further information or coding categories [17].

All interviews were conducted by the same interviewer. The interviewer was trained in qualitative research. The interviews ranged in duration from 20-60 min and were conducted one-to-one either by phone or in person from January to March 2016. Of the 16 women interviewed, only 14 provided demographic data (see Table 1 for information on participants). Participants received and were asked to read a participant information sheet informing them about the study and what was required of them. All participants were required to sign a consent form before the interviews took place. The consent form informed participants that the interview was voluntary and that they were free to stop the interview at any time during the interview, without giving a reason as to why. All participants were also required to complete a demographics form asking information about age and ethnicity. The form also sought information related to the women's current pregnancy, such as trimester, any pregnancy related and nonpregnancy related medical conditions, and use of the Web and support networks.

The interview schedule consisted of 4 sections. Section 1: online information seeking, with 7 questions (ie, what online sources do you use to find out information about your pregnancy?). Section 2: how you search, with 5 questions (ie, how do you know what to trust online?). Section 3: impact of online searching, with 9 questions (ie, what makes online information reassuring? How long does that last? What makes online information worrying? How long does that last?). Section 4: support and coping, 6 questions (ie, do you ever access any online health care professionals?).

Ethical and Research Approvals

Approval for the study was given by University of Bolton Research Ethics Committee in October 2015, United Kingdom.

Data Analysis

All interviews were digitally recorded and transcribed verbatim. All qualitative data were analyzed using Nvivo version 10 software. The qualitative approach used to analyze the interviews was an inductive thematic analysis [18, 19] since the analysis was grounded in the data rather than existing theories. The data was analyzed using Braun and Clarkes approach [18] for using thematic analysis in psychology. In order to ascertain and increase intercoder reliability and the reliability of the results, 2 transcripts where independently coded by both authors to develop a coding framework and code book. Any subsequent additional theming was discussed during independent analysis. The stories from other women who had experienced the same or similar experiences helped pregnant women to feel reassured and less like they were the only ones going through something.

Results

Overview of Results

In general, pregnant women tended to find the Web a useful source of information and advice during their pregnancy. Although there were things that worried them when they searched on the Web, overall the accessibility of Web-based information and the stories and experiences of other women were viewed as reassuring. Four main themes emerged from the interviews. Table 2 shows the main themes, topics, and an example quote from the interview analysis.

Online Forums: Reading the Stories and Experiences of Other Pregnant Women

From the analysis, it was evident that the participants tended to search and engage with online forums to read and get advice from other pregnant women. All the women talked about other women's stories they read on the Web and these were viewed positively and seen as providing pregnant women with reassurance.

Reassurance, Not Alone, and Normalized

Other women's stories and experiences had a positive impact in terms of reassurance and helping to reduce worry as shown in Table 2. The stories from other women who had experienced the same or similar experiences helped pregnant women to feel reassured and less like they were the only ones going through something alone. Not feeling as if they were the only one's experiencing something also normalized symptoms and issues for women as highlighted in the quotes in Table 2. Interestingly, the following quote shows how some women found reading other women's experiences as supportive and which enabled them to ask questions they did not want to ask their midwife:

> *If I had to ask the midwife like everything I was worried about, I wouldn't worry her that much and ask her that. And I think also the fact that there's so many people who are talking about their own experience, so it can be really supportive so it's like if you're worried about your next scan or whatever it is then you can do a post saying I'm worried has anyone else been, and then you've got lots of people who kind of supporting you.* [Participant 14, multiparous]

Table 1. Demographics or characteristics of participants.

Demographics		n
Age in years		
	16-20	1
	21-25	6
	26-36	2
	31-35	3
	>36	2
Pregnancy status		
	Primiparous	6
	Multiparous	8
Trimester at the time of interview		
	1	0
	2	7
	3	7
Medical complications in this pregnancy (including pre-eclampsia, depression, diabetes, and sickness)		
	Yes	7
	No	7
Medical complications in a previous pregnancy		
	Yes	4
	No	10
Web use		
	Daily	12
	A few times a week	2
Support during pregnancy		
	Midwife	14
	General practitioner	7
	Family	14
	Friends	12

Forums were also viewed by women as a place to gain information and keep up to date:

> *Like you know again if you think they might be a bit of new information out or, another person might have written online that they've got the same illness that day and you know you want to just read up about them and what they're going through and how they're feeling so.* [Participant 1, primiparous]

Others viewed searching online as a specific place for seeking reassurance rather than information as expressed in the following quote:

> *I always go on for reassurance really rather than getting the information. That's what I use it for.* [Participant 9, multiparous]

Table 2. Themes and topics.

Theme	Topics	Example quote
Sharing experiences	Reassurance	*I'd say they reassure me, because you know I'll search for if I'm feeling down and you'll see other women that feel the same as you.* [Participant 1, primiparous]
	Not alone	*Just knowing somebody else has been through the same kind of thing as you and you're not on your own, knowing that you're not the only person going through a certain thing.* [Participant 4, primiparous]
	Normalize	*I mean there's a lot of stuff on there, on Netmums...and when I went on Netmums there is a lot of women out there who say it's you know—perfectly normal, not to worry about it. And then I did check it with my midwife and it is you know just part of the babies growing development.* [Participant 16]
	Reliability of information	*I tend to not rely on the information too much because there are you know anybody can put anything on the internet and it's all sorts of blogs and it's err, so yeah I do like to read opinions of other people but I don't, I try not to depend on them.* [Participant 5, primiparous]
	Worst case scenarios	*The fact that say for example on the forum you're not going to get people going on saying I'm having a perfectly healthy pregnancy, you're going to have the people that have had problems. So the fact it's almost going to kind of over-represent, it's always going to over-represent the people who've had issues.* [Participant 14, multiiparous]
Affordance of the Web	Quick and accessible	*Well it's immediate so you're not concerned until you receive or get a doctor's appointment, so you can pretty much get an answer straight away can't you, I think it's a really useful source* [Participant 2, multiparous]
	Not a substitute for midwife	*Online used for (short term) reassurance but offline for definitive answers, so I would trust kind of forums like Mumsnet and Netmums and things if it was just about how people felt about something. But if I actually wanted an answer to something I wouldn't trust them. I'd trust it if it was a website where, someone would like BabyCenter I think it's called, where someone's asked a question and they've had approved people answering it. Whereas if I'm just kind of looking for reassurance like is there anyone else who was on anti-depressants and their baby was fine, then I'm more reassured by kind of anecdotal accounts for that.* [Participant 14, multiiparous]
	Good source of information and advice	*I generally actually tend to go online and like find out information first for myself before I go, go and speak about it with other people. I don't, because that's, that's sort of like my nature that I feel like I bother people if I ask them something so I try not to ask questions I try to find things for myself.* [Participant 5, primiparous]
Pregnancy-specific focus	Search on the Web now pregnant	*I've looked more and researched more about pregnancy since I've been pregnant but beforehand I wouldn't say I used the website if I was poorly, if I was poorly I just used to go to the doctors. I wouldn't try and erm, you know, think of what condition I've got, I'd just go straight to the doctors.* [Participant 8]
	Baby development	*Online like I said I just use my app, I sometimes use google but it's more just the app to be honest with you, just having a nosey at the changes to the baby and things like that.* [Participant 4, multiparous]
	Reliable sources	*I usually go refer back to the NHS pages as well.* [Participant 9, multiparous]
When to stop	Information not helping	*Just when you feel like you now well this is not really helping me anymore so just shut it down.* [Participant 5, primiparous]
	Saturation	*I think the more that the same thing that comes up then I trust it, so the more it occurs.* [Participant 9, multiparous]
	Reassured	*I just think when you feel like you've got an answer.* [Participant 2, multiparous]

Questions of Reliability

Despite the benefits gained from reading other women's pregnancy stories and experiences, there was the recognition by the participants that there were issues with searching and reading online and caution should be taken as expressed in the following quote on the unreliability of information sourced online. This caution is due to the Web s' often contradictory nature, and the fact that information is based on peoples, often limited experiences and are opinion based, rather than always fact based.

> *Contradicts itself a lot, some people say "oh I was bleeding, I didn't have a miscarriage," "I was bleeding, I did have a miscarriage." So you never know what to believe. So then if a load of people have said "oh I've had a miscarriage these pains the same as you, I had a miscarriage," you're going to think you're going to have one even though you might not. So that can really worry some people.* [Participant 10, primiparous]

Worst Case Scenarios

There is a suggestion that stories on the Web are biased toward the negative view point and there was a strong acknowledgment throughout the interviews that many of the stories people posted online were often "horror stories" or worst case scenarios and viewed as something one would not necessarily have access to if it was not for the Web. This perhaps highlights that although easily accessible, the Web is not constrained by geographic boundaries.

> *Because you can see obviously worst case scenario and people that might then think that it's worse than it actually is. You know, and it isn't always reliable is it, so and I'm not a chatroom person at all but sometimes you just referred to these aren't you because that specific symptom you mentioned. And I think the people that generally write on these chatrooms are the people that have had the bad thing happen to them, you know the worst case scenario. You're rarely going to hear off somebody that actually it turned out fine.* [Participant 2, multiparous]

These worse case scenarios or extreme cases would often cause concern for participants and increase worry in pregnant women as acknowledged clearly by participant 1:

> *I suppose when you read things that you know aren't particularly nice, you know again with my pre-eclampsia you know and then if I read, you know a lady passed away from it or you know someone's had life-long effects from it and stuff like that then you obviously, then I'm going to worry...and think you know well that could happen to me or you know well what if it does when really the chances of that are quite slim, but I suppose it is just the things you read and the things that has happened to people.* [Participant 1, primiparous]

Online Affordances

Quick and Accessible

The affordances of the Web were viewed as important especially in terms of the accessibility and the speed with which one can get answers to concerns or worries online. This was especially important when it may be difficult to access offline health professionals. A number of the pregnant women in this study stated that they accessed the Web daily for nonwork issues, suggesting they are regular Web users. Supporting the literature that the Web is becoming increasingly popular for health information [1], participant 15 expresses this point: "When you're looking for something you just search online." The women tended to google symptoms as their method of searching symptoms, as research has also previously found [7], with the same participant stating: "If I've got pains or anything like that I'll just Google it. Yeah so Google's a brilliant thing."

Online Information Not Viewed as a Substitute to Offline Health Professional (Midwife)

In terms of information online, participants recognized that the information received online should not be trusted and that you needed to also go offline. Searching for information online was not a substitute for seeking offline support or advice, more an accessible, easy first port of call as highlighted by the quote in Table 2.

Good Source of Information and Advice

Despite the acknowledgment of caution required when seeking information online, the Web was viewed as a good place to get advice and perhaps gain a feeling or sense of control with their pregnancy. In particular, it was viewed by participants as a good place to get information before seeking help from a health professional, again linking into the accessibility of online information.

In many cases it was also down to the fact that they did not want to bother health professionals, and women felt comfortable gaining reassurance or being more informed before they contacted a health professional such as a midwife.

> *Yeah, well, err I think, I generally actually tend to go online and like find out information first for myself before I go, go and speak about it with other people. I don't, because that's, that's sort of like my nature that I feel like I bother people if I ask them something so I try not to ask questions I try to find things for myself.* [Participant 5, primiparous]

Interestingly, this participant felt the information sought online encouraged her to seek help offline:

> *Yeah, erm like the preeclampsia one last week. That scared me. Just because it's, I'm only 28 weeks now, I was 27 last week. And to have a baby at 27 weeks I was like I'm not ready, like there's no way this baby is ready to come out so I really hope it's not that. Erm, so that probably worried me, so like the night before, and then I was at the doctors the next day. But I think if I hadn't of read that I wouldn't have been at the doctors the next day. I wouldn't have been so worried, if that makes sense? Like I would have*

> *been ignorant in a way.) The fact that if like it is scary, so like with the, I hadn't got a clue about preeclampsia, didn't know what it was and then soon as I put my symptoms up and it's in your face, you can't ignore it. It's like wow, it's there, and you're just like warning signs go off. But like if I hadn't had that information I wouldn't probably have been as worried, I wouldn't have known what it was. Or have to know to go to the doctors. So it's good and bad I suppose.* [Participant 3, primiparous]

Pregnancy-Specific Focus

Online Information Seeking Different Now Pregnant

Some participants had indeed found that their online health seeking behavior had changed, in terms of searching more, since becoming pregnant. Perhaps, indicating that searching online for health and pregnancy information does indeed increase, or at least change, during pregnancy.

Baby Development

Many women enjoyed the website and mobile apps that informed them of their babies' development and symptoms associated with pregnancy development.

Reliable Sources

The participants in this study had specific sites and apps they viewed as reliable or official sources in which to seek pregnancy-related information from. The most popular website for pregnancy-related information was the National Health Service (NHS), with 2 women mentioning the NHS Start4Life in particular. Table 3 lists the sites the women used and felt where reliable sources of online information. Only 1 participant mentioned being recommended a site (Association for Improvements in the Maternity Services [AIMS]) by a midwife. Unsurprisingly the official NHS website was mentioned most reliable as it is the most widely known and reflects the current National Institute for Health and Care Excellence (NICE) guidance on treatment.

Table 3. Websites and apps used by the participants.

Website or app	n
NHS[a] (NHS Start4Life)	16 (2)
The bounty app	8
Net mums	7
Mums net	4
Baby centre	4
Emma's diary	1
AIMS[b]	1

[a] NHS: National Health Service.

[b] AIMS: Association for Improvements in the Maternity Services.

Knowing When to Stop

In regard to any insights into how women skillfully surf, it was vital to gain an understanding and some insight into how pregnant women knew when to stop searching online for pregnancy-related health information. The most popular themes were (1) the information is no longer helpful or helping them, (2) saturation, and when the participant felt they were reading the same information and searches were no longer producing new information, but the same information was being repeated and (3)reassuring, in that they feel they had sought and gained an appropriate answer and they felt reassured. Other participants in the study indicated that they often found it difficult knowing when to stop:

> *It depends how worried I am about something. So I am a bit of a worrier like I do have anxiety so I will like sometimes go through a few pages and read everything. I'd probably carry on searching and sometimes make it worse. Because the more you search the further you go down the search results the more unclear and the more worrying the information gets.* [Participant 14, multiparous]

Whereas others found it much easier to get the information they required and could stop searching:

> *I think sometimes it's just about reassurance sometimes, reading it and thinking right okay, I'll be satisfied with that. And then that will be done until something else.* [Participant 7, multiparous]

Discussion

Principal Findings

Knowing when to stop searching was different for different women. The Web was considered an extremely quick and accessible source, compared with offline official sources used for support and reassurance from those going through the same experience—normalizing and used for self-triage, possibly prior to seeking medical advice. The information sought online was not a replacement for a midwife or doctor's advice, supporting previous findings [5,7]. In general, pregnant women tended to find the Web a useful source of information and advice during their pregnancy. Although there were things that worried them when they searched online, overall, the accessibility of online information and the stories and experiences of other women were viewed as reassuring. This reassurance was provided through both practical information and advice as well as through emotional support; both equally reassuring for the participants in this study.

The findings from this research indicate that searching online does indeed increase during pregnancy for women. However, this online searching was not necessarily a bad thing. It was clear that the women gained reassurance from hearing the stories and experiences of other women, either to gain information and knowledge or to gain a “second opinion.” This supports Bernhardt and Felter’s research on (female) parents’ health information seeking for their children [10]; allowing women to feel more in control and informed during their pregnancy [13]. This study does give an indication that women liked to have some form of control in their pregnancy. This is shown through women wanting to seek out information on their own, perhaps enabling them to feel empowered and informed with some level of control. This may be especially pertinent for women with complications or who have had complications in their pregnancy.

Not feeling as if they were the only ones experiencing something also normalized symptoms and issues for women, supporting previous research [7]. However, it was evident that this online information was not a substitute or a replacement for offline information [5,7], confirming the view [5] that information sought online is secondary in terms of both priority and reliability toward offline sources. However, in this study, women sought information online before going offline. In one particular instance, searching online convinced one woman that she did indeed need to seek offline care. In particular, women enjoyed the websites and mobile apps that informed them of their babies’ development and symptoms associated with pregnancy development, again supporting previous findings [7,10]. The Web, in particular the online forums for pregnant women, were viewed as good place to communicate and connect with others, providing women with a feeling of support as well as enabling women to develop networks with others going through pregnancy at the same time. This is valuable as women may not have this network of support offline.

However, the convenience of the Web and the fact that “Google it” is a well-known and established term of phrase, highlights that the Web is, for most people, a first port of call. This was found to be the case for pregnant women, who for one reason or other would turn to the Web first for answers or reassurance before seeking advice and reassurance offline. Numerous reasons were provided from the interviews as to why it was the first port of call aside from it being quick and accessible, namely, not wanting to bother their midwife or another HCP, wanting to be more informed about an issue or concern before going to see their midwife, not wanting to ask what may be deemed as silly questions when with their midwife, not wanting to waste their midwife’s time with a list of questions, not being able to contact or see a midwife and needing more speedy reassurance, feeling uncomfortable asking health professionals questions, or forgetting to ask certain questions when with the midwife. Online information provided short-term reassurance whereas information sought offline through a midwife or HCP provided long-term reassurance. Online information provided by a professional did also provide long-term reassurance but it is unclear from this research how the women knew if any claims of information coming from a HCP were indeed genuine.

Not surprisingly the NHS was the most reliable source of information for participants. Only one pregnant women in this study had been recommended another website by a HCP (in this case their midwife). It would be beneficial for health professionals to provide pregnant women with information on reliable online sources and highlight them to the potential dangers of forums. HCP should suggest suitable sites to empower people [12], especially since previous research highlighted this as a concern for midwives that more and more pregnant women are searching information online [13]. This and prior research [5] note the dangers of seeking health information online due to the element of bias toward having access to “worst case scenarios.” Interestingly, participants in the study did acknowledge that the worst case scenarios are an issue online and that this is an affordance of the Web, allowing access to a wider network of people who you would not have access to, and hear stories from, offline. This acknowledgement and insight suggests a good understanding of the Web and how to skillfully surf.

In some ways it was almost expected that when searching online about pregnancy-related issues, one would come across worst case scenarios or as some women referred to them, horror stories. The pregnant women in the study referred a lot to reading online forums, and it was particularly recognized that forums were a place people who have, or have had, problems tend to go to share their story. It was also recognized and accepted that healthy people do not generally go on forums to say all is well, unless it was in response to a worst case scenario in a supportive way, such as experiencing something similar and now having a healthy baby. This also leads to the acknowledgment of reliability of information, in that it was frequently acknowledged that anybody can post on forums and blogs about their issues and experiences and this is often unmoderated. However, these are people’s opinions and personal experience which may not apply to others and certainly not everyone. Again this shows a level of how skilful surfing is and how people navigate the vast amount of information online. Knowing what certain information online does and does not provide differs depending on the source of that information.

No real difference was evident from the findings of any difference in the online searching approaches or styles in terms of pregnancy status between primiparous and multiparous participants. However, the authors would suggest further research in this area to perhaps investigate any change in confidence or anxiety levels during pregnancy and how this may impact online searching. It was noted that 1 multiparous participant had noticed a difference in her online searching between her two pregnancies. As she knew much more during her second pregnancy that certain symptoms or concerns were normal, her previous experience had notably, to her, reduced her need to search for information online.

Strengths and Limitations

This methodology did present a number of limitations. First, the sample was self-selecting and was only relevant to pregnant women who did use the Web for information and support. The sample may reflect only those who feel competent with Web-based activities and who have reasonable information technology (IT) literacy skills. Due to the inclusion criteria, the study missed out on the insights of pregnant women who did not consider Web-based resources as useful during pregnancy. The women in the study all talked about googling symptoms and reading posts from other women who had posted on forums. It would therefore be interesting for future work to understand the main issues posted on forums that pregnant women discuss and also how the information received via forums and other pregnant women's stories in particular, impact on decision making during pregnancy. Previous research has considered the impact on decision making of online information [5]; however, more research is needed in this area especially in terms of support online.

A second limitation of the study is the sample size. Since this is a qualitative study, a sample size of 16 interviews is sufficient. However, triangulation of these findings with a larger more quantitative study would add weight to the findings discussed here. The authors have data from a Web-based questionnaire but this was beyond the scope of discussion in this paper but will add weight to these findings in the future. A major strength of this study is that not only did the authors conduct interviews with currently pregnant women but the health information and other demographic data was also gathered. Having knowledge of information pertaining to women's pregnancy at the time of interview, such as trimester and medical conditions, enabled the researchers to consider factors which may influence use of the Web and how women searched. In terms of demographics, all but one of the women interviewed was white; perhaps a more ethnically diverse sample would yield differing results. In addition, this research gained knowledge of the participant's experience of pregnancy-related complications, either during this pregnancy or a previous pregnancy. Previous research [6] has found health-related anxiety to increase in women during pregnancy who have experienced complications during pregnancy. More qualitative research into this area may provide insight into this area.

None of the women interviewed were in their first trimester of pregnancy. Due to the sampling technique employed, it was observed by the authors that the women approached to take part in the study who were in their first trimester of pregnancy did not want to be interviewed until they had reached the second trimester. This is understandable due to the increased risks of pregnancy loss in the first 12 weeks (first trimester). However, since the first trimester is a particularly worrying time for pregnant women and a time when symptoms such as morning sickness are often most persistent, it would be interesting for research to gain an insight of how women may change their online information or health searching during this period.

No longer helping, saturation, and feeling reassured where the main reasons why women stopped their online pregnancy-related searches. Women in the study felt confident they could navigate through the mass of data and there was an element of knowing when to stop searching. However, more research is needed to dig deeper and understand how pregnant women, and indeed other people seeking health information online, evaluate the quality of information sourced in order to gain a more complete insight into what constitutes skillful surfing.

Conclusions

Searching for health information and advice online during pregnancy is viewed as quick, easy, and accessible. The affordances of the Web provide pregnant women the opportunity to go online as a first port of call. Knowing they were not alone, ascertained through the reading of online posts by other pregnant women sharing their experiences or symptoms, enabled women to normalize their experience and despite some caution required, was ultimately reassuring for pregnant women.

Acknowledgments

The authors would like to thank the University of Bolton for funding the project through the graduate intern scheme and the Jenkinson Award. The authors would like to thank research assistant Candice Simms for her work on the project, the National Childcare Trust; in particular Dr Abigail Easter for all her support and promotion of the study, and all the women who took part in the study; both the qualitative and quantitative elements.

Authors' Contributions

Both the authors were involved in the concept and design of the study as well as data analysis. All authors had major contributions to the write-up and editing of the manuscript. Both authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Fox S, Duggan M. Health online health-online-/. 2013. Pew Research Center URL: http://www.pewinternet.org/2013/01/15/ [accessed 2017-05-18]
2. Pew Research Center. Health topics: Food safety, drug safety, and pregnancy information are among eight new topics included in our survey URL: http://www.jmir.org/2017/6/e194/ [accessed 2017-05-18]
3. Fox S. Health topicsight in ten adult internet users look for information online pewinternet. Pew Research Center URL: http://www [accessed 2017-05-18]

4. Rowlands IJ, Loxton D, Dobson A, Mishra GD. Seeking Health Information Online: Association with Young Australian Women's Physical, Mental, and Reproductive Health. J Med Internet Res 2015 May 18;17(5):e120 [FREE Full text] [doi: 10.2196/jmir.4048] [Medline: 25986630]
5. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
6. Kowalyk KM, Hadjistavropoulos HD, Jones SL. What impact does pregnancy have on anxiety about health? J Psychosom Obstet Gynaecol 2009 Dec;30(4):223-230. [doi: 10.3109/01674820903276453] [Medline: 19863458]
7. Song F, West J, Lundy L, Smith-Dahmen N. Women, pregnancy, and health information online: the making of informed patients and ideal mothers. Gend Soc 2012;26(5):773-798. [doi: 10.1177/0891243212446336]
8. Madden M, Rainie L. Pew Research Center 2003.
9. White M, Dorman SM. Receiving social support online: implications for health education. Health Educ Res 2001 Dec;16(6):693-707. [Medline: 11780708]
10. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 01;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
11. Lagan BM, Sinclair M, Kernohan WG. A Web-based survey of midwives' perceptions of women using the Internet in pregnancy: a global phenomenon. Midwifery 2011 Apr;27(2):273-281. [doi: 10.1016/j.midw.2009.07.002] [Medline: 19700228]
12. Huberty J, Dinkel D, Beets MW, Coleman J. Describing the use of the internet for health, physical activity, and nutrition information in pregnant women. Matern Child Health J 2013 Oct;17(8):1363-1372. [doi: 10.1007/s10995-012-1160-2] [Medline: 23090284]
13. Larsson M. A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery 2009;25(1):20. [doi: 10.1016/j.midw.2007.01.010.PMID:17408822]
14. Glaser B, Strauss A. The Discovery of Grounded Theorytrategies for Qualitative Research. Chicago: Aldine; 1967:A.
15. O'Reilly M, Parker N. 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. Qualitative Research 2012 May 17;13(2):190-197. [doi: 10.1177/1468794112446106]
16. Walker JL. The use of saturation in qualitative research. CJCN . Medline 2012;22(2):37-46.
17. Guest G. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Clarke V, Braun V, Hayfield N. Thematic Analysis. Inmith JA. editor. Qualitative psychology: a practical guide to research methods, 3rd edition. Londonage Publications Ltd; 2015:a.

Abbreviations

AIMS: Association for Improvements in the Maternity Services
HCP: health care professional
IT: information technology
NHS: National Health Service
NICE: National Institute for Health and Care Excellence

Edited by G Eysenbach; submitted 01.07.16; peer-reviewed by S Mosadeghi, D Da Costa, M Larsson, K Linden, AL Wennberg; comments to author 17.11.16; revised version received 12.01.17; accepted 03.04.17; published 02.06.17.

Please cite as:
Prescott J, Mackie L
"You Sort of Go Down a Rabbit Hole...You're Just Going to Keep on Searching": A Qualitative Study of Searching Online for Pregnancy-Related Information During Pregnancy
J Med Internet Res 2017;19(6):e194
URL: http://www.jmir.org/2017/19/e194/
doi:10.2196/jmir.6302
PMID:28583906

Original Paper

Improving Consensus Scoring of Crowdsourced Data Using the Rasch Model: Development and Refinement of a Diagnostic Instrument

Christopher John Brady[1], MD, MHS; Lucy Iluka Mudie[1], MBBS, MPH; Xueyang Wang[1], BS; Eliseo Guallar[2], MD, MPH, DrPH; David Steven Friedman[1,2], MD, MPH, PhD

[1]Dana Center for Preventive Ophthalmology, Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, United States
[2]Bloomberg School of Public Health, Department of Epidemiology, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Christopher John Brady, MD, MHS
Dana Center for Preventive Ophthalmology
Wilmer Eye Institute
Johns Hopkins University School of Medicine
600 N Wolfe St
Baltimore, MD,
United States
Phone: 1 410 502 2789
Email: brady@jhmi.edu

Abstract

Background: Diabetic retinopathy (DR) is a leading cause of vision loss in working age individuals worldwide. While screening is effective and cost effective, it remains underutilized, and novel methods are needed to increase detection of DR. This clinical validation study compared diagnostic gradings of retinal fundus photographs provided by volunteers on the Amazon Mechanical Turk (AMT) crowdsourcing marketplace with expert-provided gold-standard grading and explored whether determination of the consensus of crowdsourced classifications could be improved beyond a simple majority vote (MV) using regression methods.

Objective: The aim of our study was to determine whether regression methods could be used to improve the consensus grading of data collected by crowdsourcing.

Methods: A total of 1200 retinal images of individuals with diabetes mellitus from the Messidor public dataset were posted to AMT. Eligible crowdsourcing workers had at least 500 previously approved tasks with an approval rating of 99% across their prior submitted work. A total of 10 workers were recruited to classify each image as normal or abnormal. If half or more workers judged the image to be abnormal, the MV consensus grade was recorded as abnormal. Rasch analysis was then used to calculate worker ability scores in a random 50% training set, which were then used as weights in a regression model in the remaining 50% test set to determine if a more accurate consensus could be devised. Outcomes of interest were the percent correctly classified images, sensitivity, specificity, and area under the receiver operating characteristic (AUROC) for the consensus grade as compared with the expert grading provided with the dataset.

Results: Using MV grading, the consensus was correct in 75.5% of images (906/1200), with 75.5% sensitivity, 75.5% specificity, and an AUROC of 0.75 (95% CI 0.73-0.78). A logistic regression model using Rasch-weighted individual scores generated an AUROC of 0.91 (95% CI 0.88-0.93) compared with 0.89 (95% CI 0.86-92) for a model using unweighted scores (chi-square *P* value<.001). Setting a diagnostic cut-point to optimize sensitivity at 90%, 77.5% (465/600) were graded correctly, with 90.3% sensitivity, 68.5% specificity, and an AUROC of 0.79 (95% CI 0.76-0.83).

Conclusions: Crowdsourced interpretations of retinal images provide rapid and accurate results as compared with a gold-standard grading. Creating a logistic regression model using Rasch analysis to weight crowdsourced classifications by worker ability improves accuracy of aggregated grades as compared with simple majority vote.

(J Med Internet Res 2017;19(6):e222) doi:10.2196/jmir.7984

KEYWORDS

crowdsourcing; diabetic retinopathy; Rasch analysis; Amazon Mechanical Turk

Introduction

Overview

Diabetes mellitus (DM) is a highly prevalent disease affecting over 415 million individuals worldwide, 80% of whom reside in low- and middle-income countries [1]. By 2040, the prevalence of DM is expected to reach 642 million, with the largest increases seen in countries with developing economies [1]. In the United States, 21.0 million people had known diabetes in 2012 and another 8.1 million had undiagnosed diabetes [2]. Diabetic retinopathy (DR) is an important complication of DM, currently affecting approximately 93 million people worldwide, with 28 million of these suffering from vision-threatening DR [3]. It is estimated that the number of Americans with DR will reach 16 million by 2050, with 3.4 million of these individuals afflicted with vision-threatening DR [4].

While DR is the leading cause of vision loss in working age individuals [4], screening for DR is an effective and cost-effective means of identifying the disease early, referring affected individuals for appropriate therapies, and preventing vision loss [5-8]. Despite the increasing prevalence of DR, the annual increase in the number of practicing ophthalmologists is only 2% [9], largely in high-income countries [10]. As a way of overcoming human resource shortfalls and in order to increase adherence with DR screening recommendations more broadly, telehealth programs using nonmydriatic fundus photography and remote interpretation are increasing [11-13].

In addition to improving screening uptake, telehealth may provide ways to reduce provider, payer, and societal costs [14-16]. Among the costs of a telehealth program for DR screening are the fundus camera, the telehealth software package, and the human resources needed for image acquisition and interpretation. Fundus photo interpretation costs in DR screening may be high given labor-intensive interpretation protocols and the need to interpret multiple images per patient. Computerized, semiautomated image analysis techniques have been developed which may be able to reduce physician workload and screening costs [17-19]; however, these methods are not approved by the US Food and Drug Administration nor in wide use clinically at this time. As telehealth expansion continues, novel low-cost methods will be needed to interpret the large volume of fundus images expected with rising incidence of diabetes, especially in resource-poor settings and in large public health screenings.

The use of crowdsourcing in biomedical research is in its infancy, although some groups have used this method in public health research [20] and to interpret biomedical images [21,22]. Crowdsourcing has been used to categorize a number of fundus photos with a variety of diagnoses as normal or abnormal [23]. In a trial conducted in the United Kingdom using untrained graders, the sensitivity was ≥96% for normal versus severely abnormal and from 61% to 79% for normal versus mildly abnormal [23]. In a proof-of-concept study, we have demonstrated that untrained crowdsourced workers can rapidly and accurately identify images with DR [24]. We have also demonstrated that crowdsourcing workers can improve their ability to identify characteristic glaucomatous changes in optic nerve photographs [25]. In this study we seek to perform an external validation of our method of crowdsourcing DR identification using a public dataset of 1200 retinal photographs and explore methods of improving the determination of a consensus score from multiple individual crowdsourced grades including creating a logistic regression model that includes other data points collected at the time of the grading and a second model that weights the responses of graders based on ability in a training dataset using the Rasch model.

Crowdsourcing Background

Crowdsourcing is “an online, distributed problem-solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals” [26]. Distributed human intelligence tasking [26], a subset of crowdsourcing, can involve subdividing larger tasks into small portions and then recruiting a group of individuals to each complete these small portions, and only collectively, the entire task. Amazon Mechanical Turk (AMT) is an online distributed human intelligence market that allows access to thousands of people who can quickly accomplish small, discrete tasks for small amounts of money. Typical AMT tasks include tagging photos, translating words, or writing very short articles for websites. AMT has its own vocabulary used by workers (Turkers) and task administrators (Requestors). A human intelligence task (HIT) is a small job which may be performed in a matter of seconds or minutes and, once the work is approved by the Requestor, may pay $0.01 to $0.25 or more per task depending on the complexity of the HIT. A group of HITs is called a batch and is made up of similar HITs. Depending on the complexity of the task and the payment offered by the Requestor, a batch is often completed within minutes or hours of posting [27]. One particular application in the recent literature has been the use of crowdsourcing to generate ground-truth annotations for deep learning algorithm training and validation [28], which have themselves been recently demonstrated to be quite effective at DR retinal image classification [28,29]. Other types of crowdsourcing such as broadcast search have also been applied to retinal image grading through a 2015 Kaggle competition [30].

Finding the Consensus Grade

Several methods for aggregating multiple grades into a consensus score have been described in the biomedical literature, dating back several decades [31]. The simplest method, termed majority vote (MV), involves promoting the modal response to the crowdsourced determination, as described in Whitehill et al [32]. In a binary classification scheme, whichever response is selected by half or more of respondents becomes the consensus. While this approach is computationally simple, the differential ability of workers is ignored as is differential difficulty of the unique tasks. Therefore, other methods of aggregating scores have been explored that rely on patterns of individual worker responses over multiple tasks and comparisons with or incorporation with expert annotations where available [22,32-36]. Additionally, several investigators have explored incorporation of artificial intelligence or deep learning methodologies to aggregation of data [37]. Others have turned to methods of item response theory to improve aggregation by specifically looking for inattentive or malicious users and eliminating their data [38]. In this study, we hypothesized an improved consensus grade would be found using Rasch analysis–determined weights applied to logistic regression models.

Methods

Crowdsourcing Platform

An interface for fundus photo classification has been previously described for the AMT crowdsourcing platform [24]. The United Kingdom national screening program grading scale [39] was chosen due its broad clinical telemedicine deployment. For the purposes of the study, terms from this scale were translated into plain language: background retinopathy was called mild, preproliferative retinopathy was called moderate, and proliferative retinopathy was called severe. Maculopathy is defined as abnormal on a training image with otherwise moderate disease but is not coded separately. The AMT interface was designed to provide training on grading of DR within each HIT. This training includes 7 images annotated with the salient features of each level of retinopathy in plain language. Turkers are presented with the following text: “This is a photo of the inside of the eye. We are looking to label eyes as healthy or unhealthy with respect to diabetes. Rate this eye.” Turkers can hover their mouse over the adjacent training images (2 normal, 1 mild, 1 moderate, 3 severe) while reviewing the active test image (Multimedia Appendix 1). This layout allows for all of the training and grading to occur in one browser window. Turkers receive US $0.10 per image, with a 40% commission going to Amazon, for a total cost of US $1.40 per image.

Baseline Trial with Majority Vote Analysis

For the first phase of this project, 1200 images from the Messidor public dataset [40] were posted for 10 unique binary annotations to provide external validation of the prior proof-of-concept study (yielding a dataset of n=12,000). We previously found 10 annotations per image appears to produce the maximal area under the receiver operating characteristic (AUROC), with little benefit seen for >10 gradings per image [24]. The Messidor dataset is composed of 800 mydriatic and 400 nonmydriatic retinal fundus photos of universally high quality and resolution. The images are supplied with ground truth grading on the following scale:

- 0—normal: no microaneurysms, no hemorrhages (n=546)
- 1—1-5 microaneurysms, but no hemorrhages (n=153)
- 2—6-14 microaneurysms OR 1-4 hemorrhages, but no neovascularization (n=247)
- 3—15 or more microaneurysms OR 5 or more hemorrhages OR presence of neovascularization (n=254)

For the purposes of this study, the presence of 5 or fewer microaneurysms was felt to be clinically insignificant and thus we classified Messidor 0 and 1 images as normal (58%, n=699) and Messidor 2 and 3 images as abnormal (42%, n=501), to mimic the accuracy of an American Telemedicine Association Category 1 screening program [41]. Turkers were not made aware of the source of the images.

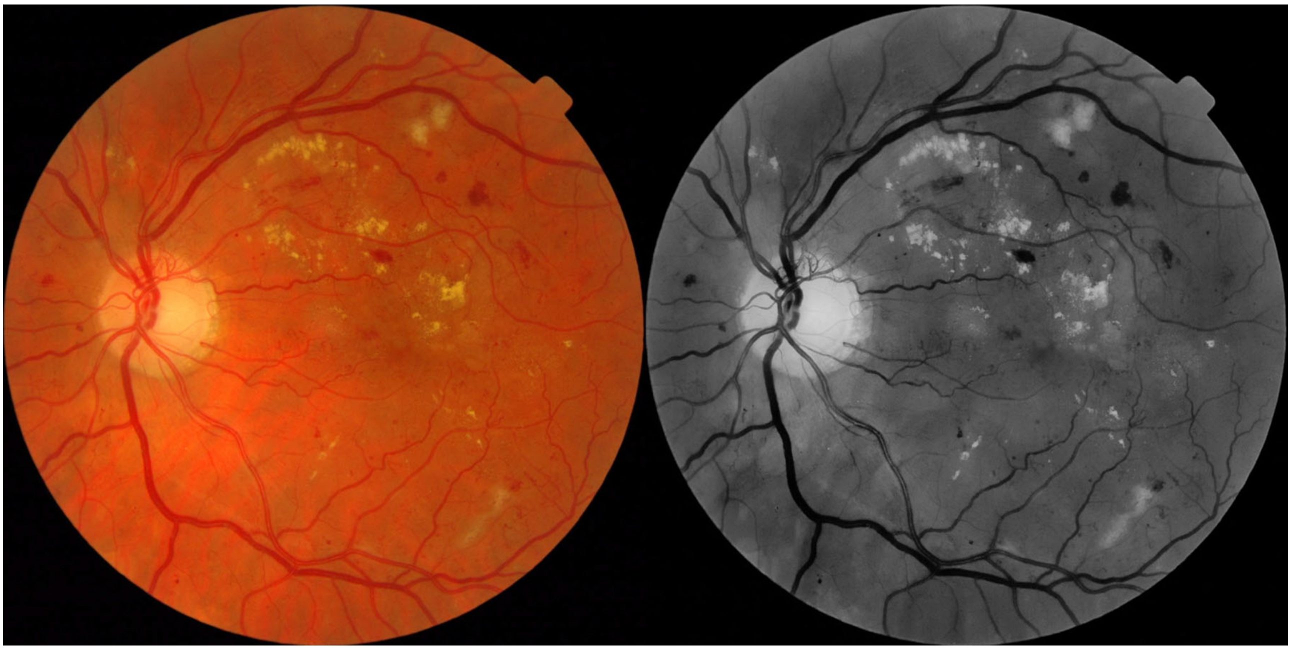
To create the MV consensus, each image was assigned the grade of abnormal if half or more Turkers deemed an image abnormal, otherwise the image was classified as normal. Sensitivity, specificity, and AUROC were calculated. This batch and grading scheme served as the baseline results for comparison with the regression models used in later phases of the research.

As there was no a priori rationale to suggest that the mean Turker score (with rounding toward abnormal) would provide the most accurate approximation of the ground truth classification, additional methods of generating consensus were explored.

Weighted Logistic Regression Model

For this phase, we recognized that among Turkers there is a range of ability, and among images there is a range of difficulty. In order to improve throughput, Turkers were asked to grade images in multiples of 10 (rather than single images), and to collect more data about the Turkers’ interactions with the task for future phases, the project was migrated to a new online interface using Volunteer Science (Multimedia Appendix 2). In the new interface, the 1200 Messidor images were posted for binary grading first using the full color images and then again with the images converted to grayscale with the red color channel removed in Adobe Lightroom (applied B&W Preset with green filter, in B&W Mix reduce red to –75) (Figure 1). This was done to simulate red-free images, which may allow for better detection of DR [42]. This allowed us to have a dataset with up to 30 grades per image across the 3 different batches (baseline, phase 2 color, and phase 2 black and white), albeit captured under slightly different circumstances. Throughout, Turkers were paid $0.10 per image.

Figure 1. Example color image and simulated red-free retinal photograph created by deleting the red channel in Adobe Lightroom.



The dataset of 1200 images was randomly divided into 600 training and 600 test images. The distribution of Messidor categories within these 2 subsets was within 2% of the entire dataset. Using the training images, a matrix of images and individual Turkers was created with each cell either being a missing datapoint (if that particular Turker did not grade that particular image), a 1 for a correct classification, or a 0 for an incorrect classification. Rasch analysis was then performed to determine the image measures and Turker measures based solely on the information in this matrix in the training set (see Figure 2 for the Rasch model [43], where, in this study, P_{ni} is the probability of a given image *n* of difficulty B_n having a correct response provided by Turker *i* of skill level D_i). Therefore, the Turker's ability measure and the image's difficulty measure are expressed as log-odds units (logits), theoretically ranging from –∞ to +∞. The negative exponentiated Turker ability measure, then, is the odds that an image of average difficulty (ie, B_n=0) would be categorized correctly by that particular Turker. This value was then multiplied by each of that Turker's categorizations from the test set (with abnormal =1, normal =–1). The weighted scores were then summed for each image. In an initial analysis, the consensus image score was considered to be abnormal if greater than or equal to zero and normal otherwise. Sensitivity, specificity, and AUROC were calculated as above with comparison to the baseline MV results. In a subsequent analysis, the consensus image score was included as a continuous variable in a logistic regression model to determine the ideal cut-off value for different values of percent correct, sensitivity, and specificity.

Data were analyzed using Stata Statistical Software: Release 14 (StataCorp LLC) and Winsteps Rasch measurement computer program (winsteps.com). The Johns Hopkins University Institutional Review Board (IRB) deemed this research IRB-exempt as nonhuman subjects research.

Figure 2. The Rasch model formula.

$$\log_e\left(\frac{P_{ni}}{1-P_{ni}}\right)=B_n-D_i$$

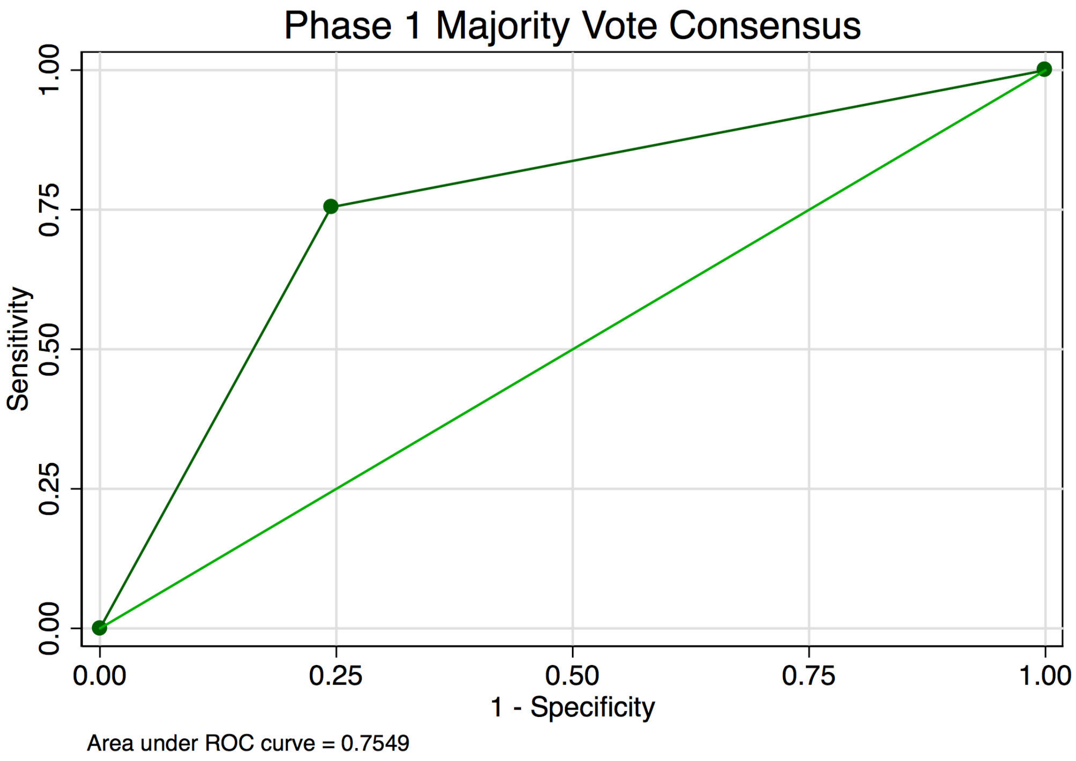
Results

Baseline Majority Vote

A batch of 12,000 (1200 images × 10 repetitions) tasks was posted on AMT March 13, 2015, 11:00 AM Eastern Time for a total cost of US $1440 ($1200 for Turker compensation, $240 for Amazon commission). The grading was complete in 68 minutes, with 97% of tasks (11,640/12,000) graded within 35 minutes. Tasks submitted without image grades were immediately reposted so there were no missing data. The tasks were submitted by 281 unique Turkers, with each submitting a mean of 42.7 tasks (median 28, mode 1). Turkers were only able to grade each image once.

The MV consensus was correct in 75.5% (906/1200) of images. Sensitivity and specificity were both 75.5%. The AUROC was 0.755 (Figure 3).

Figure 3. Receiver operating characteristic for the diagnosis of abnormal retinal photograph in the phase 1 baseline analysis.



Weighted Logistic Regression Model

For this phase, the focus shifted to the perspective of the Turkers rather than on the images themselves. Exploration of Turker accuracy motivated an attempt to incorporate Turker ability into a predictive model. There is expected to be a distribution of Turker accuracy that is not necessarily related to the number of tasks performed [44]. As such, any method that implicitly weights a consensus score based on number of tasks performed as does MV may reduce accuracy. In the phase 1 baseline task, among the 281 unique Turkers median percentage of images graded correctly was 64.7% (18 correct out of the median graded of 28) with an interquartile range of 55.5% to 74.4%.

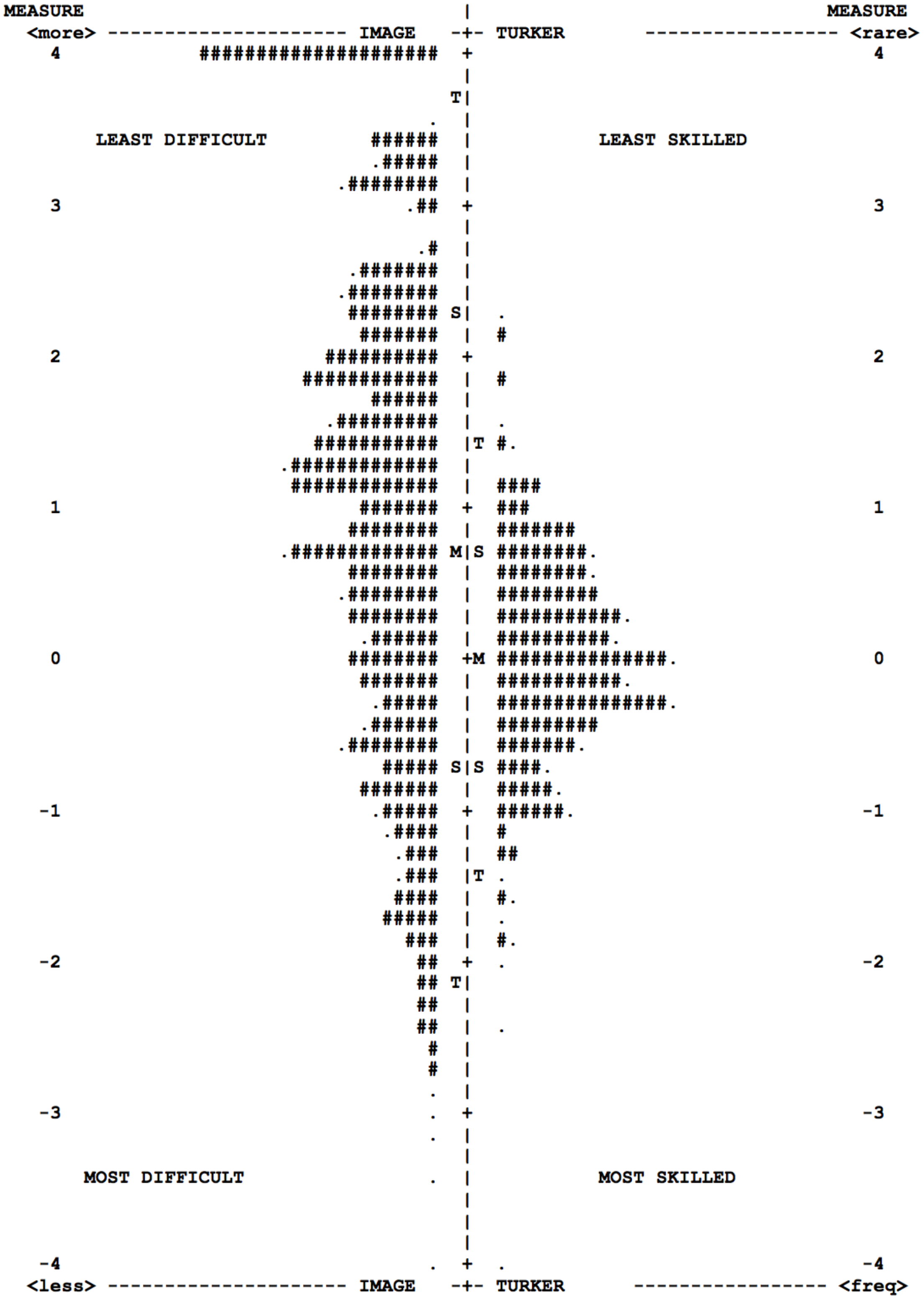
Prior to performing Rasch analysis, the results of the improved Volunteer Science interface (1200 color images + 1200 red-free images; cost US $2558, completed over 10 days) were merged with the phase 1 baseline classifications to permit as many grades as possible. In essence, we treated the Turkers as test takers taking a test involving grading multiple images. Duplicate grades of the same image by an individual Turker were deleted (6.7%; 2228/33,319 grades deleted). For stability, we also excluded 1027 grades by 227 Turkers who had graded fewer than 10 images within the training set of 600 images (leaving 301/528 Turkers and 14,539/15,566 grades). No images were excluded. Using Rasch analysis, we found Turker ability ranging from the most highly skilled at –3.75 logits to the least skilled at 1.9 logits. The median ability is set in the model as zero, and the interquartile range of ability was –0.40 to 0.47 (Figure 4).

After transformation, the Turker measure scores from log-odds to odds of correctly classifying an average difficulty image, weights outside the top and bottom centiles (1%) were truncated to the level of the 1st and 99th centile to increase stability and minimize the effect of outliers.

When the Turker weights were applied to the classifications in the color images in the test set using an arbitrarily determined cut-off (0), the percent correctly classified improved to 80.7% with an AUROC of 0.817 (from 74.0% correct and AUROC 0.739 for the test set images only in the phase 1 baseline task).

To determine if the arbitrary cut-off could be improved, a logistic regression model using the consensus image score determined by the weighted Turker classifications was generated. Using this model, a much more granular receiver operating characteristic (ROC) could be generated. A similar ROC was generated from a separate regression model using the unweighted consensus classifications from the same batch (Figure 5). The AUROCs were 0.908 (95% CI 0.883-0.933) and 0.889 (95% CI 0.862-916), respectively (chi-square *P* value<.001). A post hoc sensitivity analysis exploring the decision to dichotomize the Messidor dataset between grades 1 and 2 was performed by excluding all Messidor grade 1 images. In this analysis, the AUROCs were 0.919 (95% CI 0.896-0.943) for the Rasch-weighted scores and 0.888 (95% CI 0.861-0.915) for the unweighted scores (chi-square *P* value<.001). An additional sensitivity analysis was performed by using the entire 1200 images to generate Rasch Turker measures, and then the AUROC was calculated by using a Jackknife cross-validation excluding 1 image each time for 1199 repetitions [45]. Using this method, the AUROC was 0.909 (95% CI 0.892-0.926).

Figure 4. Image-Turker map illustrating distribution of measure scores for image grading difficulty and Turker ability (#=2 images/Turkers, .=1 image/Turker, M=mean score, S=1 standard deviation, T=2 standard deviations).



```
MEASURE                                  |                               MEASURE
  <more> ---------------------  IMAGE   -+- TURKER       ----------------- <rare>
    4             #####################  +                                   4
                                         |
                                        T|
                                      .  |
        LEAST DIFFICULT          ######  |           LEAST SKILLED
                                 .#####  |
                              .########  |
    3                               .##  +                                   3
                                         |
                                     .#  |
                               .#######  |
                              .########  |
                               ######## S|  .
                                #######  |  #
    2                        ##########  +                                   2
                           ############  |  #
                                 ######  |
                             .#########  |  .
                            ###########  |T #.
                         .#############  |
                          #############  |  ####
    1                           #######  +  ###                              1
                               ########  |  #######
                         .############# M|S ########.
                               ########  |  ########.
                              .########  |  #########
                               ########  |  ###########.
                                .######  |  ##########.
    0                          ########  +M ###############.                 0
                                #######  |  ###########.
                                 .#####  |  ###############.
                                .######  |  #########
                              .########  |  #######.
                                  ##### S|S ####.
                                #######  |  #####.
   -1                            .#####  +  ######.                         -1
                                  .####  |  #
                                   .###  |  ##
                                   .###  |T .
                                   ####  |  #.
                                  #####  |  .
                                    ###  |  #.
   -2                                ##  +  .                               -2
                                     ## T|
                                     ##  |
                                     ##  |  .
                                      #  |
                                      #  |
                                      .  |
   -3                                 .  +                                  -3
                                      .  |
                                         |
       MOST DIFFICULT                 .  |           MOST SKILLED
                                         |
                                         |
                                         |
   -4                                 .  +  .                               -4
  <less> ---------------------  IMAGE   -+- TURKER      ----------------- <freq>
```

Examination of multiple dichotomization cut-points revealed that choosing a cut-off that would permit a minimum sensitivity of 90.3% allows for specificity of 68.5% and percent correctly classified at 77.5% with an AUROC of 0.79 (95% CI 0.76-0.83) (Table 1,Figure 6).

Rasch analysis also allowed for a qualitative analysis of the retinal images. The images were sorted by image measure on the logit scale as generated by the Rasch analysis described earlier. The 20 images with the lowest measures ranged from –4.75 to –2.56 logits, which corresponds to the log odds that a Turker of average ability (ie, Turker measure = 0) would grade these images correctly. Among these 20 images, 0, 1, or 2 Turkers (out of maximum of 30) graded each correctly, and thus these images were designated the most difficult to grade. The 20 images with the highest measures were selected as the easiest to grade. A total of 20 sequential images were then selected at the 3 image measure quartiles as successively less difficult images to grade (Table 2,Figure 7). The hardest images were largely Messidor grade 0 and 1 images with some abnormal features but without significant DR (eg, chorioretinal atrophy, choroidal nevus) that had been graded as abnormal by Turkers. Intermediate images were mostly Messidor grade 2 images with extrafoveal microaneurysms of subtle hard exudates as well as Messidor grade 0 images without any nondiabetic pathology or distracting features. The easiest images were generally Messidor grade 3 with prominent hard exudates apparent.

Because data on the time spent completing the task and prior exposure to similar tasks is collected in addition to the grade for the current image task when a crowdsourcing worker completes a task, a separate logistic regression model that incorporated variables for time spent on each task and prior experience with ophthalmic HITs was run but did not improve diagnostic accuracy (data not shown).

Table 1. Characteristics of different cut-point values using the weighted logistic model, as compared with the majority vote weighted cut-point and the phase 1 baseline task.

		Correct %	Sensitivity %	Specificity %	AUROC[a] 95% CI
Phase 1 MV[b] baseline		75.5	75.5	75.5	0.75 (0.73-0.78)
MV weighted arbitrary cut-point		80.7	87.1	76.1	0.82 (0.79-0.85)
Weighted regression					0.91 (0.88-0.93)
	Maximizing % correct	85.0	81.1	87.8	0.84 (0.81-0.87)
	Sensitivity ≈ 90%	77.5	90.3	68.5	0.79 (0.76-0.83)
	Specificity ≈ 90%	84.5	76.6	90.1	0.83 (0.80-0.86)

[a] AUROC: area under the receiver operating characteristic.

[b] MV: majority vote.

Table 2.

Difficulty	Measure score range logits	Images graded correctly %	Messidor grade mode
Hardest	–4.74 to –2.56	0-8.3	1
Intermediate 1	–0.14 to –.04	43.4-53.9	0
Intermediate 2	1.01-1.1	69.2-76.9	0
Intermediate 3	2.04-2.09	85.2-88.9	0
Easiest	4.5-4.91	100	3

Figure 5. Receiver operating curve generated from a logistic regression model using weighted consensus scores of the random 50% (600 images) test set and a second using the nonweighted scores from the same data.

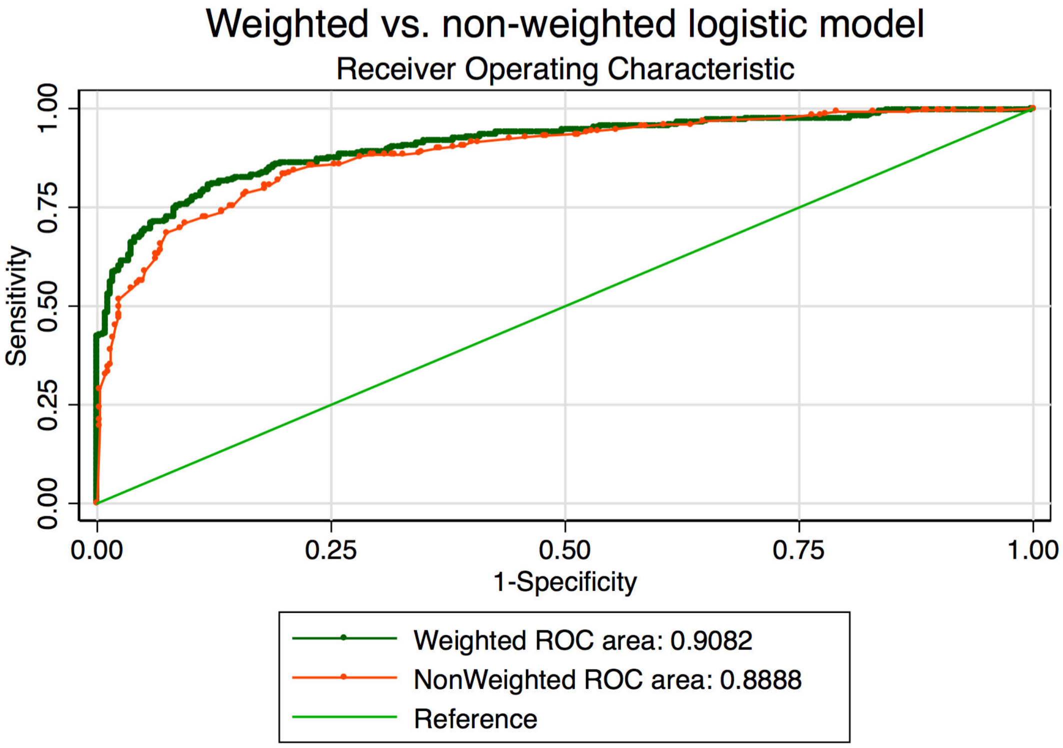


Figure 6. Receiver operating curve from logistic regression model using weighted consensus scores using a dichotomization cut-point designed to permit sensitivity of 90% shown alongside unweighted and Rasch-weighted majority vote cut-points.

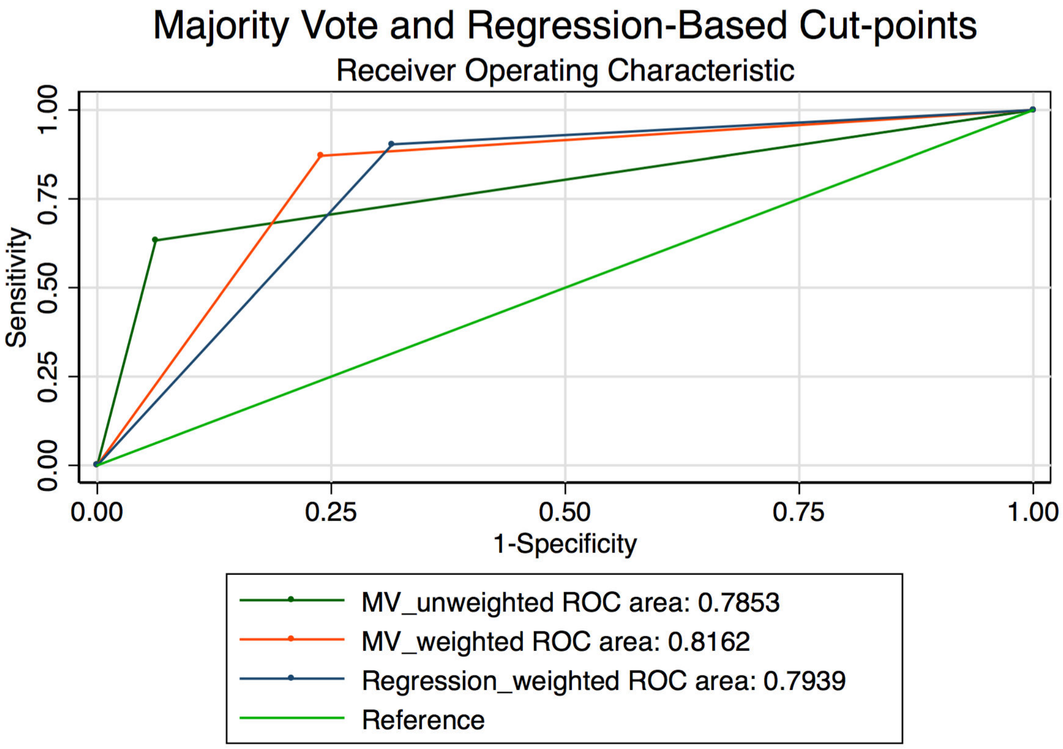
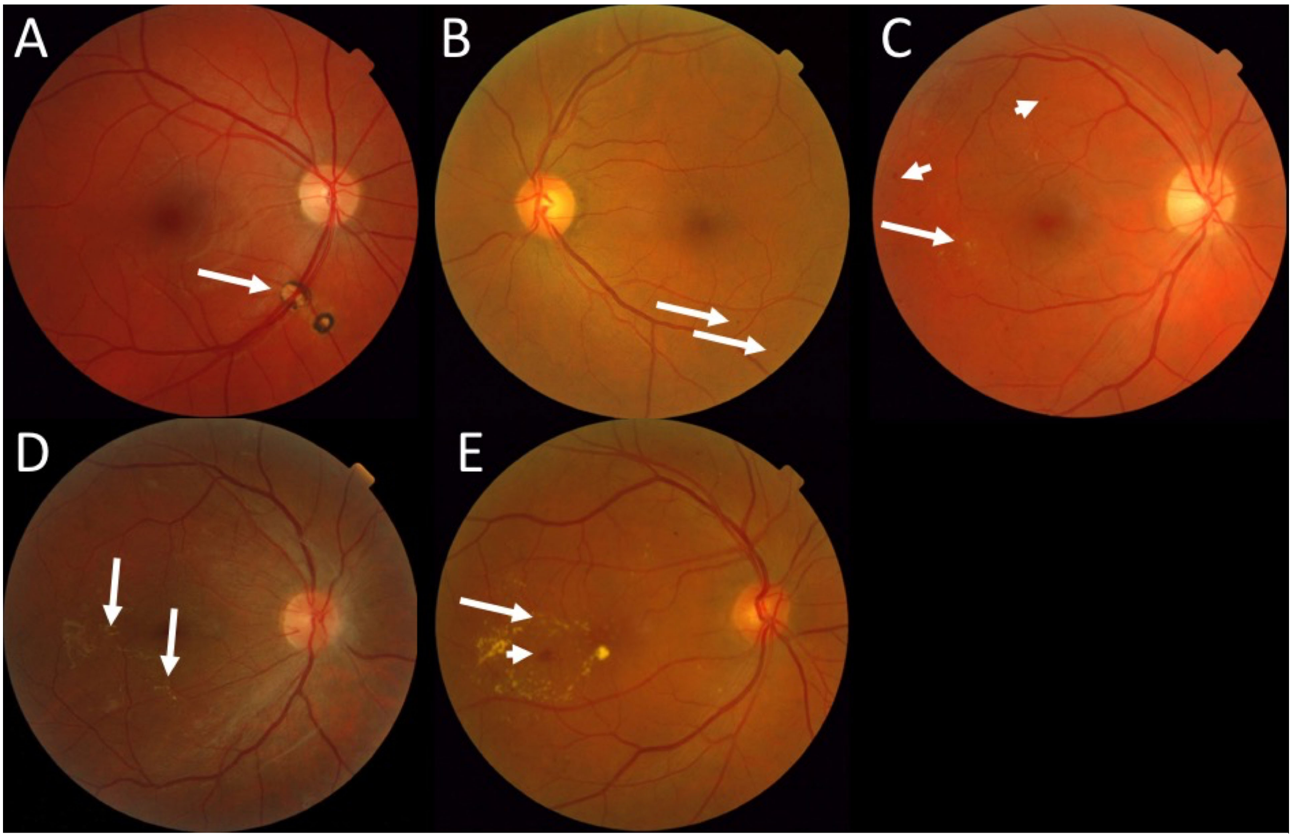


Figure 7. Representative retinal fundus images organized by progressive ease of grading correctly (A-E). (A) The image reveals areas of chorioretinal atrophy (arrow) but is without lesions of diabetic retinopathy. (B) This image reveals very subtle microaneurysms (arrows). (C) This image reveals more obvious microaneurysms (arrowheads) and subtle hard exudates (arrow). (D) This image reveals more apparent hard exudates (arrow). (E) This image reveals obvious hard exudates (arrow) and more obvious hemorrhagic microaneurysms (arrowhead).



Discussion

We have shown that workers on a popular crowdsourcing platform AMT are able to rapidly and accurately identify mild to moderate DR in a large public dataset and that weighting Turker responses by their demonstrated ability improves the accuracy of their crowdsourced grades.

There are many ways of defining a crowdsourcing consensus, or “divining the wisdom of the crowd.” For binary tasks or categories that can be rationally dichotomized (as was done in this study by reducing 4 levels of disease to disease or no disease), one could take a simple MV approach such that the image receives the categorization rendered by half or more of respondents. To reach a consensus with categorical data, using the modal response may reduce the influence of outlier or inattentive/malicious users. Both methods involve a post hoc analysis of the data. Alternatively, one could allow consensus to be determined on the fly, such that if enough workers render the same or similar judgment of an image, the image is immediately coded with this classification so that the full 10 responses need not be completed.

In this study, we sought to determine whether knowledge of an individual Turker’s ability on a training set of images could be used to improve accuracy of the consensus grade in a separate test set of images. We chose to use the Rasch model with image difficulty as the latent trait. In this way, we were able to determine the odds of each image being correctly classified by a Turker of average ability and the odds of each Turker being able to correctly grade an image of average difficulty. Using the entire 600 image training set, we saw reasonable targeting of Turker ability and image difficulty. This allowed us to weight a Turker’s response to the images in the test set for use in a logistic regression model. This also allowed for a qualitative assessment of the retinal images from a unique perspective, ranked from difficulty to grade correctly rather than ranked by disease severity.

While we were not able to meet all assumptions of the Rasch model for this study, we noted that the use of weighting Turkers’ responses showed a small but significant improvement in the AUROC as compared with unweighted aggregation. This result was very encouraging and suggests several possible improvements that can be made to our crowdsourcing method. For example, if a returning Turker has previously had their ability calculated, this can be immediately applied to their new categorizations. If a new Turker begins a retinal grading task, they can be asked to perform a brief quiz to determine their ability prior to officially grading images. This method may allow for a reduction in the number of annotations per image required to generate a stable estimate for each image. Moreover, we believe the use of regression methods to be of benefit as they allow diagnostic cut-points to be set based on the specific needs of the clinical or research paradigm (ie, to select the balance between sensitivity and specificity). We have also demonstrated the utility of a brief training prior to completing annotation of glaucomatous optic nerve images [25]. The relative (or combined) utility of these 2 approaches remains to be tested.

Additionally, in a recent study exploring the use of deep learning artificial intelligence for retinal image interpretation published by researchers at Google [28], a stated limitation was their use of MV consensus grading of several ophthalmologists in both their 128,000+ image training set and 11,700+ image test sets. The authors acknowledged that much of the residual imprecision of the algorithm likely resides in feeding better gold standard data into the algorithm, creating an opening for similar methods as described here.

There are several limitations to crowdsourcing retinal image processing. Because users are anonymous and cannot be directly selected by the researcher, there is no way to ensure high quality, conscientious workers each time work is posted. Since we did not collect demographic information from our Turkers, there was no way to use Turker factors to predict worker accuracy. Indeed, the pool of workers can vary substantially over time and different trends in how workers engage with the site have become apparent to us over the course of the 3 years of this experiment. For example, we have recently noticed that many workers use automated scripts to accept or reserve large numbers of tasks at once, and then they can proceed at their own pace without concern for there being few tasks left for them. This hoarding has made metrics of time spent per image rather meaningless, but it is not clear that it has led to worse outcomes overall (data not shown). Regardless, researchers who wish to use crowdsourcing need to be aware of the culture of the crowdsourcing marketplace they choose.

Our current method used the supplied Messidor grade as the gold standard. While this is a high-quality, well-known dataset, there were dramatic differences in how the images were graded compared to standard clinical and telemedicine grading schemes such as the one we used for training. Particularly, while we tried to mitigate clinically insignificant disease by defining the very mild disease category (Messidor 1) as normal, there was still the possibility of clinically very mild disease in the most severe Messidor category (eg, 16 microaneurysms is Messidor 3 but could be considered minimal retinopathy on most clinical grading scales). Since our sensitivity analysis suggested slightly better diagnostic accuracy when completely excluding all Messidor level 1 images, we believe our dichotomization was appropriately conservative.

There are several potential benefits to the use of crowdsourcing for the interpretation of visual data in ophthalmology. First, an inexpensive, rapid, and accurate system to reduce the number of images needing human grading in large public health screenings is needed. Importantly, this model should also be scalable; although the cost of grading per image here was greater than we have previously reported due to increased Amazon fees, the Turker compensation may have some elasticity which could be formally tested in the future. Despite these increased costs, crowdsourcing may be less expensive than other models of automated retinal image analysis and may be combined with other models to save costs and increase scalability. An approach which accurately identifies normal (or very mildly abnormal, allowing for some false negatives) fundi would be of great value and could reduce the skilled grader burden by up to 26% to 38% or more according to some investigators using artificial intelligence programs [19]. A first pass to remove normal images is currently being done with an artificial intelligence solution in Scotland’s national screening program [46]. A similar first pass or low-level annotation scheme was also validated using crowdsourcing to improve the accuracy and efficiency of expert grading of pathology slides for breast cancer [22]. If appropriately validated, crowdsourcing retinal images could provide a similar service at lower cost and with less infrastructure in all settings but could be particularly attractive in resource-poor settings. Likewise, a means to rapidly interrogate existing datasets could allow for nimble hypothesis generation for secondary data analyses. Overall, our results suggest that generating weighted classifications with Rasch analysis, which are then used in a weighted logistic regression model, may improve the accuracy of information obtained by crowdsourcing to grade retinal images for diabetic retinopathy.

Acknowledgments

This publication was made possible by the Johns Hopkins Institute for Clinical and Translational Research (ICTR), which is funded in part by grant KL2TR001077 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the Johns Hopkins ICTR, NCATS, or NIH. Funding was also received from the Johns Hopkins–University of Maryland Diabetes Research Center Pilot and Feasibility Award under National Institute of Diabetes and Digestion P30DK079637 (principal investigator Hussain). The retinal images used in this study were kindly provided by the Messidor program partners (www.adcis.net/en/DownloadThirdParty/Messidor.html).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the Amazon Mechanical Turk Web interface for fundus photo grading.

[JPG File, 562KB - jmir_v19i6e222_app1.jpg]

Multimedia Appendix 2

Screenshot of the Volunteer Science hosted Web interface for fundus photo grading.

[JPG File, 570KB - jmir_v19i6e222_app2.jpg]

References

1. IDF Diabetes Atlas. Brussels, Belgium: International Diabetes Federation; 2015. URL: http://www.diabetesatlas.org/ [accessed 2017-04-24] [WebCite Cache ID 6pySYEPZV]
2. National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014. Atlanta: US Department of Health and Human Services URL: http://www.cdc.gov/diabetes/pubs/statsreport14/national-diabetes-report-web.pdf [accessed 2017-05-04] [WebCite Cache ID 6TIZswZP6]
3. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, Meta-Analysis for Eye Disease (META-EYE) Study Group. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 2012 Mar;35(3):556-564 [FREE Full text] [doi: 10.2337/dc11-1909] [Medline: 22301125]
4. Saaddine JB, Honeycutt AA, Narayan KM, Zhang X, Klein R, Boyle JP. Projection of diabetic retinopathy and other major eye diseases among people with diabetes mellitus: United States, 2005-2050. Arch Ophthalmol 2008 Dec;126(12):1740-1747. [doi: 10.1001/archopht.126.12.1740] [Medline: 19064858]
5. Beck RW. The burgeoning public health impact of diabetes: the role of the ophthalmologist. Arch Ophthalmol 2011 Feb;129(2):225-229. [doi: 10.1001/archophthalmol.2010.331] [Medline: 21320972]
6. Javitt JC, Canner JK, Frank RG, Steinwachs DM, Sommer A. Detecting and treating retinopathy in patients with type I diabetes mellitus. A health policy model. Ophthalmology 1990 Apr;97(4):483-494. [Medline: 2109299]
7. Javitt JC, Aiello LP. Cost-effectiveness of detecting and treating diabetic retinopathy. Ann Intern Med 1996 Jan 01;124(1 Pt 2):164-169. [Medline: 8554212]
8. Jones S, Edwards RT. Diabetic retinopathy screening: a systematic review of the economic evidence. Diabet Med 2010 Mar;27(3):249-256. [doi: 10.1111/j.1464-5491.2009.02870.x] [Medline: 20536486]
9. Silva PS, Cavallerano JD, Aiello LM, Aiello LP. Telemedicine and diabetic retinopathy: moving beyond retinal screening. Arch Ophthalmol 2011 Feb;129(2):236-242. [doi: 10.1001/archophthalmol.2010.365] [Medline: 21320974]
10. Bastawrous A, Hennig BD. The global inverse care law: a distorted map of blindness. Br J Ophthalmol 2012 Oct;96(10):1357-1358 [FREE Full text] [doi: 10.1136/bjophthalmol-2012-302088] [Medline: 22740107]
11. Mansberger SL, Sheppler C, Barker G, Gardiner SK, Demirel S, Wooten K, et al. Long-term comparative effectiveness of telemedicine in providing diabetic retinopathy screening examinations: a randomized clinical trial. JAMA Ophthalmol 2015 May;133(5):518-525. [doi: 10.1001/jamaophthalmol.2015.1] [Medline: 25741666]
12. Sim DA, Mitry D, Alexander P, Mapani A, Goverdhan S, Aslam T, et al. The evolution of teleophthalmology programs in the United Kingdom: beyond diabetic retinopathy screening. J Diabetes Sci Technol 2016 Feb 01;10(2):308-317 [FREE Full text] [doi: 10.1177/1932296816629983] [Medline: 26830492]
13. Kim J, Driver DD. Teleophthalmology for first nations clients at risk of diabetic retinopathy: a mixed methods evaluation. JMIR Med Inform 2015;3(1):e10 [FREE Full text] [doi: 10.2196/medinform.3872] [Medline: 25705814]
14. Brady CJ, Villanti AC, Gupta OP, Graham MG, Sergott RC. Tele-ophthalmology screening for proliferative diabetic retinopathy in urban primary care offices: an economic analysis. Ophthalmic Surg Lasers Imaging Retina 2014;45(6):556-561. [doi: 10.3928/23258160-20141118-11] [Medline: 25423636]
15. Au A, Gupta O. The economics of telemedicine for vitreoretinal diseases. Curr Opin Ophthalmol 2011 May;22(3):194-198. [doi: 10.1097/ICU.0b013e3283459508] [Medline: 21460727]
16. Rein DB, Wittenborn JS, Zhang X, Allaire BA, Song MS, Klein R, Vision Cost-Effectiveness Study Group. The cost-effectiveness of three screening alternatives for people with diabetes with no or early diabetic retinopathy. Health Serv Res 2011 Oct;46(5):1534-1561 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01263.x] [Medline: 21492158]
17. Abràmoff MD, Folk JC, Han DP, Walker JD, Williams DF, Russell SR, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol 2013 Mar;131(3):351-357. [doi: 10.1001/jamaophthalmol.2013.1743] [Medline: 23494039]
18. Trucco E, Ruggeri A, Karnowski T, Giancardo L, Chaum E, Hubschman JP, et al. Validating retinal fundus image analysis algorithms: issues and a proposal. Invest Ophthalmol Vis Sci 2013 May 01;54(5):3546-3559 [FREE Full text] [doi: 10.1167/iovs.12-10347] [Medline: 23794433]
19. Goatman K, Charnley A, Webster L, Nussey S. Assessment of automated disease detection in diabetic retinopathy screening using two-field photography. PLoS One 2011;6(12):e27524 [FREE Full text] [doi: 10.1371/journal.pone.0027524] [Medline: 22174741]
20. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. Am J Prev Med 2014 Feb;46(2):179-187. [doi: 10.1016/j.amepre.2013.10.016] [Medline: 24439353]
21. Luengo-Oroz MA, Arranz A, Frean J. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. J Med Internet Res 2012 Nov 29;14(6):e167 [FREE Full text] [doi: 10.2196/jmir.2338] [Medline: 23196001]

22. Eickhoff C. Crowd-powered experts: helping surgeons interpret breast cancer images. 2014 Presented at: Proceedings of the First International Workshop on Gamification for Information Retrieval; 2014; Amsterdam. [doi: 10.1145/2594776.2594788]
23. Mitry D, Peto T, Hayat S, Morgan JE, Khaw K, Foster PJ. Crowdsourcing as a novel technique for retinal fundus photography classification: analysis of images in the EPIC Norfolk cohort on behalf of the UK Biobank Eye and Vision Consortium. PLoS One 2013;8(8):e71154 [FREE Full text] [doi: 10.1371/journal.pone.0071154] [Medline: 23990935]
24. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res 2014;16(10):e233 [FREE Full text] [doi: 10.2196/jmir.3807] [Medline: 25356929]
25. Wang X, Mudie LI, Baskaran M, Cheng C, Alward WL, Friedman DS, et al. Crowdsourcing to evaluate fundus photographs for the presence of glaucoma. J Glaucoma 2017 Jun;26(6):505-510. [doi: 10.1097/IJG.0000000000000660] [Medline: 28319525]
26. Brabham D. Crowdsourcing. Cambridge: MIT Press; 2013:0262518473.
27. Amazon Mechanical Turk. URL: https://www.mturk.com/mturk/welcome [accessed 2017-04-24] [WebCite Cache ID 6pySlfcfD]
28. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
29. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 2016 Oct 01;57(13):5200-5206. [doi: 10.1167/iovs.16-19964] [Medline: 27701631]
30. Kaggle. Diabetic retinopathy detection URL: https://www.kaggle.com/c/diabetic-retinopathy-detection [accessed 2017-06-07] [WebCite Cache ID 6r2qdTk45]
31. Dawid AP, Skene AM. Maximum likelihood estimation of observer error-rates using the EM algorithm. Applied Statistics 1979;28(1):20. [doi: 10.2307/2346806]
32. Whitehill J, Wu T, Bergsma J, Movellan J, Ruvolo P. Whose vote should count more? Optimal integration of labels from labelers of unknown expertise.: Advances in Neural Information Processing Systems; 2009. URL: http://papers.nips.cc/paper/3644-whose-vote-should-count-more-optimal-integration-of-labels-from-labelers-of-unknown-expertise [accessed 2017-05-04] [WebCite Cache ID 6qDSv83Ed]
33. Tang W, Lease M. Semi-supervised consensus labeling for crowdsourcing. 2011 Presented at: Special Interest Group on Information Retrieval, Workshop on Crowdsourcing for Information Retrieval; 2011; Beijing p. 1-6 URL: https://www.ischool.utexas.edu/~ml/papers/tang-cir11.pdf
34. Sheng V, Provost F, Ipeirotis P. Get another label? improving data qualitydata mining using multiple, noisy labelers. 2008 Presented at: Proceedings of the 14th ACM SIGKDD international conference on Knowledge Discovery and Data Mining; 2008; Las Vegas. [doi: 10.1145/1401890.1401965]
35. Dalvi N, Dasgupta A, Kumar R, Rastogi V. Aggregating crowdsourced binary ratings. 2013 Presented at: Proceedings of the 22nd International Conference on World Wide Web; 2013; Rio de Janeiro URL: http://www2013.org/proceedings/p285.pdf
36. Passonneau R, Carpenter B. Transactions of the Association for Computational Linguistics. The benefits of a model of annotation URL: https://www.transacl.org/ojs/index.php/tacl/article/view/389/54 [WebCite Cache ID 6r39tbmwX]
37. Rose S. Mortality risk score prediction in an elderly population using machine learning. Am J Epidemiol 2013 Mar 01;177(5):443-452. [doi: 10.1093/aje/kws241] [Medline: 23364879]
38. Hovy D, Berg-Kirkpatrick T, Vaswani A, Hovy E. Learning whom to trust with MACE. URL: http://www.cs.cmu.edu/~hovy/papers/13HLT-MACE.pdf [accessed 2017-06-07] [WebCite Cache ID 6r3AGPArU]
39. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
40. Decencière E, Zhang X, Cazuguel G, Laÿ B, Cochener B, Trone C. Image Analysis and Stereology. Feedback on a publicly distributed image database: the Messidor database URL: https://hal.archives-ouvertes.fr/hal-01082570/document [accessed 2017-05-04] [WebCite Cache ID 6qDTeXrpg]
41. Li HK, Horton M, Bursell S, Cavallerano J, Zimmer-Galler I, Tennant M, American Telemedicine Association Diabetic Retinopathy Telehealth Practice Recommendations Working Group, et al. Telehealth practice recommendations for diabetic retinopathy, Second edition. Telemed J E Health 2011 Dec;17(10):814-837. [doi: 10.1089/tmj.2011.0075] [Medline: 21970573]
42. Venkatesh P, Sharma R, Vashist N, Vohra R, Garg S. Detection of retinal lesions in diabetic retinopathy: comparative evaluation of 7-field digital color photography versus red-free photography. Int Ophthalmol 2015 Oct;35(5):635-640. [doi: 10.1007/s10792-012-9620-7] [Medline: 22961609]
43. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. San Diego: Mesa Press; 1993.

44. Snow R, O'Connor B, Jurafsky D, Ng A. Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks. 2008 Presented at: EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2008; Honolulu URL: http://dl.acm.org/citation.cfm?id=1613751
45. Rabe-Hesketh S, Everitt B. Handbook of Statistical Analyses Using Stata, 4th Edition. Boca Raton: Chapman and Hall; 2007.
46. Medalytix Retinal Screening. URL: http://www.medalytix.com/ [accessed 2017-04-24] [WebCite Cache ID 6pySuwkJl]

Abbreviations

AMT: Amazon Mechanical Turk
AUROC: area under the receiver operating characteristic
DM: diabetes mellitus
DR: diabetic retinopathy
HIT: human intelligence task
ICTR: Institute for Clinical and Translational Research
IRB: institutional review board
MV: majority vote
NCATS: National Center for Advancing Translational Sciences
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 06.05.17; peer-reviewed by J Pearson, Z Yan, C Kurz, J Chamberlain; comments to author 31.05.17; revised version received 07.06.17; accepted 08.06.17; published 20.06.17.

Please cite as:
Brady CJ, Mudie LI, Wang X, Guallar E, Friedman DS
Improving Consensus Scoring of Crowdsourced Data Using the Rasch Model: Development and Refinement of a Diagnostic Instrument
J Med Internet Res 2017;19(6):e222
URL: http://www.jmir.org/2017/6/e222/
doi:10.2196/jmir.7984
PMID:28634154

Original Paper

“Click for Closer Care”: A Content Analysis of Community Pharmacy Websites in Four Countries

Sandra Zwier[1], MSc, PhD

[1] Amsterdam School of Communication Research ASCoR, Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Sandra Zwier, MSc, PhD
Amsterdam School of Communication Research ASCoR
Department of Communication Science
University of Amsterdam
Post box 15796
Amsterdam, 1001 NG
Netherlands
Phone: 31 20 5255094
Fax: 31 20 5253681
Email: s.m.zwier@uva.nl

Abstract

Background: Combinations of professional and commercial communication are typically very controversial, particularly in health care communication on the Internet. Websites of licensed community pharmacies on the other hand tend to raise remarkably little controversy, although they typically contain controversial combinations of clinical and commercial services previously unprecedented in professional health care communication.

Objective: The aim of this study was to fill the void of knowledge about the combination of clinical and commercial services presented on the websites of licensed community pharmacies.

Methods: A content analysis of clinical and commercial services presented in a random sample of 200 licensed community pharmacy websites from Great Britain, the Netherlands, the Canadian provinces British Columbia and Manitoba, and the Australian states New South Wales and Western Australia was conducted.

Results: The top five specific services mentioned on the community pharmacy websites were cosmetic products (126/200, 63.0%), medication refill request options (124/200, 62.0%), over-the-counter medicine (115/200, 57.5%), complementary and alternative medicine (107/200, 53.5%), and home medical aids (98/200, 49.0%). On average, 72.5% (145/200) of the community pharmacy websites across the 4 countries included a combination of clinical and commercial services. A combination of clinical and commercial services was more often present on chain pharmacy websites (120/147, 82.8%) than single pharmacy websites (25/53, 47%; $P<.001$), and most often on the Canadian community pharmacy websites, followed by the Australian, British, and Dutch pharmacy websites, respectively ($P<.02$). Furthermore, more than half of the pharmacies’ homepages contained a combination of clinical and commercial images (107/200, 53.5%), and almost half of the homepage menus contained a combination of clinical and commercial items (99/200, 49.5%). The latter were, again, more common on chain pharmacy than single pharmacy websites ($P<.001$), with significant differences between countries ($P<.001$).

Conclusions: A considerable share of websites of licensed community pharmacies in Great Britain, the Netherlands, Canada, and Australia combine clinical services with commercial services. Previous research into the presence of a combination of commercial and professional services suggests that such a combination may lead to increased interest in commercial services that may be unnecessary or inappropriate to patients’ health.

(J Med Internet Res 2017;19(6):e205) doi:10.2196/jmir.6899

KEYWORDS

community pharmacy services; pharmaceutical services; online pharmacies; marketing of health services; commerce; pharmacy ethics

Introduction

Combinations of Professional and Commercial Health Communication

Combinations of professional and commercial communication are typically very controversial, particularly in health care, where vulnerable patient groups may become interested in goods or services that may be unnecessary, inappropriate, or even dangerous to their health [1-3]. That is why such a combination often gives rise to calls for stringent regulatory action, meanwhile generating significant volumes of research. Examples include a combination of clinical and commercial communication in cosmetic surgery [4,5], medical tourism [6,7], and robotic surgery [8,9]. Recent years have also witnessed a lot of alarm surrounding the surge of illicit online pharmacies that sell prescription drugs without a physician's prescription, and are notorious for the use of hard-line marketing techniques such as mail spamming and search engine contamination [10-12]. On the other hand, remarkably little controversy or research surrounds the websites of regular, licensed community pharmacies. However, as we will argue, community pharmacy websites are gaining increasing importance as online "health care hubs" where patient consumers may submit electronic prescriptions, order prescription and over-the-counter medication, consult pharmacists about a chronic condition, obtain a prescription for a minor ailment, download software to monitor their health, and order personal care and household products, among others. They are also one of the major mainstream venues where professional clinical and commercial communication come together on the Internet [13-17], and this could be a potential minefield where many patients can easily confuse commercial communication for professional advice.

This study aimed to fill the currently large void in knowledge about the contents of licensed community pharmacy websites, guided by the following research question: *To what extent do present-day licensed community pharmacy websites combine clinical and commercial communication*? This question was investigated through a systematic content analysis of licensed community pharmacy websites from Australia, Canada, Great Britain, and the Netherlands. Besides gauging a picture of the combination of clinical and commercial content on officially licensed community pharmacies' websites in a number of Western countries nowadays, this study also provided an overview of the range of clinical and commercial services put forward on present-day community pharmacy websites, guided by the subquestions: *What types of clinical services are mentioned on present-day licensed community pharmacy websites*? and *What types of commercial services are mentioned on present-day licensed community pharmacy websites*? Finally, although cross-national comparisons were not the study's aim, the study could also answer the subquestion: *What cross-national differences exist in the types of clinical and commercial services mentioned on community pharmacy websites*? Implications for public health education are also discussed.

Background

The ideal clinical professional stays away from involvement in commercial communication, or so it seems. Professional clinical associations often promote straight bans or use highly restrictive codes of conduct for members' commercial communication, sometimes against pressure from government administrations to loosen their regulations. The American Federal Trade Commission (FTC), for instance, has fought a lengthy statutory battle against the American Medical Association for restricting free competition through its persistent ban on physician advertising. This was later followed up by comparable cases of the FTC against other professional associations for restricting advertising, such as dentists, optometrists, accountants, and lawyers [18]. Also for instance the Israel Medical Association upheld a long-lasting public doctors' strike after the Israeli government proposed a loosening of the absolute ban on physician advertising in the country [19]. Professional clinical providers also typically have strong adverse attitudes toward marketing and advertising their services [20-22]. On the Internet, this is mirrored in the continuous effort by diverse medical and health associations to establish quality standards for professional health information. This can be seen with examples such as the "Health on the Net Foundation Code of Conduct" (HONcode) which states that "advertising should be clearly distinguished from editorial content." Similarly, the guidelines of the American Medical Association state that on the Internet "advertising and commercial sponsorship must not influence any editorial content and advertising must be easily discernible from editorial content" [23-25].

The principal argument against clinical professionals' involvement with commercial communication revolves around a perceived conflict between commercial communication on the one hand and the clinical profession's "fiduciary responsibility" on the other, that is, the duty to prioritize and protect patients' interests at all times [26,27]. The premise is that trust is a central feature of clinical professionals' relationship with their patients, and commercial communication could compromise this trust [3,20]. In persuasive communication literature, the trust that people have in the accuracy and reliability of an information source is referred to as "source credibility." It is one of the oldest [28] and most frequently studied factors in persuasive communication research [29]. Research exploring the role of source credibility in the acceptance of online health information often shows that online health information is seen as more credible and convincing when the source is a perceived expert, such as a physician or professional health organization, than a lay person [30-34]. A commercial service mentioned on a physician or hospital's website will thus be likely to be seen as a higher-quality service than the same service featured on, for instance, the website of a brand or an online retailer, such as Amazon.

Clinical and Commercial Services in Community Pharmacy

This study focused on the combination of clinical and commercial communication on community pharmacy websites. On the basis of the existing strong norms against combining clinical and commercial communication in the clinical professions, most people would be at least surprised, and possibly dismayed, if their family physician's website would promote a branded pain reliever or line of mascaras. However, it is not uncommon to see this on a community pharmacy website. For example, the website of Service Apotheek, one of the largest community pharmacy chains in the Netherlands, mentions clinical consultation services for patients with chronic conditions such as diabetes and chronic obstructive pulmonary disease (COPD). At the same time, the website presents promotion for a branded facial cream that is claimed to "fill the skin with youth" and a slimming belt that will allegedly "redefine the body contours instantly." Such a combination of clinical and commercial communication can be found in community pharmacies across the world today, from Australia [35], to Kenya [36], Peru [37], and the Vatican [38].

Given today's proliferation of commercial communication in community pharmacies, it may be assumed that a community pharmacy is largely evolving into a commercial retail business, with little in the way of an enduring role in the provision of clinical health care. Nonetheless, this is only one side of the coin. Community pharmacies' clinical care activities largely waned during the course of the 20th century, when the "golden age of doctoring" was in its heydays and pharmacists' role largely hinged on those of physicians [15,39,40]. As one pharmacist said about his time in a community pharmacy in the 1970s, "We were discouraged from talking to patients for fear we would disrupt the doctor-patient relationship" [41]. Pharmacies since, however, have taken on a more patient-centered role and are becoming increasingly more involved in patient therapy [42,43]. Moreover, in recent years many countries such as China, Great Britain, New Zealand, and Switzerland have installed programs for clinical services to be delivered by community pharmacies, in an effort to control the steep rise in costs of clinical care delivered by family physicians and hospitals [44-46]. Community pharmacies' clinical role thus has risen "like a phoenix from the ashes" [39]. Present-day clinical services offered through community pharmacies may range from patient education, to consultations for patients with chronic conditions, to primary care provision, to actual prescribing [39,47-50]. For instance, Walgreens pharmacy in the United States currently has 400 so-called "health care clinics" across the country where the public may obtain treatment for minor illnesses. Similarly, Boots pharmacy in the United Kingdom features a number of "online clinics" where people can obtain prescription-only medication without going to the doctor for conditions such as erectile dysfunction, period delay, and malaria prevention.

Study Aims

All in all, developments in community pharmacies over the last few years have led to the presence of a combination of clinical and commercial services that were previously unprecedented in the domain of professional clinical care. To what extent this is mirrored on community pharmacy websites however is unknown. Extant research on community pharmacy websites is scant, with very few exceptions [51,52]. Thus, this study aimed to fill the void in knowledge by documenting the types of clinical and commercial services present on licensed community pharmacies' websites in 4 Western countries and the extent to which the websites contain a combination of clinical and commercial services.

Methods

Content Analysis

A systematic content analysis was conducted to allow for an unobtrusive observation of representative samples of licensed community pharmacies' websites and their contents. To strengthen the generalizability of the findings beyond the specific context of one particular country, community pharmacy websites from 4 different countries and different continents were included in the study, namely from Great Britain, the Netherlands, Canada, and Australia. The choice for these 4 countries was in part based on practical reasons of the researchers' command of the languages. Furthermore, low and middle-income countries were not included because pharmacy legislation may often be fragmented with limited enforcement of regulations in these countries [53], and this would significantly hinder identification of the population of licensed pharmacies in these countries. Finally, the United States and New Zealand were excluded because they currently are the only two countries in the world where direct-to-consumer advertising (DTCA) of prescription medication is permitted [3]. This exceptional regulatory status may restrict the generalizability of the findings regarding commercial content on community pharmacy websites, to most other countries in the world where DTCA of pharmaceuticals is prohibited. All in all, this study involved a content analysis of community pharmacy websites from Great Britain, the Netherlands, Canada, and Australia as countries with highly institutionalized and enforced pharmacy legislation and a ban on DTCA for medical drugs. Without making unsubstantiated claims regarding the representativeness of these 4 countries for community pharmacy websites worldwide, this likely renders a conservative rather than exaggerated picture of combinations of clinical and commercial services on community pharmacy websites worldwide nowadays.

Sample

Sampling Method

The sample consisted of 200 licensed community pharmacy websites: 50 from Great Britain, 50 from The Netherlands, 50 from the Canadian provinces British Columbia and Manitoba, and 50 from the Australian states New South Wales and Western Australia.

The sample of British community pharmacies was randomly drawn from a directory of all 14,437 licensed community pharmacies from the British General Pharmaceutical Council, and the sample of Dutch pharmacies from a directory of all 1981 licensed pharmacies in the Netherlands from the Dutch Healthcare Inspectorate. Community pharmacy licensing and directories in Australia and Canada are organized by the given state or province, which is why the sampling of community pharmacies proceeded with the use of the directories of two states or provinces per country. For Canada, 25 community pharmacy websites were drawn from the directory of the 1256 licensed community pharmacies of the College of Pharmacists of British Columbia and 25 from the 415 registered pharmacies by the College of Pharmacists of Manitoba. For Australia, 25 community pharmacy websites were drawn from the directory of the 2514 licensed community pharmacies by the Pharmacy Council of New South Wales and 25 from the 637 licensed pharmacies by the Pharmacy Registration Board of Western Australia.

Since this study endeavored to study community pharmacy websites, community pharmacies in the directories without a (working) website were excluded from the sample; this concerned 33.3% (67/200) community pharmacies of the original selection (26 Australian, 13 British, 28 Canadian, and 0 Dutch original selections). Also, the pharmacy departments of mass merchants and health centers had to be excluded. This is because the websites of these merchants and centers do not allow a valid separation between services offered through the pharmacy department and services offered through other departments, such as those of family physicians or a nurse on a health care center's website or banking or petrol services on a retail centers' website. Inclusion of these websites thus would have led to a significant overestimation of combinations of clinical and commercial services on community pharmacy websites. This concerned 8.5% (17/200 community pharmacies of the original selection (4 Australian, 4 British, 7 Canadian, and 2 Dutch original selections). When a pharmacy in the random selection had to be excluded from the sample, the next one in the directory was always chosen.

Sample Characteristics

Table 1 provides an overview of the share of chain pharmacies versus single pharmacies per country in the sample, with the former defined as community pharmacies belonging to a group of four or more community pharmacies that operate under a joint brand name [54]. It can be seen that almost three quarters (73.0%; 147/200) of the community pharmacies in the sample operated under a chain pharmacy retail brand name, with little variation across countries.

Table 1. Share of single and chain community pharmacies in the sample.

Pharmacy business type	Community pharmacies				
	Australia, N=50	Great Britain, N=50	Canada, N=50	Netherlands, N=50	All, N=200
Single pharmacy, n (%)	15 (30)	10 (20)	12 (24)	16 (32)	53 (27.0)
Chain pharmacy[a], n (%)	35 (70)	40 (80)	38 (76)	34 (68)	147 (73.0)

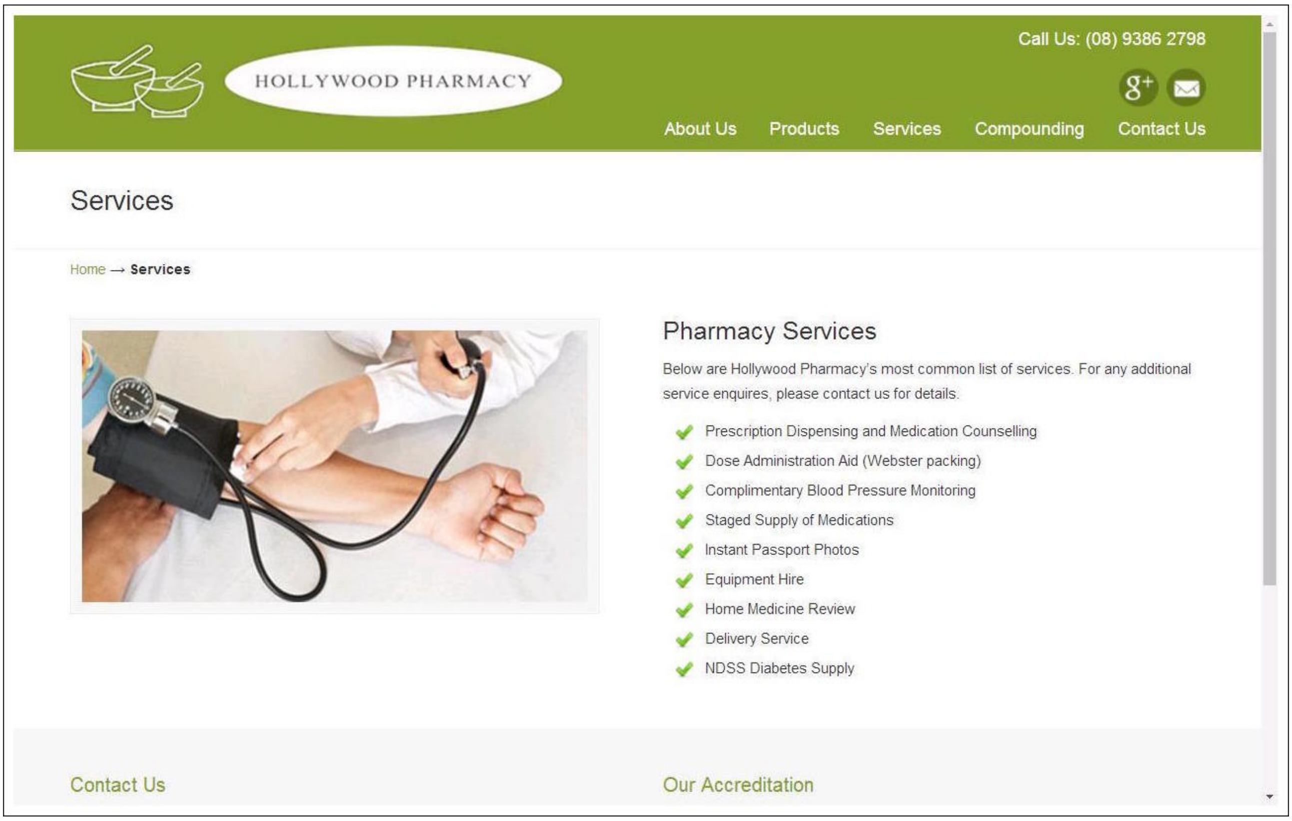
[a] Chain pharmacies are community pharmacies belonging to a group of four or more community pharmacies that operate under a joint brand name and often participate in one or more centralized programs [55].

Considering the present study's aims, it is of relevance to note that community pharmacies operating under a chain pharmacy retailer name often participate in one or more centralized programs, such as supply chains, pharmacist training schemes, or store brands. This could impact the variation in the services mentioned on the websites. However, the extent to which individual pharmacies partake in the pharmacy chain's centralized programs varies between, as well as within, pharmacy chains, while individual pharmacies may also include their own local services on the website. For instance, the Boots community pharmacy from the town of St Austell in the United Kingdom in our sample listed 10 services on its website, whereas the Boots pharmacy from the British town of Peebles in our sample listed only 2 services. That is why each community pharmacy website in the sample was coded separately and often contributed to variation in the data.

Coding Instrument and Procedure

Figures 1-4 show screenshots of community pharmacy websites from the sample. Each community pharmacy website in the sample obtained a binary code for the presence or absence of specific categories of clinical and commercial services on the website. Since a number of discrete products and services are concerned that obtain their meaning from the trade's definition of these products/services, measures of intersubjective agreement were not deemed relevant here and the coding was done by a single author. Below is a rationale for and overview of the coding categories employed.

Figure 1. Screenshot of website Hollywood Pharmacy, Nedlands (Australia), taken on March 29, 2017.



Pharmacy Services

Below are Hollywood Pharmacy's most common list of services. For any additional service enquires, please contact us for details.

- ✓ Prescription Dispensing and Medication Counselling
- ✓ Dose Administration Aid (Webster packing)
- ✓ Complimentary Blood Pressure Monitoring
- ✓ Staged Supply of Medications
- ✓ Instant Passport Photos
- ✓ Equipment Hire
- ✓ Home Medicine Review
- ✓ Delivery Service
- ✓ NDSS Diabetes Supply

Figure 2. Screenshot of website Roadnight Pharmacy, Sidcup (Great Britain), taken on March 29, 2017.

88 Station Road, Sidcup, Kent, DA15 7DU. Tel: 02083 002 632.

☎ Call 0845 548 8556 for Online Shopping Customer Services.

£-0
0 items
Your cart is empty

Mepore Dressing 7 ..	Live Color Xxl Per..	Cow & Gate Nutripr..
£ 0.16	£ 3.90 RRP: £ 7.10 | YOU SAVE £ 3.20	£ 1.89

Figure 3. Screenshot of website Medsen pharmacy, Hilversum (Netherlands), taken on March 29, 2017.

Tijd en aandacht voor uw gezondheid

Welkom op de website van Apotheek Stephenson. We zijn gevestigd aan de Stephensonlaan 51 in Hilversum.

Kies voor service en gemak

Apotheek Stephenson biedt u een handig aanbod van services aan. U kunt bij ons 24 uur per dag en 7 dagen per week uw medicatie ophalen. Ook kunt u uw medicatie online bijbestellen, óf dit automatisch door de apotheek laten verzorgen, zodat u altijd voldoende voorraad heeft en de medicatie steeds weer voor u klaar ligt. Ook kunt u gemakkelijk via de Apotheek App uw handverkoop (met korting)

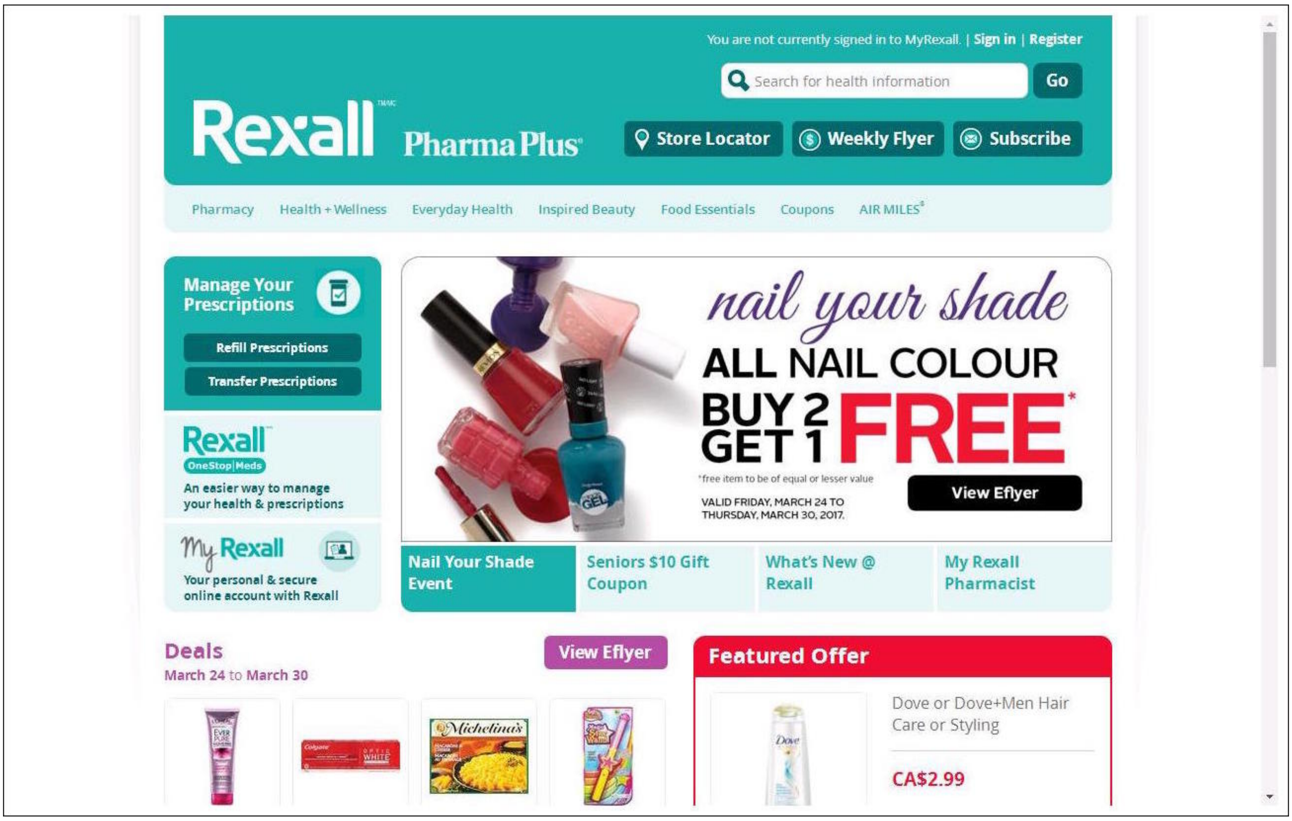
Contracten met alle zorgverzekeraars voor 2017

Ook in 2017 heeft uw apotheek weer contracten met alle zorgverzekeraars in Nederland. Heeft u vragen of wilt u meer informatie? Vraag het ons gerust.

De apotheek app

Met deze geheel

Figure 4. Screenshot of website Rexall pharmacy (Canada), taken on March 29, 2017.



Clinical and Commercial Services

Clinical Services

Clinical services were defined as clinical and pharmaceutical goods and services that are typically restricted to qualified pharmacists and other health care professionals. Three categories of clinical goods and services were distinguished in the coding, based on how current community pharmacy commonly defines its own roles [39,41-45]. The first of these is community pharmacy's traditional role of supervising the dispensing of medication. As community pharmacy in more recent years has expanded beyond the traditional function of dispensing and distributing medicine to more patient-based care, a second commonly recognized role of community pharmacy is disease detection and prevention. This includes services to maintain health such as vaccinations and stop smoking counseling, but also the timely detection of conditions such as hypertension or diabetes to combat the adverse consequences for patients more effectively. Third and finally, community pharmacy assumes a role in consulting patients about the effective and safe use of medication to achieve optimal patient outcomes, particularly for patients with chronic conditions such as hypertension, asthma, or diabetes [39,41-45].

In the coding procedure, each community pharmacy website was coded for the presence of the 3 above-mentioned roles of community pharmacy, specifically: *Dispensing services*, that is, services surrounding the dispensing of prescription medication (eg, dispensing of prescription medication, dispensing hormone replacement therapy), *Disease prevention and detection services*, that is, methods for diagnosing conditions and avoiding illness (eg, diabetes test, flu vaccinations), and *Clinical consultations*, that is, consultations for patients with specific conditions (eg, consultations for asthma patients, diabetes patients). Specific subcategories within these 3 coding categories, such as the testing of blood pressure, cholesterol, diabetes, and so on, were derived from the categorizations on the pharmacy websites themselves. Clinical services that typically belong to other professional health care domains such as veterinarian, optometric, or dental care services were excluded from coding. The screenshot of the website of Hollywood pharmacy from the town of Nedlands in Australia in Figure 1 illustrates the 3 categories of clinical services. Specifically, we see that the list of pharmacy services on this website includes dispensing services ("prescription dispensing," "dose administration aid"), disease detection and prevention services ("complementary blood pressure monitoring"), and clinical consultation services ("home medicine review," a consultation to review a patient's medication use to check for possible overlaps or interactions).

Commercial Services

Commercial services were defined as all services mentioned on the community pharmacy websites that are not typically restricted to licensed health care professionals. Again, 3 categories of commercial goods and services were distinguished. The first category reflects pharmacies' traditional role of the sale of health-related goods and services such as cold and cough medicine or pain relievers to ease and combat ills. The second category, the sale of nonhealth services, reflects the rise of the "drugstore" and "convenience store" concept in community pharmacy with an expanded focus on the sale of a range of everyday products such as cosmetics, snacks, and cleaning supplies. Finally, the third category reflects the move away from the brick-and-mortar commercial sale only, to an e-commerce model with the integration of a Web shop in community pharmacy websites.

In the coding procedure, each community pharmacy website was coded for the presence of the above-mentioned commercial roles of community pharmacy, specifically: *Health services*, specifically the sale of over-the-counter medicine (eg, pain relievers, first aid kits), complementary and alternative medicine (eg, homeopathic products, aromatherapy), diet products (eg, slimming shakes, gluten-free products), and home medical aids (eg, mattress covers, shower stools). *Non-health services*, specifically cosmetic products (eg, soaps, make-up products), food and drinks (eg, candy, snacks), and other goods or services (eg, greeting cards, pet products, kitchen appliances, photo services). *E-commerce*, that is, options to purchase services online through the pharmacy website, often through a Web shop. Again, the specific subcategories were derived from the categorizations provided by the pharmacy websites themselves. The screenshot of the website of Roadnight Pharmacy from Sidcup in Great Britain in Figure 2 is an example of a pharmacy website with an integrated Web shop. The Web shop offers a range products for sale, including health products such as the wound dressing product at the bottom of the screenshot and nonhealth services such as the hair dye product also at the bottom of the screenshot.

Menu and Images

Another part of the coding gauged the combination of clinical and commercial services on (1) the website menu and (2) through website images. The focus thereby was on the homepage as the place where users "land" after opening a pharmacy website.

Homepage Menu

The menu on the homepage of a community pharmacy website is the main tool that allows users to navigate to specific services. The menu can list items that connect to clinical services on the website (eg, "prescriptions," "immunizations"), items that connect to commercial services ("fragrances," "gifts"), or items that connect to other services (eg, "site map," "home"). In the screenshot of the website of Medsen Pharmacy from the Netherlands in Figure 3, for instance, the menu at the top of the page has 5 categories. From left to right, these are "About us," "Health ABC," "Our services," "Permission" (to share medical information with other health providers), and finally "Registration."

To gauge the presence of a combination of clinical and commercial services on the homepage menus, each item on the homepage menu was coded as connecting to a clinical service or a commercial service, using the definitions of clinical and commercial services as given above. A category "other" was used for menu items connecting to other utilities or services such as "contact" or "pharmacy locator."

Homepage Images

Images on community pharmacy homepages are usually either static pictures or sliding pages ("carousel format"). They can serve to highlight clinical pharmacy services, such as a photo of a patient having his blood pressure taken (see example of Hollywood pharmacy in Figure 1). They can also serve to highlight commercial services, such as images of nail polishes (see example of Rexall pharmacy in Figure 4). To gauge the presence of a combination of clinical and commercial services in homepage images, each image on the homepage menu was coded as referring to a clinical service or a commercial service, again using the definitions of clinical and commercial services as given above. A third coding category of "other" was used for images that referred to neither clinical nor commercial services, such as a photo of the pharmacy premises or staff members.

Data Processing and Analysis

The units of analysis were the community pharmacy websites. Chi-square tests of independence were used to test if the presence of a combination of clinical and commercial services on the community pharmacy websites was contingent on pharmacy types (chain pharmacies and single pharmacies), or country (Australia, Canada, Great Britain, and the Netherlands). Phi coefficients gauged associations between the presence of different categories of services on the websites. Significance levels of $P<.05$ were used for all tests.

Results

Clinical and Commercial Services Mentioned on the Websites

Clinical Services

An overview of the specific clinical services listed on the community pharmacy websites can be found in Table 2, whereby it should be noted that only clinical services mentioned on more than one website are listed in the table. In line with community pharmacies' traditional function as a dispensaries of prescription medication, all (N=200) community pharmacy websites made mention of this service. It can also be seen in Table 2 that the most commonly mentioned dispensing service, other than medication dispensary at the pharmacy location, tended to emphasize convenience for patients by presenting options for online ordering of prescription medication refills, multi-dosage packaging, and delivery of prescription medication to patients' home.

Disease prevention and detection services came as the second most often category of clinical services presented on the community pharmacy websites, with 79.5% (N=159/200) of the websites mentioning specific services of this type. Among these, blood pressure tests, stop smoking counseling, and flu vaccinations were the most frequently mentioned services. Clinical consultation services were presented on less but still a substantial share of 65.0% (N=130/200) community pharmacy websites. The number one specific clinical consultation service thereby was "medication therapy review and management," which is used here as an umbrella term for a recognized clinical pharmacy service aimed at optimizing therapeutic outcomes through an assessment of a patient's patterns of medication use, therapeutic responses, and potentially adverse reactions. Depending upon country, the service is known as medication use review, medication therapy management, or home medicines review [56,57]. Consultations for asthma and diabetes patients came second and third, respectively, as the most often mentioned clinical consultation services.

Table 2. Presence of clinical services on community pharmacy websites in Australia, Great Britain, Canada, and the Netherlands.

Clinical services[a]	Location of community pharmacies					
	Australia	Great Britain	Canada	Netherlands	All	*P*[b]
Dispensing services	50 (100%)	50 (100%)	50 (100%)	50 (100%)	200 (100%)	N/A
Online request for refill prescription medication[c]	0 (0%)	45 (90%)	30 (60%)	49 (98%)	124 (62.0%)	<.001
Medication multidosage packaging[d]	21 (42%)	8 (16%)	36 (72%)	26 (52%)	91 (45.5%)	<.001
Medication home delivery[e]	11 (22%)	14 (28%)	33 (66%)	28 (56%)	86 (43.0%)	<.001
Medication compounding[f]	10 (20%)	0 (0%)	27 (54%)	4 (8%)	41 (20.5%)	<.001
Emergency contraception	0 (0%)	27 (54%)	0 (0%)	0 (0%)	27 (13.5%)	<.001
Online physician for ordering prescription medication[g]	0 (0%)	23 (46%)	0 (0%)	0 (0%)	23 (11.5%)	<.001
Minor ailment service[h]	1 (2%)	13 (26%)	7 (14%)	0 (0%)	21 (11.0%)	<.001
Hormone replacement therapy[i]	2 (4%)	0 (0%)	9 (18%)	0 (0%)	11 (5.5%)	<.001
Opioid substitution[j]	4 (8%)	1 (2%)	2 (4%)	0 (0%)	7 (3.5%)	<.16
Disease prevention and detection services	41 (82%)	43 (86%)	37 (74%)	38 (76%)	159 (79.5%)	<.44
Stop smoking counseling	7 (14%)	30 (60%)	13 (26%)	9 (18%)	59 (29.5%)	<.001
Travel health advise	4 (8%)	0 (0%)	3 (6%)	20 (40%)	27 (13.5%)	<.001
Healthy eating and weight advise	10 (20%)	6 (12%)	5 (10%)	2 (4%)	23 (11.5%)	<.09
Medication use during Ramadan advise	0 (0%)	0 (0%)	0 (0%)	11 (22%)	11 (7.3%)	<.001
Diagnostic tests and checks						
blood pressure	33	15	14	0	62	
diabetes/glucose	14	9	5	14	42	
cholesterol	15	2	2	1	20	
body mass	12	0	0	0	12	
bowel cancer	6	2	0	0	8	
spot check	1	6	0	0	7	
Chlamydia/gonorrhea	4	1	1	0	6	
anticoagulation	0	2	0	0	2	
Allergy	0	2	0	0	2	
Vaccinations						
flu	15	21	22	0	58	
travel	0	14	20	0	34	

Clinical services[a]	Location of community pharmacies					
	Australia	Great Britain	Canada	Netherlands	All	P [b]
shingles	0	0	7	0	7	
Clinical consultation services	35 (70%)	36 (72%)	41 (82%)	18 (36%)	130 (65.0%)	<.001
Medication therapy review and management[k]	20 (40%)	28 (56%)	40 (78%)	2 (4%)	90 (45.0%)	<.001
Consultations for patients with specific conditions	22 (44%)	20 (40%)	20 (40%)	18 (36%)	80 (40.0%)	<.85
asthma/COPD	11	9	7	14	41	
diabetes	8	0	13	13	34	
pain	1	10	5	0	16	
sleep apnea	11	0	0	0	11	
cancer	0	4	4	0	8	
urinary incontinence	1	0	0	7	8	
post-hospital discharge	3	2	1	1	7	
hair loss	0	2	1	0	3	
osteoporosis	0	0	3	0	3	
coeliac	3	0	0	0	3	
erectile dysfunction	0	1	1	0	2	
Absence from work certificates[l]	16 (32%)	0 (0%)	0 (0%)	0 (0%)	16 (8.0%)	<.001

[a] Only clinical services that were mentioned on more than one website are included.

[b] On the basis of the chi-square test of independence χ^2_3 (N=200): N/A=nonapplicable.

[c] Patients submit an online request for a refill of a medicine that a physician prescribed to them earlier, after which the pharmacy arranges the order for the refill directly with the physician and makes the refill medication ready for the patient.

[d] The pharmacy promises to deliver medication to the patient’s home, or alternatively patients receive a code that will allow them to open a locker where their medication was deposited.

[e] A blister package (“webster pak”) or roll of sachets containing the medications a patient should take on a specific day and time, often used for patients on a multiple medication regime.

[f] The pharmacy prepares a medicine in-house that is tailored to the specific needs of an individual patient.

[g] Patients obtain prescription medication for specific conditions by filing an online questionnaire that is reviewed and then ordered by a physician hired by the pharmacy.

[h] Pharmacist advises and prescribes medicine for common, non–life threatening conditions such as eczema, headaches, or coughs.

[i] A therapy for menopause-related symptoms involving medications to artificially boost hormone levels.

[j] Substitution medication for patients with an opioid dependence, including the dispensing of methadone.

[k] Consultation to review a patient’s medication use to check for possible overlaps or interactions, identify and diminish side effects, and improve a patient’s understanding and medication adherence.

[l] Pharmacist issues a certificate as proof of legitimate absence from work, part of the 2009 Australian Fair Work Act.

Table 3. Presence of commercial services on community pharmacy websites in Australia, Great Britain, Canada, and the Netherlands.

Commercial services	Location of community pharmacies					P [a]
	Australia	Great Britain	Canada	Netherlands	All	
Health-related services	25 (50%)	31 (62%)	44 (88%)	25 (48%)	119 (59.5%)	<.001
Over-the-counter medicine[b]	30 (60%)	31 (62%)	29 (58%)	25 (50%)	115 (57.5%)	ns
Complementary and alternative medicine	24 (48%)	30 (60%)	29 (58%)	24 (48%)	107 (53.5%)	ns
Home medical aids[c]	33 (66%)	19 (38%)	24 (48%)	22 (44%)	98 (49.0%)	.03
Diet products[d]	28 (56%)	27 (54%)	17 (34%)	22 (44%)	94 (47.0%)	ns
Nonhealth-related services	32 (76%)	30 (60%)	32 (88%)	32 (64%)	126 (63.0%)	ns
Cosmetic products	32 (64%)	30 (60%)	32 (64%)	32 (64%)	126 (63.0%)	ns
Foods and beverages	0 (0%)	5 (10%)	13 (26%)	6 (12%)	24 (16.0%)	<.001
E-commerce	20 (40%)	30 (60%)	9 (18%)	32 (64%)	91 (45.5%)	<.001

[a] On the basis of the chi-square test of independence χ^2_3 (N=200): ns=nonsignificant.

[b] Products dispensing of some of the specific classes of over-the-counter drugs can be restricted to pharmacists or drug store owners only. This concerns only a very small portion of the wide range of over-the-counter medicines available.

[c] Tools to mitigate medical treatment or impairment, such as mattress covers, crutches, or shower seats. Usually sold, but sometimes also for hire, through the pharmacy.

[d] Products that replace or complement conventional foods, taken for health or cosmetic reasons, such as slimming shakes, vitamins, and gluten-free foods.

Commercial Services

The overview of the commercial services presented on the community pharmacy websites in Table 3 shows that nearly 60% of the community pharmacy websites (N=119/200, 59.5%) presented the sale of health products on their websites. Particularly the sale of over-the-counter medicine such as cold and cough medicine, as well as complementary and alternative medicine such as homeopathic products, were popular items on the community pharmacy websites. The sale of nonhealth services was even more common on the websites (N=126/200, 63.0%). Commonly mentioned in this category of services was the sale of cosmetic products such as shampoos or hand creams; however, the offer of nonhealth-related services on the pharmacy websites could also extend into a very wide range of goods or services, including public transport and lottery tickets, walking sticks, computers, or an organic juice bar inside the pharmacy. Finally, nearly half (N=91/200, 45.5%) of the pharmacy websites included e-commerce options (ie, a Web shop) for the online sale of commercial products.

Cross-National Differences

Although cross-national comparisons were not this study’s primary aim, the last columns in Tables 2 and 3 show the extent to which the mention of specific clinical and commercial services on the community pharmacy websites was contingent upon the country where the pharmacy is located, based on chi-square tests of independence. A notable finding here was that there were few differences across countries in the categories of clinical services mentioned on the community pharmacies websites. Both dispensing services and disease detection and prevention services were mentioned equally often on the community pharmacy websites across countries. On the other hand, a rather large share of the specific clinical services within these categories proved country-specific. Thus, many specific clinical services, be it medication compounding, consultations for diabetes patients, or stop smoking counseling services, were mentioned on the pharmacy websites in one country significantly more often than in other countries. Furthermore, a notable cross-national difference in the mention of commercial services on websites was the relatively low share of Canadian pharmacies that integrated e-commerce in their websites compared with the three other countries.

Table 4. Presence of services on chain pharmacy websites and single pharmacy websites.

Presence of services	Pharmacy business type		
	Chain pharmacy (N=147)	Single pharmacy (N=53)	P [a]
Clinical services			
Dispensing services	147 (100%)	53 (100%)	N/A
Clinical consultation services	101 (68.7%)	30 (57%)	ns
Disease prevention and detection services	120 (81.6%)	31 (59%)	<.001
Commercial services			
Health-related commercial services	104 (70.7%)	22 (42%)	<.001
Nonhealth-related commercial services	114 (77.6%)	13 (25%)	<.001
E-commerce	85 (57.8%)	7 (13%)	<.001

[a] On the basis of the chi-square test of independence χ^2_1 (N=200): N/A=nonapplicable, ns=nonsignificant.

Differences Between Single and Chain Pharmacies

Table 4 shows that chain pharmacy websites presented disease prevention and detection services as well as all categories of commercial services significantly more often than single pharmacy websites. Note that this does not mean that single pharmacy websites did not mention these services, as all categories of clinical services were mentioned on more than half of the single pharmacy websites. Additionally, health-related commercial services such as self-care products were present on 42% (22/53) of the single pharmacy websites in the sample compared with 70.7% (104/147) of the chain pharmacy websites.

Combinations of Clinical and Commercial Services

The analyses focusing on a combination of clinical and commercial services on community pharmacy websites showed that nearly three quarters (N=145/200, 72.5%) of the community pharmacy websites were a mix of clinical and commercial services. The remaining pharmacy websites presented clinical services only (N=55/200, 27.5%). This confirms that having a combination of clinical and commercial services on the website of licensed community pharmacies is widespread. Although the share of pharmacy websites containing a combination of clinical and commercial services was at least 62% (31/50) in each country, a chi-square test of independence further showed that combinations were not equally common across the 4 countries, χ^2_3 (N=200)=10.91, P=.01. They were most common on the Canadian community pharmacy websites (N=44/50, 88%), followed by the Australian (N=38/50, 76%), British (N=32/50, 64%), and finally the Dutch (N=31/50, 62%) community pharmacy websites. Moreover, whereas more than 4 in every 5 websites from chain pharmacies contained a combination of clinical and commercial services (N=120/147, 82.8%), slightly less than half of the single pharmacy websites did (N=25/53, 47%), resulting in a significant chi-square relationship (χ^2_1 (N=200)=24.12, P<.001). A combination of clinical and commercial services on community pharmacy websites was thus more commonly seen on the websites of pharmacies operating under a chain brand than single pharmacies.

Table 5 renders phi coefficients for the degree of association between the presence of specific categories of clinical services and specific categories of commercial services on the community pharmacy websites. It can be seen that specifically the mention of disease prevention and detection services (eg, stop smoking counseling, flu vaccinations, diabetes tests) was significantly associated with the mention of commercial services on the websites. The mention of dispensing services, on the other hand, was unrelated to the mention of commercial services on the websites, whereas there was a significant negative association with the presence of e-commerce services.

Table 5. Phi coefficients for the presence of categories of clinical and categories of commercial services on the community pharmacy websites (Φ, 200).

Clinical services	Commercial services			
	Health-related	Nonhealth-related	E-commerce	All commercial services
Dispensing services	.09	.01	−.17[a]	.00
Consultation services	.19[b]	.00	−.15[a]	.09
Disease prevention and detection services	.14[a]	.26[b]	.11	.26[b]

[a] P<.05.

[b] P<.01.

Menu and Images

Homepage Menu

One of the main elements of community pharmacy websites is the homepage menu that allows users to navigate to specific services on the website. On nearly half of the pharmacy websites' homepages, namely 99/200 homepages (49.5%), the homepage menu was a combination of clinical and commercial service items. The occurrence of this combination of clinical and commercial services on the homepage menus was further contingent on country and occurred most often on the Canadian websites (37/50, 74%), followed by the British (N=30/50, 60%), Australian (N=24/50, 48%) and finally the Dutch (N=6/50, 12%) websites, χ^2_3 (N=200)=42.71, *P*<.001. Also, a combination of clinical and commercial homepage menu items were significantly more common on chain pharmacy websites (N=82/147, 55.8%) than single pharmacy websites (N=15/53, 28%), χ^2_1 (N=200)=11.75, *P*<.001.

Homepage Images

Another way of combining clinical and commercial content on the homepages of the community pharmacies' websites is through the inclusion of clinical and commercial images on the homepage. More than half of the homepages of community pharmacy websites contained a combination of clinical and commercial images (N=107/200, 53.5%). A combination of clinical and commercial images occurred most often on the homepages of the British pharmacy websites (N=40/50, 80%), followed by the Australian (N=25/50, 50%), Canadian (N=24/50, 48%), and finally the Dutch (N=24/50, 48%) homepages, χ^2_3 (N=200)=21.12, *P*<.001). Again, it was significantly more common on the homepages of chain pharmacy websites (N=97/147, 66.0%) than on the homepages of single pharmacy websites (N=10/53, 19%), χ^2_1 (N=200)=35.40, *P*<.001).

Discussion

Principal Findings

Community pharmacy websites are playing an increasingly important role in clinical health care nowadays as they have moved from a static reflection of local premises for dispensing medication to an online "health care hub." Here, patient consumers can submit electronic prescriptions, order prescription and over-the-counter medication, consult pharmacists on the management of a disease, download software to monitor one's health, and order vaccinations, personal care, and household products, among other things. Pharmacy licensing however lags behind these developments and often does not include specific regulations about a combination of clinical and commercial services on pharmacy websites. The present study is situated amidst these developments, which may be a significant "social force" as online retail continues to grow but has not received much attention in the literature so far in the context of pharmacy websites [47-50]. This study addressed the pertinent question as to what types of clinical and commercial services are mentioned on present-day licensed community pharmacy websites and to what extent the websites currently present a combination of clinical and commercial services. Such a combination is regarded as highly controversial in the clinical professions and tends to frequently raise calls for codes of conduct, regulatory action, or straightforward bans to protect vulnerable patient groups from the undue influence on their preferences for treatment [1-3].

Through a content analysis of representative samples of licensed community pharmacy websites in Great Britain, the Netherlands, Australia, and Canada, it was found that around three quarters of the websites of licensed community pharmacies currently present a combination of clinical and commercial services. Particularly disease prevention and detection services such as stop smoking counseling and travel health advice were often combined with the sale of commercial services on the websites. This shows, first, that a controversial combination of clinical and commercial services is a common feature of licensed community pharmacy websites nowadays. Furthermore, particularly the inclusion of disease prevention and detection services on the pharmacy websites signals an appeal to a broad group of customers to maintain their health through a combination of timely disease detection and prevention and purchasing health and wellness products such as over-the-counter medicine, diet shakes, and aromatherapy oils.

Up to three quarters of community pharmacies in our sample operate under a chain pharmacy's retail brand name, and the findings showed that a combination of clinical and commercial services is particularly common on chain pharmacies' websites, although also nearly half of the single pharmacy websites included a combination of clinical and commercial services. One explanation is that commercial content for chain pharmacy websites is created in the head offices, whereas single pharmacy owners may not have similar resources. Another explanation may be that single pharmacy owners attempt to avoid the controversial combination of clinical and commercial contents on their websites more than chains in an attempt to emphasize adherence to the trade's traditional values and present as different from the chain pharmacy competition in this respect. Single community pharmacies have been found to prioritize the profession's fiduciary role more than chain pharmacies [16,58].

A final notable finding from this study is that there were few differences across countries in the categories of clinical services mentioned on the community pharmacies websites, and both dispensing services and disease detection and prevention services were mentioned equally often on the community pharmacy websites across countries. This suggests that community pharmacy websites generally include the same scope of services across the four high-income countries. However, there was substantial variation within, as well as across, countries in regard to specific clinical services mentioned on the community pharmacy websites. This can in part be explained by the specific national context. For instance, online pharmacies that are (actually or supposedly) "Canadian" obtained a negative reputation in recent years for allegations of illegitimate e-commerce practices [59]. It appears that licensed community pharmacies from Canada therefore more often did not include e-commerce options than other countries. Other cross-national differences can be explained by national pharmacy legislation. For instance, receiving a prescription from an online physician through a community pharmacy website occurred only on British pharmacy websites because it is prohibited in the three other countries in our sample [60], while vaccinations through pharmacies, for instance, are not allowed in the Netherlands [61]. The larger share of the cross-national variation however cannot be explained by nation-specific context or legislation, suggesting that the community pharmacies' offerings are not only shaped by clinical considerations, but also by other factors. Possible factors include insurer-based remuneration programs [62], as well as government- and industry-sponsored health and disease campaigns [1,63].

Limitations

Regarding the generalizability of the findings, it is important to note that the present study endeavored to gain insight into the combination of clinical and commercial communication on present-day community pharmacy websites. No claim regarding a one-to-one relationship with the actual services offered at a pharmacy's premises is made, as this falls outside this study's aims. It is of further relevance to note that the community pharmacy websites in our sample likely render a somewhat conservative picture of the combination of clinical and commercial content on community pharmacy websites worldwide. First, because the focus of this study was on websites of licensed community pharmacies, rather than the more heavily studied phenomenon of online pharmacies involved in illicit practices [10-12]. Second, regulations surrounding commercial communication in health care are typically enforced more heavily in the countries in our sample than in middle- or low-income countries [53]. Third and lastly, mass merchants and supermarkets with pharmacy departments were excluded from the present sample since these websites would not allow a clear separation between services mentioned as part of pharmacy departments, versus other store departments. The websites of some of the largest mass merchants in the world such as Wal-Mart, Tesco, and Costco frequently have pharmacy sections as well, which leads to the combination of clinical services with a vast range of commercial services. Although caution is warranted due to a lack of empirical data, the combination of clinical and commercial communication found in this study likely forms a lower bound, rather than an upper bound, of those found on community pharmacy websites across the board.

Implications

This paper has drawn attention to the websites of licensed community pharmacies as one of the major online venues in health care where a combination of clinical and commercial services is common practice. One implication is a further recognition in the literature on online DTCA and medical marketing that this is an important area of health care communication on the Internet nowadays. Effects on online patients' reactions have yet to be studied but, as mentioned in the introductory section, patients often see online health information as more credible and persuasive when it comes from a source that is perceived as having relevant expertise [30-33]. With the increasingly expanding role of community pharmacies in clinical health care provision, this suggests for policy makers and regulatory agencies that the combination of clinical and commercial services on community pharmacy websites is able to impact patients' beliefs and preferences. It should thereby be noted that, likening to contemporary advertising formats such as sponsored content and native advertising [64,65], the distinction between services restricted to licensed pharmacists versus commercial services is nearly never clearly marked on community pharmacy websites. Due to this, patients cannot immediately tell if a service or product, such as a magnetic therapy service or insect repellent, featured on a community pharmacy website is clinical or commercial. The extant research shows that people are rarely motivated, or able, to distinguish commercial content from editorial content in the absence of clear disclosures [65,66]. All in all, community pharmacy websites clearly are no longer just an online version of a medication dispensing facility alone and should be regarded as a significant party in the controversy over combinations of clinical and commercial communication on the Internet nowadays.

Conflicts of Interest

None declared.

References

1. Moynihan R, Heath I, Henry D. Selling sickness: the pharmaceutical industry and disease mongering. BMJ 2002 Apr 13;324(7342):886-891 [FREE Full text] [Medline: 11950740]
2. Kemp E, Min KS, Joint E. Selling hope: the role of affect-laden health care advertising in consumer decision making. Journal of Marketing Theory and Practice 2015 Jul 24;23(4):434-454. [doi: 10.1080/10696679.2015.1049688]
3. Schenker Y, Arnold RM, London AJ. The ethics of advertising for health care services. Am J Bioeth 2014;14(3):34-43. [doi: 10.1080/15265161.2013.879943] [Medline: 24592839]
4. Mazzeo SE, Trace SE, Mitchell KS, Gow RW. Effects of a reality TV cosmetic surgery makeover program on eating disordered attitudes and behaviors. Eat Behav 2007 Aug;8(3):390-397. [doi: 10.1016/j.eatbeh.2006.11.016] [Medline: 17606237]
5. Vittoria A, Paino G, La ME, Budetta G, Cecio A. Chromogranin- and somatostatin-containing neuroendocrine cells in the porcine uterus. An immunocytochemical study. Anat Histol Embryol 1989 Jun;18(2):136-142. [Medline: 2569277]
6. Horowitz MD, Rosensweig JA, Jones CA. Medical tourism: globalization of the healthcare marketplace. MedGenMed 2007 Nov 13;9(4):33. [Medline: 18311383]

7. Turner L. ‘First world health care at third world prices’: globalization, bioethics and medical tourism. BioSocieties 2007 Sep;2(3):303-325. [doi: 10.1017/S1745855207005765]
8. Jin LX, Ibrahim AM, Newman NA, Makarov DV, Pronovost PJ, Makary MA. Robotic surgery claims on United States hospital websites. J Healthc Qual 2011 Nov;33(6):48-52. [doi: 10.1111/j.1945-1474.2011.00148.x] [Medline: 22059902]
9. Schiavone MB, Kuo EC, Naumann RW, Burke WM, Lewin SN, Neugut AI, et al. The commercialization of robotic surgery: unsubstantiated marketing of gynecologic surgery by hospitals. Am J Obstet Gynecol 2012 Sep;207(3):174.e1-174.e7. [doi: 10.1016/j.ajog.2012.06.050] [Medline: 22835493]
10. Fittler A, Bősze G, Botz L. Evaluating aspects of online medication safety in long-term follow-up of 136 internet pharmacies: illegal rogue online pharmacies flourish and are long-lived. J Med Internet Res 2013 Sep 10;15(9):e199 [FREE Full text] [doi: 10.2196/jmir.2606] [Medline: 24021777]
11. Katsuki T, Mackey TK, Cuomo R. Establishing a link between prescription drug abuse and illicit online pharmacies: analysis of twitter data. J Med Internet Res 2015 Dec 16;17(12):e280 [FREE Full text] [doi: 10.2196/jmir.5144] [Medline: 26677966]
12. Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res 2011 Sep 30;13(3):e74 [FREE Full text] [doi: 10.2196/jmir.1795] [Medline: 21965220]
13. Brown D. The paradox of pharmacy: a profession's house divided. J Am Pharm Assoc (2003) 2012;52(6):e139-e143. [doi: 10.1331/JAPhA.2012.11275] [Medline: 23143300]
14. Bush J, Langley CA, Wilson KA. The corporatization of community pharmacy: implications for service provision, the public health function, and pharmacy's claims to professional status in the United Kingdom. Res Social Adm Pharm 2009 Dec;5(4):305-318. [doi: 10.1016/j.sapharm.2009.01.003] [Medline: 19962674]
15. Roland M, Dixon M. Randomized controlled trial of an educational booklet for patients presenting with back pain in general practice. J R Coll Gen Pract 1989 Jun;39(323):244-246 [FREE Full text] [Medline: 2556518]
16. Perepelkin J, Dobson RT. Influence of ownership type on role orientation, role affinity, and role conflict among community pharmacy managers and owners in Canada. Res Social Adm Pharm 2010 Dec;6(4):280-292. [doi: 10.1016/j.sapharm.2009.11.001] [Medline: 21111386]
17. Sleath B, Campbell W. American pharmacy: a profession in the final stage of dividing? J Pharm Mark Manage 2011 Dec 04;14(1):1-25. [doi: 10.3109/J058v14n01_01]
18. Kwoka JE. The Federal Trace Commission and the professions: a quarter century of accomplishment and some new challenges. SSRN Journal 2004:- American Antitrust Institute Working Paper No. 04-04 [FREE Full text] [doi: 10.2139/ssrn.1103513]
19. Fishman RH. New Israeli law cancels ban on physicians advertising. The Lancet 2000 May;355(9218):1895. [doi: 10.1016/S0140-6736(05)73345-6]
20. Hite RE, Fraser C. Meta-analyses of attitudes toward advertising by professionals. J Marketing 1988 Jul;52(3):95. [doi: 10.2307/1251453]
21. Tomycz ND. A profession selling out: lamenting the paradigm shift in physician advertising. J Med Ethics 2006 Jan;32(1):26-28 [FREE Full text] [doi: 10.1136/jme.2005.012617] [Medline: 16373520]
22. Zwier S. The neglected repercussions of a physician advertising ban. J Med Ethics 2014 Mar;40(3):198-201. [doi: 10.1136/medethics-2012-101078] [Medline: 23557909]
23. Devine T, Broderick J, Harris LM, Wu H, Hilfiker SW. Making quality health websites a national public health priority: toward quality standards. J Med Internet Res 2016;18(8):e211 [FREE Full text] [doi: 10.2196/jmir.5999] [Medline: 27485512]
24. Risk A, Dzenowagis J. Review of internet health information quality initiatives. J Med Internet Res 2001;3(4):E28 [FREE Full text] [doi: 10.2196/jmir.3.4.e28] [Medline: 11772543]
25. Winker MA, Flanagin A, Chi-Lum B, White J, Andrews K, Kennett RL, et al. Guidelines for medical and health information sites on the internet: principles governing AMA web sites. American Medical Association. JAMA 2000;283(12):1600-1606. [Medline: 10735398]
26. Cassel CK. The patient-physician covenant: an affirmation of Asklepios. Conn Med 1996 May;60(5):291-293. [Medline: 8998907]
27. ABIM Foundation. American Board of Internal Medicine, ACP-ASIM Foundation. American College of Physicians-American Society of Internal Medicine, European Federation of Internal Medicine. Medical professionalism in the new millennium: a physician charter. Ann Intern Med 2002;136(3):243-246. [Medline: 11827500]
28. Hovland CI, Weiss W. The influence of source credibility on communication effectiveness. Public Opin Q 1951 Jan 01;15(4):635-650 [FREE Full text] [doi: 10.1086/266350]
29. Pornpitakpan C. The persuasiveness of source credibility: a critical review of five decades' evidence. J Appl Social Pyschol 2004 Feb;34(2):243-281. [doi: 10.1111/j.1559-1816.2004.tb02547.x]
30. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the internet. Med Inform Internet Med 2006 Mar;31(1):45-52. [doi: 10.1080/14639230600552601] [Medline: 16754366]
31. Eastin M. 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]

32. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Comm Res 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
33. Yang Q, Beatty M. A meta-analytic review of health information credibility: belief in physicians or belief in peers? HIM J 2016 Aug;45(2):80-89. [doi: 10.1177/1833358316639432] [Medline: 27105476]
34. Yi M, Yoon J, Davis J, Lee T. Untangling the antecedents of initial trust in Web-based health information: the roles of argument quality, source expertise, and user perceptions of information quality and risk. Decision Support Systems 2013 Apr;55(1):284-295. [doi: 10.1016/j.dss.2013.01.029]
35. Quigley G. i2p. 2016. From community pharmacy to commodity pharmacy? URL: http://i2p.com.au/community-pharmacy-or-commodity-pharmacy/ [accessed 2016-10-26] [WebCite Cache ID 6lXxdFg06]
36. Mulupi D. How we made it in Africa. 2016. The man who built Kenya's Mimosa pharmacy chain URL: http://www.howwemadeitinafrica.com/the-man-who-built-kenyas-mimosa-pharmacy-chain/ [accessed 2016-10-26] [WebCite Cache ID 6lXy7njTX]
37. Peru Reports. 2016. Pharmacy chain buys competitor in Peru's fragmented industry URL: http://perureports.com/2016/03/23/pharmacy-chain-buys-competitor-in-perus-fragmented-industry/ [accessed 2016-10-26] [WebCite Cache ID 6lY5LKfYp]
38. Perrottet T. Slate. 2016 Oct 26. The Pope's supermarket, and other Vatican curiosities URL: http://www.slate.com/articles/life/welltraveled/features/2011/vatican_inside_the_secret_city/vatican_guide_inside_the_pope_s_supermarket.html [accessed 2016-10-26] [WebCite Cache ID 6lY5wOfAy]
39. Anderson S. Community pharmacy and public health in Great Britain, 1936 to 2006: how a phoenix rose from the ashes. J Epidemiol Community Health 2007 Oct;61(10):844-848 [FREE Full text] [doi: 10.1136/jech.2006.055442] [Medline: 17873218]
40. McKinlay JB, Marceau LD. The end of the golden age of doctoring. Int J Health Serv 2002;32(2):379-416. [Medline: 12067037]
41. Alston G. Multibriefs. 2016. Is clinical pharmacy really the future of our profession? URL: http://exclusive.multibriefs.com/content/is-clinical-pharmacy-really-the-future-of-our-profession/pharmaceutical [accessed 2016-10-26] [WebCite Cache ID 6lY6ciaxG]
42. Kieć EF, Stepniewski M, Kolarzyk E, Targosz D, Lagan W. Comparison of airway resistance of inhabitants of two distinctly different polluted areas. Pol J Occup Med 1990;3(3):311-321. [Medline: 2136538]
43. Farris KB, Fernandez-Llimos F, Benrimoj SI. Pharmaceutical care in community pharmacies: practice and research from around the world. Ann Pharmacother 2005 Sep;39(9):1539-1541. [doi: 10.1345/aph.1G049] [Medline: 16014373]
44. Fang Y, Yang S, Zhou S, Jiang M, Liu J. Community pharmacy practice in China: past, present and future. Int J Clin Pharm 2013 Aug;35(4):520-528. [doi: 10.1007/s11096-013-9789-5] [Medline: 23661172]
45. Bidwell S, Thompson L. GPs, community pharmacists and shifting professional boundaries. N Z Med J 2015 May 15;128(1414):19-26. [Medline: 26117387]
46. Hersberger KE, Messerli M. Development of clinical pharmacy in Switzerland: involvement of community pharmacists in care for older patients. Drugs Aging 2016 Mar;33(3):205-211. [doi: 10.1007/s40266-016-0353-6] [Medline: 26884391]
47. Adamcik BA, Ransford HE, Oppenheimer PR, Brown JF, Eagan PA, Weissman FG. New clinical roles for pharmacists: a study of role expansion. Soc Sci Med 1986;23(11):1187-1200. [Medline: 3810205]
48. Benrimoj SI, Frommer MS. Community pharmacy in Australia. Aust Health Rev 2004 Nov 8;28(2):238-246. [Medline: 15527404]
49. Doucette WR, Kreling DH, Schommer JC, Gaither CA, Mott DA, Pedersen CA. Evaluation of community pharmacy service mix: evidence from the 2004 National Pharmacist Workforce study. J Am Pharm Assoc 2006 May;46(3):348-355. [doi: 10.1331/154434506777069471]
50. Edmunds J, Calnan MW. The reprofessionalisation of community pharmacy? an exploration of attitudes to extended roles for community pharmacists amongst pharmacists and General Practioners in the United Kingdom. Soc Sci Med 2001 Oct;53(7):943-955. [Medline: 11522139]
51. Zehnder S, Bruppacher R, Ruppanner H, Hersberger KE. Swiss community pharmacies' on the Web and pharmacists' experiences with e-commerce: longitudinal study and Internet-based questionnaire survey. J Med Internet Res 2004 Mar 03;6(1):e9 [FREE Full text] [doi: 10.2196/jmir.6.1.e9] [Medline: 15111275]
52. Yegenoglu S, Sozen B, Aslan D, Calgan Z, Cagirci S. An evaluation of the quality of Turkish community pharmacy web sites concerning HON principles. Telemed J E Health 2008 May;14(4):375-380. [doi: 10.1089/tmj.2007.0065] [Medline: 18570568]
53. Lowe RF, Montagu D. Legislation, regulation, and consolidation in the retail pharmacy sector in low-income countries. Southern Med Review 2009;2(2):35-44 [FREE Full text]
54. American Pharmacist Association. Pharmacist. 2016. Chain community pharmacy: Staff URL: https://www.pharmacist.com/sites/default/files/files/Profile_06_chain_pharmacy-staff.pdf [accessed 2016-10-26] [WebCite Cache ID 6lY7YljRd]
55. International Diabetes Federation. International Diabetes Federation. 2015. DF Diabetes Atlas, 7th edition URL: http://www.diabetesatlas.org/ [accessed 2017-05-22] [WebCite Cache ID 6qeFj4dgz]

56. American Pharmacist Association, National Association of Chain Drug Stores Foundation. Pharmacist. Medication therapy management in pharmacy practice: core elements of an MTM service model URL: http://www.pharmacist.com/sites/default/files/files/core_elements_of_an_mtm_practice.pdf [accessed 2016-10-26] [WebCite Cache ID 6lY8JCyfu]
57. Viswanathan M, Kahwati LC, Golin CE, Blalock SJ, Coker-Schwimmer E, Posey R, et al. Medication therapy management interventions in outpatient settings: a systematic review and meta-analysis. JAMA Intern Med 2015 Jan;175(1):76-87. [doi: 10.1001/jamainternmed.2014.5841] [Medline: 25401788]
58. Perepelkin J, Di Zhang D. Brand personality and customer trust in community pharmacies. Intl J of Pharm & Health Mrkt 2011 Sep 06;5(3):175-193. [doi: 10.1108/17506121111172194]
59. LegitScript Folks. LegitScript. Why “Canadian” internet pharmacies aren’t what they seem URL: https://www.legitscript.com/blog/2009/08/but-its-a-canadian-internet-pharmacy-telling-the-truth/ [accessed 2016-10-26] [WebCite Cache ID 6lY9KO9u9]
60. Fung CH, Woo HE, Asch SM. Controversies and legal issues of prescribing and dispensing medications using the Internet. Mayo Clin Proc 2004 Feb;79(2):188-194. [doi: 10.4065/79.2.188] [Medline: 14959914]
61. European Pharmacist Forum. The role of pharmacy in supporting the public's health: an EPF white paper and call to action URL: http://www.europeanpharmacistsforum.com/documents/10429/a18fce74-71e3-4d93-b68d-7377023e3956 [accessed 2016-10-26] [WebCite Cache ID 6lYA02ZUP]
62. Houle SK, Grindrod KA, Chatterley T, Tsuyuki RT. Paying pharmacists for patient care: a systematic review of remunerated pharmacy clinical care services. Can Pharm J (Ott) 2014 Jul;147(4):209-232 [FREE Full text] [doi: 10.1177/1715163514536678] [Medline: 25360148]
63. Cleland R, Finskud L, Raimondi V. Nine prescriptions for brand health. Market Leader 2004(Spring):50-55 [FREE Full text]
64. van Reijmersdal E, Neijens P, Smit EG. A new branch of advertising. JAR 2009 Dec 01;49(4):429-449. [doi: 10.2501/S0021849909091065]
65. Wojdynski BW. The deceptiveness of sponsored news articles: how readers recognize and perceive native advertising. Am Behav Sci 2016 Jul 28;60(12):1475-1491. [doi: 10.1177/0002764216660140]
66. de Pelsmacker P, Neijens PC. New advertising formats: how persuasion knowledge affects consumer responses. Journal of Marketing Communications 2012 Feb;18(1):1-4. [doi: 10.1080/13527266.2011.620762]

Abbreviations

COPD: chronic obstructive pulmonary disease
DTCA: direct-to-consumer advertising
FTC: American Federal Trade Commission

Edited by G Eysenbach; submitted 30.10.16; peer-reviewed by J Perepelkin, P Schulz, K Grindrod, Q Yang; comments to author 15.01.17; revised version received 23.02.17; accepted 30.03.17; published 14.06.17.

Please cite as:
Zwier S
“Click for Closer Care”: A Content Analysis of Community Pharmacy Websites in Four Countries
J Med Internet Res 2017;19(6):e205
URL: http://www.jmir.org/2017/6/e205/
doi:10.2196/jmir.6899
PMID:28615153

Original Paper

Designing, Prototyping and Evaluating Digital Mindfulness Applications: A Case Study of Mindful Breathing for Stress Reduction

Bin Zhu[1,2]; Anders Hedman[1], PhD; Shuo Feng[1]; Haibo Li[1], PhD; Walter Osika[3], MD, PhD

[1]School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden

[2]School of Film and Animation, China Academy of Art, Hangzhou, China

[3]Department of Clinical neuroscience, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Bin Zhu
School of Computer Science and Communication
KTH Royal Institute of Technology
Lindstedtsvägen 3
Stockholm, 10044
Sweden
Phone: 46 723068282
Fax: 46 87909099
Email: binz@kth.se

Abstract

Background: During the past decade, there has been a rapid increase of interactive apps designed for health and well-being. Yet, little research has been published on developing frameworks for design and evaluation of digital mindfulness facilitating technologies. Moreover, many existing digital mindfulness applications are purely software based. There is room for further exploration and assessment of designs that make more use of physical qualities of artifacts.

Objective: The study aimed to develop and test a new physical digital mindfulness prototype designed for stress reduction.

Methods: In this case study, we designed, developed, and evaluated HU, a physical digital mindfulness prototype designed for stress reduction. In the first phase, we used vapor and light to support mindful breathing and invited 25 participants through snowball sampling to test HU. In the second phase, we added sonification. We deployed a package of probes such as photos, diaries, and cards to collect data from users who explored HU in their homes. Thereafter, we evaluated our installation using both self-assessed stress levels and heart rate (HR) and heart rate variability (HRV) measures in a pilot study, in order to measure stress resilience effects. After the experiment, we performed a semistructured interview to reflect on HU and investigate the design of digital mindfulness apps for stress reduction.

Results: The results of the first phase showed that 22 of 25 participants (88%) claimed vapor and light could be effective ways of promoting mindful breathing. Vapor could potentially support mindful breathing better than light (especially for mindfulness beginners). In addition, a majority of the participants mentioned sound as an alternative medium. In the second phase, we found that participants thought that HU could work well for stress reduction. We compared the effect of silent HU (using light and vapor without sound) and sonified HU on 5 participants. Subjective stress levels were statistically improved with both silent and sonified HU. The mean value of HR using silent HU was significantly lower than resting baseline and sonified HU. The mean value of root mean square of differences (RMSSD) using silent HU was significantly higher than resting baseline. We found that the differences between our objective and subjective assessments were intriguing and prompted us to investigate them further.

Conclusions: Our evaluation of HU indicated that HU could facilitate relaxed breathing and stress reduction. There was a difference in outcome between the physiological measures of stress and the subjective reports of stress, as well as a large intervariability among study participants. Our conclusion is that the use of stress reduction tools should be customized and that the design work of mindfulness technology for stress reduction is a complex process, which requires cooperation of designers, HCI (Human-Computer Interaction) experts and clinicians.

(J Med Internet Res 2017;19(6):e197) doi:10.2196/jmir.6955

KEYWORDS

respiration; biofeedback; mindfulness; stress; device design; sound; light; breathing; heart rate; relaxation

Introduction

High levels of stress and information overload are increasingly contributing to the global burden of disease [1,2] at great societal costs. For example, in Europe, the total costs of mental health disorders are estimated to be €240 billion per year [3], and globally the costs are most certainly much higher. There is therefore an urgent need for both preventive interventions aimed at reducing the adverse effects of stress, taking into account biological, psychosocial, and environmental risk factors, as well as new effective treatment strategies.

The human stress response is partially mediated through the autonomic nervous system (ANS), divided into the sympathetic and parasympathetic systems. The majority of our internal organs are, directly or indirectly, controlled through these systems. The sympathetic system prepares us for dangerous situations and triggers the stress response. The parasympathetic system governs rest and digestion [4]. High levels of stress and prolonged stress without recuperation result in cognitive, emotional, and somatic symptoms [5-7].

Western stress reduction methods have lately been CBT-oriented (building on insights from Cognitive Behavioral Therapy), with mechanistic approaches, and have emphasized behavioral changes. Eastern meditative traditions have in contrast focused more on present-centered attention and awareness, that is, what has been termed mindfulness in the West [8,9], and non-judgmental acceptance of encountered life situations. Practicing compassion and selflessness in search of balance and resilience, appreciation of life and nature in a fluid and dynamic state of constant imbalance are also major topics of eastern meditative traditions such as Buddhism [10]. Balance and resilience can be achieved through the self-healing capacity of individuals particularly when confronting stress and adversity [11]. Eastern meditative approaches have been deployed for stress reduction in westernized contexts, for example, through the practice of tai-chi [12,13], Qigong [14,15], and mindfulness [16,17].

Drawing on eastern meditative traditions, mindfulness-based stress reduction (MBSR) [18] has been proven effective for reducing stress and anxiety, and is currently one of the most researched treatment packages [19-25]. MBSR is based on procedures to establish increased awareness of moment-to-moment experience and develop compassionate non-judgmental acceptance of oneself, others, and encountered life situations. With digital technologies potentially revolutionizing health and well-being, people increasingly turn to technology to handle stress, anxiety, social isolation, and negative emotions [26-28]. Conventional mindfulness practice is moving to digital devices that potentially can support people's needs in the digital age [29,30]. New partnerships among psychologists, social scientists, designers, and engineers are needed to better understand the psychological and behavioral impact of these new technologies and applications [31].

In this paper, we focus on designing digital technology for MBSR, in particular, mindful breathing. Mindful breathing can be defined as calm and conscious deep breathing and is a cornerstone of MBSR [32]. We are often unaware of how we breathe and when stressed we breathe in shallow, tensed ways that furthers the stress. "Take a deep breath" is one of the first things we are likely to say to someone deeply anxious. We intuitively know that calm and deep (abdominal) breathing does us good. Can we support such healthy breathing patterns through technology? This is one of the questions we explore in this paper.

Researchers have developed technologies to support relaxation breathing. Wollmann et al developed an android flight game using resonant frequency breathing to increase heart rate variability (HRV) and improve ANS balance [33]. Khut and Muller [34] designed a biofeedback-based multisensory presentation of heart and breathing rates to support self-awareness in more direct ways than traditional quantified-self-monitoring techniques provide. Khut also developed a biofeedback-based stress-relief app called BrightHearts [35,36]. Heart rate and breathing sensors were used to visualize interactive halos and sounds. Colors, diameters, and saturation of the halos vary according to breathing and heart rate. When one's heart rate decreases, the color changes from red to green and the pitch decreases. Researchers from Stanford University found visual feedback to be effective for supporting healthy deep breathing and that auditory feedback can further subjective feelings of calm [37]. Others found that calming nature sounds were an effective way of reducing stress in an emergency department. Negative affect scores were lower whereas positive affect scores were the same or higher for participants of a nature-sound intervention group [38] on an affect test in comparison with a control group. Vidyarthi et al [39] created a meditative display by connecting respiration to a peaceful soundscape and developed a psychological framework for media immersion. Furthermore, physical and tangible artifacts have been used with digital technologies to capture and represent respiration and bodily interaction. Influenced by somaesthetics [40], Höök and her colleagues translated experiential insights from long-term practice of Feldenkrais exercises into a design prototype called Breathing Light [41], an enclosed space for reflection utilizing body movements. Sensors measure chest movements that control a lamp so the light changes according to breathing patterns. Their somaesthetic design engaged people by increasing their awareness of internal bodily sensations rather than external processes.

Inspired by the previous studies, we have the hypothesis that interactive physical artifacts based on visual and sound feedback can support mindful breathing for stress reduction. To explore how interactive visual and sound feedback could be used to support mindful and relaxing deep breathing, we designed, developed, and evaluated HU (Figure 1), a physical digital mindfulness prototype based on vapor, light, and sound. HU is a sonified vapor diffuser that expels vapor and emits sounds according to systematic relations with a meditator's heart rate.

We think of these relations as isomorphic and tunable. They are relations among heart rate, vapor expulsions and sounds, subject to experimentation and tuning through user feedback. As a device for supporting mindful breathing, HU could come to function as a part of MBSR interventions.

The paper is structured as follows: First, we look at stress and stress management from scientific and contemplative perspectives, focusing on MBSR. Second, we account for how we developed HU to support mindful breathing for well-being. Third, we describe how users experience HU, as well as how using HU affects them physiologically and psychologically with respect to stress resilience. Finally, we discuss our conclusions and possible directions for future work.

Figure 1. HU: a case study of mindful breathing for stress reduction.

Methods

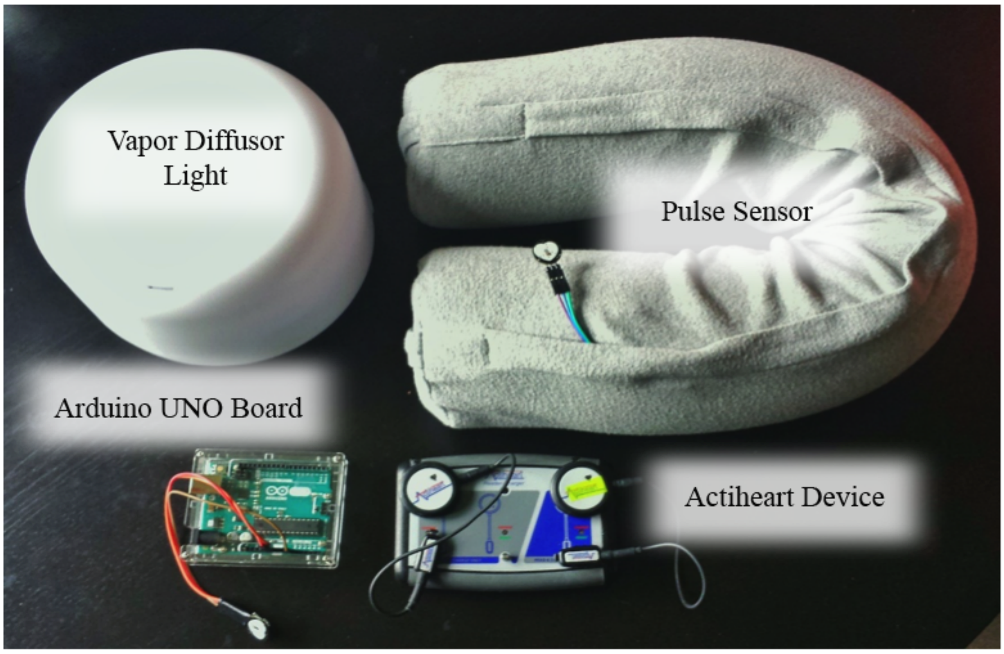
Study Design

We applied the Research through Design (RtD) method [42] in our design process of HU. In an RtD research process, we drive the exploration of both problems and solutions. We gain new knowledge through the act of making things and iterative design work. The knowledge gained during the process can be expressed and articulated in design methods and principles. Our design process consisted of two phases. In the first, initial design phase, we used vapor and light to support mindful breathing. In the second phase, we added sonification.

Initial Design Phase: The Prototype

The design of HU is inspired by the ancient Chinese incense furnace. In Chinese, HU means to breathe out. HU is a vapor diffuser modified with biosensors. HU measures the heart rate with an earlobe sensor (Figure 1) and supports resonant frequency breathing through vapor pulses. Resonant frequency breathing is a specific biofeedback training strategy based on resonance properties of the cardiovascular system [43,44]. Resonant frequency breathing can be effective in furthering relaxation and homeostatic balance [43]. HU expels vapor to support resonant frequency at 0.1 Hz (6 breaths/minute) [43,45]. The light of HU is adapted with the meditator's heart rate through the brightness. Fast heart rates are bright and slow heart rates dim.

Figure 2. HU prototype.



Initial Design Phase: Evaluation

We investigated how HU could support mindful deep breathing through a meditation intervention. We selected 25 participants through snowball sampling [46], of which 12 were males and 13 were females. The ages ranged between 23 and 60. They were originally from Sweden (10), China (7), Spain (2), the United States (2), the United Kingdom (1), Italy (1), and Australia (1), and all lived in Sweden. They represented a range of occupations including teachers, students, engineers, doctors, office workers, and mindfulness coaches. Among the 25 participants, 15 lacked the experience of mindfulness or meditative practice; 5 were beginners; and the remaining 5 were experienced mindfulness meditators. We arranged the HU in a quiet room and asked participants to try it for 10-20 min, and then interviewed them. The participants were interested in HU and enjoyed engaging with HU to support mindful breathing. About 22 of the 25 participants (88%) claimed that vapor can lead to mindful breathing and helped them relax. The meditators regarded the vapor as a good component for 4 reasons:

1. Rhythmically tuned vapor pulses can support synchronized deep breathing.

2. Vapor is aesthetically pleasing.

3. Fleeting vapor reflects the fleeting present.

4. Vapor is subtle and non-distractive.

Three meditators, however, thought vapor failed to support mindful breathing. Two of them lacked experience with mindfulness and claimed it was difficult to follow the vapor while focusing on their breath. One of them was a mindfulness coach who had practiced mindfulness for 10 years. She focused better on her "inner being" without the vapor.

Fifteen participants saw the light as a good way to enhance the experience of interaction. Light was regarded as a good medium mainly for three reasons:

1. It is visually pleasing and evokes enjoyment.

2. It enriches the breathing practice.

3. It supports body awareness.

Five participants disliked the light. They felt the connection between the light and heart rate was ambiguous and unclear. Three of them felt the light (in particular, the brightness indicating high heart rate) was stressful. Two mentioned that the rapidly changing light distracted them.

The results from the trial suggested that vapor and light can be effective ways of promoting mindful breathing, but that vapor could potentially support mindful breathing better than light (especially for mindfulness beginners). Finally, we found that, interestingly, a majority of the participants mentioned sound as an alternative medium for mindful breathing. Further to this, 13 participants closed their eyes during the trial, which indicated why sound might seem so natural.

Sonified Design Phase: The Prototype

According to our meditation intervention feedback, we supplemented HU with sound. Sound can reduce stress [47,48]. Humans normally respond with affect to rhythmical sound [49], and rhythmical sound can in turn influence human behavior [50,51]. Sound can have positive psychological and physiological effects on illness [52-56]. Modern nursing pioneer Florence Nightingale noted that wind instruments, the human voice, and stringed instruments helped patients recover [57]. Alvarsson et al investigated the way sounds of nature affected subjects physiologically and psychologically [58]. Forty trial subjects were exposed to a “sounds of nature” or a “noisy” environment after performing stressful mental arithmetic. The results demonstrated that people recovered faster from sympathetic activation in the “sounds of nature” environment. Another study suggested that a natural environment had restorative effects on patients, including more positive emotions [59] and less mental fatigue [60].

We selected suitable soundtracks from a natural environmental sound database and a music database for our prototype. The prototype has two parts: a hardware prototype and a software application. The core of our hardware prototype (the sense and control unit) includes an Arduino UNO board and a pulse sensor. Heart rate data can be collected via the pulse sensor and transmitted to the Arduino UNO board. Arduino is an open source integrated programming environment [61]. The software is an audio application running on a PC developed using Pure Data, a real-time, graphical, dataflow programming application [62]. The software prototype enabled sound to be coupled with the meditator’s heart rate, so tempo and sound volume increased and decreased synchronously with heart rate. We offered three sound themes:

1. Breathing: A female breathing sound
2. Nature: Rhythmical ocean waves
3. Music: Half Moon Serenade (meditation music)

People could choose sound themes freely.

Sonified Design Phase: Evaluation

We used two approaches to evaluate HU: an exploration study and an experiment. Here, we name HU without sound as silent HU to distinguish it from sonified HU.

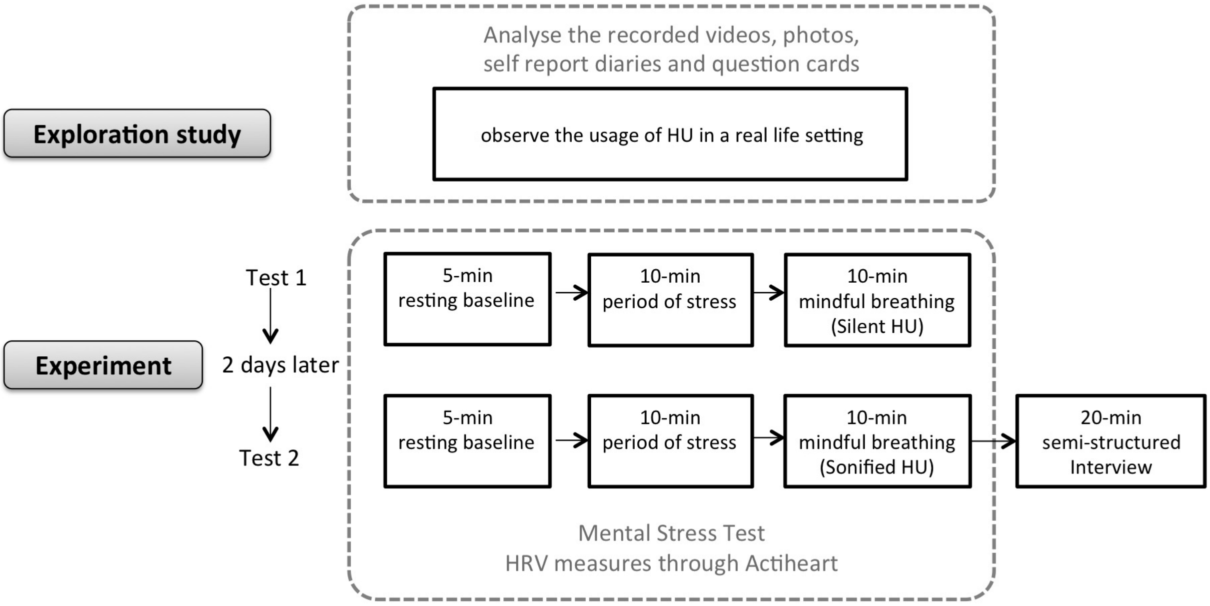
Exploration Study

To study an everyday-use scenario, the experiment was conducted in the participant’s homes (Figure 3). In the exploration stage, participants used HU for 2 days, in accordance with their own likings. During this stage, they described their experiences through diaries, photos, and short videos (of how and where they used HU). The diaries captured use conditions, including time, place, and duration, as well as experiences before and after use. To understand sound preferences for relaxation, we used picture cards on which participants attached premade cartoons to indicate their choices (Figure 4).

Figure 3. Trial photos in the context.

Figure 4. Picture cards.

Figure 5. The procedure of the test.



Experiment

We conducted an experiment to discern how HU could help reduce stress through mindful breathing and evaluated the outcome through self- assessment and HRV measures.

Measures

Mental Stress Test

A mental stress test was used as a stress provocation, building on earlier research with mental stress experiments [63]. We used a 10-min mental arithmetic test from the Wechsler Adult Intelligence Scale (WAIS) [64], as for example, the Trier social stress test [65] did not work in our setting. Some research has been performed on the stress inducing effects of WAIS [66-68].

Visual Analogue Scale (VAS) for Subjective Stress Assessment

Visual analogue scales (VAS) are often used for assessing subjective stress levels [69-71]. Our participants marked how stressed they felt on a 10 cm long line, where zero distance from the left=No stress, 5 cm=Neutral, and 10 cm=As bad as it could be.

HRV Measurements

HRV is the variability of intervals between consecutive heartbeats. It is regarded as a performance indicator of autonomic heart rate regulation and ANS balance [72]. High HRV is associated with a well-balanced ANS, whereas low HRV is associated with chronic stress and diseases such as diabetes and depression [73-75]. HRV has also been used as a measure of mental stress reactivity over short time periods. Troubat et al found that mental stress during short periods was associated with decreased mean HRV [76]. Researchers have sought methods for restoring normal HRV to achieve balance in the ANS [77].

HRV measures are usually divided into two categories: time domain measures and frequency domain measures. One commonly used time-domain measure root mean square of differences (RMSSD) between successive rhythm-to-rhythm (RR) intervals describes short-term HRV. In a study among employees of an electronics company, researchers found that RMSSD value is associated with perceived mental stress [78]. A low value indicates high stress.

Frequency domain method takes into account different frequency components of HRV. The two main frequency components that represent ANS activity are the low-frequency (LF) components (0.04-0.15Hz) and the high frequency (HF) components (0.15-0.4 Hz). The LF band reflects sympathetic activity and the HF band, the parasympathetic activity. The ratio of LF to HF power (LF:HF) has been used to calculate and assess the sympathovagal balance for indicating the function of the ANS activity [79,80].

The Actiheart device (CamNtech, UK) [81] was used for data sampling and primary data analysis in our study. It is a compact, chest-worn monitoring device that records heart rate, interbeat interval (IBI), and physical activity in one combined, lightweight waterproof unit. In the time domain analysis of HRV, the information of inter-beat interval in each measurement period was computed. In the frequency domain analysis, spectral analysis was performed.

Recruitment

Five new participants were recruited through word of mouth (3 males and 2 females, mean age 23 years). All participants were from universities in Sweden, with normal vision and hearing. Before the test, we gathered basic demographic information and their individual assessments of stress levels for the last week. The participants signed consent forms and allowed us to use their data. We used a within-subject design to evaluate silent HU and sonified HU, respectively. As perceived stress and HRV varied from individual to individual, a within-subject design was selected to reduce error variance associated with individual differences [82].

Procedure

Figure 5 shows the test procedure schematically. The experiment comprised 3 sessions:

1. 5-min resting baseline: Participants wore Actiheart and sat in a comfortable position as their resting baseline HRV was sampled.

2. 10-min period of stress: In order to observe the effect of using HU for stress reduction, we triggered a stress response in participants before the mindful breathing task. The stress-generating task was the 10-min IQ test from the Wechsler Adult Intelligence Scale (WAIS), described above.

3. 10-min mindful breathing with HU: HU was used to guide mindful breathing for 10 min. In a first test, participants used silent HU in the mindful breathing period. After 2 days, they used sonified HU in a second test.

During the tests, participants sat in an undisturbed environment. Heart rate, RMSSD, HF, and LF were measured through the Actiheart device (Figure 2). The visual analog scale described earlier was used for assessing subjective stress levels after each session of the test.

After the experiment, we performed a semistructured interview with questions organized in 3 sections. The first section contained questions about the participants' ways of relieving stress in everyday life. The second section contained questions about HU that did not have to do with the sonification (eg, how they had used HU in their everyday life and whether they could relate the vapor flow or light to their heart rate). The last section explored issues related to the sonification. First, they were asked to describe their experiences with the sonification, and that whether they could relate the sound rhythms to their heart rate. Subsequently, we asked them what characters of sound (pitch, tempo, and volume) would make them comfortable, and how the sonification could be improved.

Results

Sonified Design Phase: Results From the Exploration Study

In the exploration stage, we observed the usage of HU in home settings with 5 participants as described earlier and analyzed their recorded videos, photos, and self-report diaries. The 5 participants reported different degrees of relaxation during and after engaging with HU. Three reported that HU guided them to breathe mindfully.

> *When I used it, I know it worked with my breath and enabled me to follow it and take away my focus from the work. It makes me calm and peaceful…* [P1]
>
> *I felt relaxed when I breathed with HU. After that, I studied more efficiently…* [P4]

Three of them used HU in the evening and before going to bed. One preferred the morning and another chose the afternoon time for relaxation. They used HU for 10-20 min at a time. Most of them installed HU in their bedrooms and living rooms for mindfulness and relaxation. One used HU on a desk beside her while working. The bedroom was the most popular location for HU.

Apart from engaging with HU to support mindful breathing, they explored how HU, as an artifact, could fit into their daily life. To illustrate this last point, some found that HU could serve as a humidifier or ambient light, with the added benefit of enhancing calmness and relaxation.

> *It just makes me feel good when HU is besides me.* [P2]
>
> *I like the dim yellow light and the vapor. Especially the white vapor…looks dreamy…* [P3]
>
> *I used HU as a humidifier at home. It works for me…I like a multi-functional device...could be a breath machine for meditation, a humidifier or an ambient light…* [P5]

In terms of the 3 interactive sound themes, 2 out of 5 participants thought nature sounds best helped them relax. They commented that the sound of water such as ocean waves could help them relax. The others believed meditation music was more relaxing. The woman's breathing sound can be used for guiding the breath but was not considered as relaxing as the other two themes.

Sonified Design Phase: Results From the Experiment

A paired *t* test was used to evaluate the effects of using silent HU and sonified HU among the 5 participants. The HR, RMSSD, LF:HF, and subjective stress scores are shown in Table 1. The mean value of HR using silent HU is significantly lower than the resting baseline and sonified HU (63.4 vs 68.6, *P*=.008; 63.4 vs 68.4, *P*=.03, respectively). The mean value of RMSSD using silent HU is significantly higher than the resting baseline (58.734 vs 46.396, *P*=.04). However, the mean values of RMSSD and LF:HF during use of sonified HU showed no significant improvement compared with silent HU and the resting baseline. The subjective stress assessment was significantly improved in silent HU (3.1 vs 6.54, *P*=.01) and sonified HU (2.9 vs 6.54, *P*=.01) compared with the resting baseline. In terms of subjective stress, there was no significant difference between silent and sonified HU (*P*=.82).

Table 1. Mean and *t* test scores during resting baseline and silent HU

Measures	Baseline		Silent HU		*t* test	*P* value
	Mean	Standard error	Mean	Standard error	t_4 degrees of freedom=4	*P* significance level of .05
HR	68.6	2.32	63.4	2.77	4.8702	.008
RMSSD	46.396	8.28	58.734	8.98	−2.9317	.04
LF:HF	2.37	0.89	3.264	1.63	−0.5512	.61
Subjective stress (0-10)	6.54	0.70	3.1	0.56	4.4278	.01

Table 2. Mean and *t* test scores during resting baseline and sonified HU

Measures	Baseline		Sonified HU		*t* test	*P* value
	Mean	Standard error	Mean	Standard error	t_4 degrees of freedom=4	*P* significance level of .05
HR	68.6	2.32	68.4	3.93	0.0960	.93
RMSSD	46.396	8.28	45.094	2.42	−0.1783	.87
LF:HF	2.37	0.89	5.126	2.04	−1.3160	.26
Subjective Stress (0-10)	6.54	0.70	2.9	0.62	4.3993	.01

Table 3. Mean and *t* test scores during silent HU and sonified HU

Measures	Silent HU		Sonified HU		*t* test	*P* value
	Mean	Standard error	Mean	Standard error	t_4 degrees of freedom=4	*P* significance level of .05
HR	63.4	2.77	68.4	3.93	−3.2969	.03
RMSSD	58.734	8.98	45.094	2.42	−1.5197	.2
LF:HF	3.264	1.63	5.126	2.04	−2.3225	.08
Subjective Stress (0-10)	3.1	0.56	2.9	0.62	0.2453	.82

We compared the measures of HR, RMSSD, LF:HF, and subjective stress during each session and analyzed the results individually (in Multimedia Appendix 1).

Sonified Design Phase: Interview Findings

P1 and P4 thought silent HU worked better than sonified HU. They declared that the visual aids were more effective than the auditory for relieving stress, as the vapor seemed more connected to the breath.

> *I think it can help me some…it was a relaxing device in that sense. I would observe the light changing for a little while and then close my eyes to relax.* [P1]
>
> *I found the sound and vapor did not correspond to each other. I focused on the vapor more than the sound so that I did not notice the relationship between sounds and my bodily state.* [P4]

Three more participants were also unsure about the sound and its relation to their bodies. They guessed there might be connections between the sound and their breathing or heart rate, because the music changed when they realigned their postures:

> *Though sometimes it was a very weird pitch of the sound, I could partly find that the sound corresponded to my breathing when I took a deep breath.* [P3]
>
> *I suspect the pitch and frequency of the audio were affected by my pulse, because when my pulse was high and fast, the sound would be more rapid and at a higher pitch.* [P5]
>
> *I did not know how it worked, but I sometimes found the sound to be a little bit weird and distorted.* [P2]

Thus, the perceived clarity of mappings between body and visualizations and sonifications should be taken into account in designing interactive interfaces.

We asked concrete questions about the perceived effectiveness of the sonification for stress reduction. Three participants thought sound helped them ease tension and relax. One thought the interactive sound should be changed more gradually. One of them wanted to equip HU with surround sound for an immersive experience. One felt natural sounds such as wave sounds naturally helped that person better than music because of their rhythmic constancy. All thought the music should be simple, so they could focus on their breath and not on the music. Two participants found the changing tempo annoying. One of them put it this way:

> *Changing the tempo of the music continuously makes it weird. The music becomes distorted, and you focus on the sound. The pitch was going up and down a lot, it is better to have steady music to relax.* [P2]

One participant was annoyed with the audio clips being repeated. When asked about the dynamic volume, most said they had not noticed the volume changed with their heart rate measures.

Discussion

Principal Findings

By developing and testing HU, we found that its vapor, light, and sound could enhance stress reduction. There was much interindividual variability in terms of preferred settings and modes and perceived stress reduction, and HR and HRV measures. Subjective stress levels were significantly improved with both silent and sonified HU. The mean value of HR using silent HU was significantly lower than the resting baseline and sonified HU. The mean value of RMSSD using silent HU was significantly higher than the resting baseline. However, the mean values of RMSSD and LF and HF during the use of sonified HU showed no significant improvement compared with silent HU and the resting baseline. In terms of subjective stress, we found no significant difference between silent and sonified HU.

How the participants perceived the mappings between heart rate and visualizations and sonifications influenced their overall experience. Regarding our interactive sound program that adjusts speed and volume according to the heart rate, there are challenges that need to be addressed. The sound tempo was not the only characteristic affected by heart rate; the pitch too was changed as a side effect. From our observations and interviews, it became clear that such pitch distortions bothered the participants. It is easy to see why: tempo and pitch play effectual and esthetic roles in music and it can be difficult to change both tempo and pitch at the same time while preserving those roles. One possible solution is to keep a slow tempo all the time. We did not test with such a noninteractive sonification, but it could be interesting to compare with in the future. Another possible solution is to use advanced sound processing, so that the pitch is preserved with tempo.

Although the sound volume varied during the experiment, participants did not notice. When we asked them about how changing the volume dynamically might enhance relaxation, they all stated that such volume changes would not affect them. It is unclear to us why sound volume did not seem to matter.

An interesting finding is that P2 and P3 felt sleepy when using HU. Participants felt calm after they used sonified HU, and most of them used HU in their bedrooms. One participant said:

> *It makes me peaceful and sleepy. Therefore, I think it will be an excellent tool for sleep.* [P2]

HU could perhaps function as a sleep aid. HU was designed as digital mindfulness artifact, and mindfulness is commonly associated with alert wakefulness. However, most meditative traditions have incorporated measures for keeping mediators awake. It is common, for example, in many schools of Buddhism to hit meditators with a stick on their backs when they fall asleep. Zen masters often do this. As our users engaged with HU, participants who felt sleepy were clearly able to calm down as in mindful mediation and like many other meditators this brought them closer to the sleeping state. A similar finding was reported by Britton et al, that is, that early phases of meditation may produce more fatigue and sleep propensity whereas later stages produce greater wakefulness as a result of neuroplastic changes and more efficient processing [83]. Sleepiness may not be the state that is most desirable in mindful mediation, but it seems to be a phase novice practitioners often go through. This would be interesting to explore in detail in a future study.

Another point to be discussed is the discrepancy between Actiheart measurements and subjective experience. For example, P3 reported feeling more relaxed in the period of sonified HU with a score of 0.5 compared with 3 of silent HU. However, RMSSD and LF and HF measures indicated that sonified HU was associated with more stress than silent HU and the resting period. The post-study interview may help us understand the reason for the discrepancy between subjective and physiological measures. As P3 said:

> *I do not like that the tempo of the sound changed. If your heart rate is below your initial point, you are very relaxed. However, if your heart rate is higher than a particular point, say 79, the music you used was at very high pace, and you couldn't feel relaxed...* [P3]

There was also a discrepancy between physiological measurements and self-reported stress levels in the case of P5. P5 expressed her critical opinion on the dynamic sound:

> *I do not like this form of the sound. It is acceptable that the tempo is changing but not for the pitch. I also think the rhythm should not be modified a lot.* [P5]

What is the right way to think about the physiological HRV measures and the self-reported stress measures? We cannot make sense of these measures without taking into account several complicated factors. One factor is for example that of the psychological situation of the test subject; it could be that the test subject is reporting feelings of relaxation while the subject is in fact not experiencing such feelings. The subject might be trying to please the experiment leader. Another factor could be sampling: perhaps the HRV equipment did not sample the signals adequately. A third factor is that the physiological and psychological measures actually might come apart, that is, they do not correspond perfectly. We hope to explore further about the relations between HRV measures and self-reported stress measures in a future follow-up study. We learned from our case study how important it is to work in a cross-disciplinary team and to make sense of the data from different perspectives. Calvo et al have stressed the significance of partnership between psychology, social sciences, and technologists [31] in the domain of human-computer–interaction for well-being. We use objective and subjective (qualitative and quantitative) approaches in terms of the assessment of mindfulness-based stress reduction. HRV measurements are commonly used in stress and medical research. These objective measures of stress are seldom used by researchers in human-computer–interaction and user experience design. The results of our study suggest we should not abandon subjective reports of user experiences. Both subjective and objective methods give valuable information. We found that the differences between the objective and subjective assessments were intriguing and prompt us to investigate them further.

Limitations

The main limitation of this study was the low number of participants. Moreover, the participants were recruited by a snowball sampling technique. To prove that sonified HU is effective in comparison with silent HU in improving HRV, more users would have to be recruited and according to a random sampling scheme. The participants' home environments were chosen as the experimental settings, but naturally, it is impossible to control for potentially disturbing factors in the same way that one can do in a lab or dedicated test environment.

Future Work

This study provoked greater thinking about how to design sonified devices for relaxation and mindfulness. The question of how biodata-based interactive sound could potentially reduce stress, along with the question of whether interactive sound or regular music or sounds has a more positive influence on stress reduction are topics that warrant further exploration. We plan to continue our investigations and design headphones with a pulse sensor and sound files embedded inside. A further challenge with such a device would be to incorporate HRV functionally.

Conclusions

The primary purpose of this study was to investigate whether a newly developed tool, HU, could support mindful breathing and help users reduce stress after exposure to a mental stressor. The observations, interviews, and HRV data helped us to, if not fully answer this question, explore it. Our study indicates the potential of interaction design based on vapor, light, and sound to support mindful breathing and stress reduction. Further to this, feedback from user explorations of HU in their homes and the interviews helped us understand more about the subjective experiences of using this kind of installation.

Acknowledgments

We thank all the participants of the user study. We also thank PhD Malena Ivarsson for excellent help and advice regarding the use of Actiheart.

Authors' Contributions

BZ contributed to the design of HU. SF and BZ conducted the user study and data analysis. AH, BZ, and SF drafted the paper; WO and HL reviewed the manuscript and contributed to subsequent drafts.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of the measures of HR, RMSSD, LF/HF, and subjective stress during each session and individual analysis of the results.

[PDF File (Adobe PDF File), 150KB - jmir_v19i6e197_app1.pdf]

References

1. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020 URL: http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng.pdf [accessed 2016-09-29] [WebCite Cache ID 6q5NLmSwi]
2. Bawden D, Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies. J Inf Sci 2009;35(2):180-191. [doi: 10.1177/0165551508095781]
3. European Agency for Safety & Health at Work. 2014. Report-Calculating the cost of work-related stress and psychosocial risks URL: https://osha.europa.eu/en/publications/literature_reviews/calculating-the-cost-of-work-related-stress-and-psychosocial-risks/view [accessed 2016-09-29] [WebCite Cache ID 6q5O5FkY0]
4. Schmidt A, Thews G. Autonomic Nervous System. In: Janig W, editor. Human Physiology, 2nd edition. New York, NY: Springer-Verlag; 1989:333-370.
5. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
6. McEwen BS. Protective and damaging effects of stress mediators: central role of the brain. Dialogues Clin Neurosci 2006;8(4):367-381. [Medline: 17290796]
7. Grossi G, Perski A, Osika W, Savic I. Stress-related exhaustion disorder--clinical manifestation of burnout? a review of assessment methods, sleep impairments, cognitive disturbances, and neuro-biological and physiological changes in clinical burnout. Scand J Psychol 2015;56(6):626-636. [doi: 10.1111/sjop.12251] [Medline: 26496458]
8. Ho DY, Chan SF, Peng SQ, Ng AK. The dialogical self: converging East-West constructions. Cult Psychol 2001;7(3):393-408. [doi: 10.1177/1354067X0173008]
9. Chan CL, Ng SM, Ho RT, Chow AY. East meets West: applying Eastern spirituality in clinical practice. J Clin Nurs 2006;15(7):822-832. [doi: 10.1111/j.1365-2702.2006.01649.x]
10. Dockett KH, Dudley-Grant GR, Bankart CP. Psychology and Buddhism: From individual to global community. New York, NY: Springer Science & Business Media; 2003.
11. Chan CL, Palley HA. The use of traditional Chinese culture and values in social work health care related interventions in Hong Kong. Health & Social Work 2005;30(1):76. [Medline: 15847241]
12. Wall RB. Tai Chi and mindfulness-based stress reduction in a Boston Public Middle School. J Pediatr Health Care 2005;19(4):230-237. [doi: 10.1016/j.pedhc.2005.02.006] [Medline: 16010262]
13. Wall RB. Teaching tai chi with mindfulness-based stress reduction to middle school children in the inner city: a review of the literature and approaches. InTai Chi Chuan 2008;52:166-172. [doi: 10.1159/000134297] [Medline: 18487896]
14. Griffith JM, Hasley JP, Liu H, Severn DG, Conner LH, Adler LE. Qigong stress reduction in hospital staff. J Altern Complement Med 2008;14(8):939-945 [FREE Full text] [doi: 10.1089/acm.2007.0814] [Medline: 18823261]
15. Skoglund L, Jansson E. Qigong reduces stress in computer operators. Complement Ther Clin Pract 2007;13(2):78-84. [doi: 10.1016/j.ctcp.2006.09.003] [Medline: 17400142]
16. Germer CK, Siegel RD, Fulton PR. Mindfulness and psychotherapy. New York City: Guilford Publications; 2016:128.
17. Hyland PK, Lee RA, Mills MJ. Mindfulness at work: a new approach to improving individual and organizational performance. Ind Organ Psychol 2015;8(04):576-602. [doi: 10.1017/iop.2015.41]
18. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol: Science and Practice 2003;10(2):144-156. [doi: 10.1093/clipsy.bpg016]
19. Marchand WR. Mindfulness-based stress reduction, mindfulness-based cognitive therapy, and Zen meditation for depression, anxiety, pain, and psychological distress. J Psychiatr Pract 2012 Jul;18(4):233-252. [doi: 10.1097/01.pra.0000416014.53215.86] [Medline: 22805898]
20. Baer RA, Carmody J, Hunsinger M. Weekly change in mindfulness and perceived stress in a mindfulness‐based stress reduction program. J Clin Psychol 2012;68(7):755-765. [doi: 10.1002/jclp.21865] [Medline: 22623334]
21. Huppert FA, Johnson DM. A controlled trial of mindfulness training in schools: the importance of practice for an impact on well-being. J Posit Psychol 2010;5(4):264-274. [doi: 10.1080/17439761003794148]
22. Khoury B, Lecomte T, Fortin G, Masse M, Therien P, Bouchard V, et al. Mindfulness-based therapy: a comprehensive meta-analysis. Clin Psychol Rev 2013;33(6):763-771. [doi: 10.1016/j.cpr.2013.05.005] [Medline: 23796855]
23. Chiesa A, Serretti A. Mindfulness based cognitive therapy for psychiatric disorders: a systematic review and meta-analysis. Psychiatry Res 2011 May 30;187(3):441-453. [doi: 10.1016/j.psychres.2010.08.011] [Medline: 20846726]
24. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
25. Bränström R, Kvillemo P, Brandberg Y, Moskowitz JT. Self-report mindfulness as a mediator of psychological well-being in a stress reduction intervention for cancer patients—A randomized study. Ann Behav Med 2010;39(2):151-161. [doi: 10.1007/s12160-010-9168-6] [Medline: 20177843]
26. Calvo RA, Peters D. Positive computing: technology for wellbeing and human potential. Cambridge: MIT Press; 2014.

27. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services. Med J Aust 2010;192(11):S53. [Medline: 20528711]
28. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010;192(11):S22. [Medline: 20528703]
29. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth and uHealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
30. Chittaro L, Vianello A. Computer-supported mindfulness: evaluation of a mobile thought distancing application on naive meditators. Int J Hum Comput Stud 2014;72(3):337-348. [doi: 10.1016/j.ijhcs.2013.11.001]
31. Calvo RA, Vella-Brodrick D, Desmet P, Ryan RM. Editorial for "positive computing: a new partnership between psychology, social sciences and technologists". Psychol Well Being 2016;6(1):10 [FREE Full text] [doi: 10.1186/s13612-016-0047-1] [Medline: 27441171]
32. Busch V, Magerl W, Kern U, Haas J, Hajak G, Eichhammer P. The effect of deep and slow breathing on pain perception, autonomic activity, and mood processing—an experimental study. Pain Med 2012;13(2):215-228 [FREE Full text] [doi: 10.1111/j.1526-4637.2011.01243.x] [Medline: 21939499]
33. Wollmann T, Abtahi F, Eghdam A, Seoane F, Lindecrantz K, Haag M, et al. User-centred design and usability evaluation of a heart rate variability biofeedback game. IEEE Access 2016;4:5531-5539. [doi: 10.1109/ACCESS.2016.2601882]
34. Khut GP, Muller L. Evolving creative practice: a reflection on working with audience experience in Cardiomorphologies. 2005 Presented at: Vital Signs: Creative Practice & New Media Now; September 2005; Melbourne. [doi: 10.1109/mmul.2007.4298460]
35. Khut GP, Morrow A, Watanabe MY. The BrightHearts project: a new approach to the management of procedure-related paediatric anxiety. 2011 Presented at: OZCHI Workshop Call for Participation; November 2011; Canberra, Australia p. 17.
36. Morrow A, Yogul M, Khut PG. Brighthearts: development of a biofeedback controlled interactive artwork for the management of procedural pain and anxiety in children. Developmental Medicine & Child Neurology 2014;56(14).
37. Wongsuphasawat K, Gamburg A, Moraveji N. You can't force calm: designing and evaluating respiratory regulating interfaces for calming technology. 2012 Presented at: In Adjunct proceedings of the 25th annual ACM symposium on User interface software and technology; October 7, 2012; Cambridge, MA p. 69-70. [doi: 10.1145/2380296.2380326]
38. Short AE, Ahern N, Holdgate A, Morris J, Sidhu B. Using music to reduce noise stress for patients in the emergency department: a pilot study. Music Med 2010 Jul 07;2(4):201-207.
39. Vidyarthi J, Riecke BE, Gromala D. Sonic Cradle: designing for an immersive experience of meditation by connecting respiration to music. 2012 Presented at: In Proceedings of the designing interactive systems conference; June 11, 2012; Newcastle, UK p. 408-417. [doi: 10.1145/2317956.2318017]
40. Shusterman R. Body consciousness: A philosophy of mindfulness and somaesthetics. Cambridge: Cambridge University Press; 2008.
41. Höök K, Jonsson M, Ståhl A, Mercurio J. Somaesthetic Appreciation Design. 2016 Presented at: the SIGCHI Conference on Human Factors in Computing Systems (CHI 16); May 7, 2016; San Jose, MA. [doi: 10.1145/2858036.2858583]
42. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. 2007 Presented at: the SIGCHI conference on Human factors in computing systems; April 29, 2007; San Jose, MA p. 493-502. [doi: 10.1145/1240624.1240704]
43. Lehrer PM, Vaschillo E, Vaschillo B. Resonant frequency biofeedback training to increase cardiac variability: rationale and manual for training. Appl Psychophysiol Biofeedback 2000;25(3):177-191. [Medline: 10999236]
44. Vaschillo EG, Vaschillo B, Lehrer PM. Characteristics of resonance in heart rate variability stimulated by biofeedback. Appl Psychophysiol Biofeedback 2006;31(2):129-142. [doi: 10.1007/s10484-006-9009-3] [Medline: 16838124]
45. Sutarto AP, Wahab MN, Zin NM. Heart Rate Variability (HRV) biofeedback: A new training approach for operator's performance enhancement. Journal of Industrial Engineering and Management 2010 Jun 22;3(1):176-198. [doi: 10.3926/jiem.2010.v3n1.p176-198]
46. Goodman LA. Snowball sampling. Ann Math Stat 1961:148-170. [doi: 10.1214/aoms/1177705148]
47. Kemper KJ, Danhauer SC. Music as therapy. Southern Medical Journal 2005;98(3):282-288. [doi: 10.1097/01.smj.0000154773.11986.39]
48. Salamon E, Kim M, Beaulieu J, Stefano GB. Sound therapy induced relaxation: down regulating stress processes and pathologies. Med Sci Monit 2003;9(5):RA96-RA101 [FREE Full text] [Medline: 12761468]
49. MacDougall R. The affective quality of auditory rhythm in its relation to objective forms. Psychol Rev 1903;10(1):15-36. [doi: 10.1037/h0072426]
50. Bauer C, Waldner F. Reactive music: when user behavior affects sounds in real-time. 2013 Presented at: CHI'13 Extended Abstracts on Human Factors in Computing Systems; April 27, 2013; Paris p. 739-744. [doi: 10.1145/2468356.2468488]
51. Karageorghis C, Jones L, Stuart DP. Psychological effects of music tempi during exercise. Int J Sports Med 2008;29(7):613-619. [doi: 10.1055/s-2007-989266] [Medline: 18050063]
52. Naghdi L, Ahonen H, Macario P, Bartel L. The effect of low-frequency sound stimulation on patients with fibromyalgia: a clinical study. Pain Res Manag 2015;20(1):e21-e27.

53. Henry JA, Zaugg TL, Myers PJ, Schechter MA. Using therapeutic sound with progressive audiologic tinnitus management. Trends Amplif 2008;12(3):188-209 [FREE Full text] [doi: 10.1177/1084713808321184] [Medline: 18664499]
54. White JM. Music therapy: an intervention to reduce anxiety in the myocardial infarction patient. Clin Nurse Spec 1992;6(2):58-63. [Medline: 1617576]
55. Ellis P. Vibroacoustic sound therapy: case studies with children with profound and multiple learning difficulties and the elderly in long-term residential care. Stud Health Technol Inform 2003;103:36-42. [Medline: 15747903]
56. Augustin P, Hains AA. Effect of music on ambulatory surgery patients' preoperative anxiety. AORN journal 1996;63(4):750-758. [Medline: 8660020]
57. Nightingale F. Notes on Nursing: What it is and what it not. Cambridge: Cambridge University Press; 2010.
58. Alvarsson JJ, Wiens S, Nilsson ME. Stress recovery during exposure to nature sound and environmental noise. Int J Environ Res Public Health 2010;7(3):1036-1046. [doi: 10.3390/ijerph7031036]
59. Ulrich RS, Simons RF, Losito BD, Fiorito E, Miles MA, Zelson M. Stress recovery during exposure to natural and urban environments. J Environ Psychol 1991 Sep;11(3):201-230. [doi: 10.1016/s0272-4944(05)80184-7]
60. Kaplan S. The restorative benefits of nature: toward an integrative framework. J Environ Psychol 1995;15(3):169-182. [doi: 10.1016/0272-4944(95)90001-2]
61. Arduino. Arduino - Introduction URL: https://www.arduino.cc/en/Guide/Introduction [accessed 2016-11-12] [WebCite Cache ID 6lyHdVrxY]
62. Pure Data. Pure Data information URL: https://puredata.info/ [accessed 2016-11-12] [WebCite Cache ID 6ly2KkSIt]
63. Chen Y, Dangardt F, Osika W, Berggren K, Gronowitz E, Friberg P. Age-and sex-related differences in vascular function and vascular response to mental stress: longitudinal and cross-sectional studies in a cohort of healthy children and adolescents. Atherosclerosis 2012;220(1):269-274. [doi: 10.1016/j.atherosclerosis.2011.10.030] [Medline: 22078247]
64. Kaufman AS, Lichtenberger EO. Assessing adolescent and adult intelligence. US: John Wiley & Sons; 2005.
65. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993;28(1-2):76-81. [doi: 10.1159/000119004]
66. Chaudhry M, Ready R. Differential effects of test anxiety & stress on the WAIS-IV. Journal of Young Investigators 2012;24(5):60-66.
67. Neupert SD, Miller LM, Lachman ME. Physiological reactivity to cognitive stressors: variations by age and socioeconomic status. Int J Aging Hum Dev 2006;62(3):221-235. [doi: 10.2190/17DU-21AA-5HUK-7UFG] [Medline: 16625938]
68. Barnes RF, Raskind M, Gumbrecht G, Halter JB. The effects of age on the plasma catecholamine response to mental stress in man. J Clin Endocrinol Metab 1982;54(1):64-69. [doi: 10.1210/jcem-54-1-64] [Medline: 7054219]
69. Bond A, Lader M. The use of analogue scales in rating subjective feelings. Br J Med Psychol 1974;47(3):211-218. [doi: 10.1111/j.2044-8341.1974.tb02285.x]
70. Lesage FX, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. Occup Med 2012;62(8):600-605. [doi: 10.1093/occmed/kqs140] [Medline: 22965867]
71. Lesage FX, Berjot S. Validity of occupational stress assessment using a visual analogue scale. Occup Med 2011;61(6):434-436. [doi: 10.1093/occmed/kqr037] [Medline: 21505089]
72. Sztajzel J. Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. Swiss Med Wkly 2004;134(35-36):514-522. [Medline: 15517504]
73. Schubert C, Lambertz M, Nelesen RA, Bardwell W, Choi JB, Dimsdale JE. Effects of stress on heart rate complexity—a comparison between short-term and chronic stress. Biol Psychol 2009;80(3):325-332 [FREE Full text] [doi: 10.1016/j.biopsycho.2008.11.005] [Medline: 19100813]
74. Champaneri S, Wand GS, Malhotra SS, Casagrande SS, Golden SH. Biological basis of depression in adults with diabetes. Curr Diab Rep 2010;10(6):396-405. [doi: 10.1007/s11892-010-0148-9] [Medline: 20878274]
75. Carney RM, Freedland KE, Veith RC. Depression, the autonomic nervous system, and coronary heart disease. Psychosom Med 2005;67 Suppl 1:S29-S33. [Medline: 15953797]
76. Troubat N, Fargeas-Gluck MA, Tulppo M, Dugué B. The stress of chess players as a model to study the effects of psychological stimuli on physiological responses: an example of substrate oxidation and heart rate variability in man. Eur J Appl Physiol 2009;105(3):343-349. [doi: 10.1007/s00421-008-0908-2] [Medline: 18987876]
77. Abtahi F. Towards Heart Rate Variability Tools in P-Health: Pervasive, Preventive, Predictive and Personalized (Doctoral dissertation, KTH Royal Institute of Technology). Sweden: KTH Royal Institute of Technology; 2016.
78. Orsila R, Virtanen M, Luukkaala T, Tarvainen M, Karjalainen P, Viik J, et al. Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company. Int J Occup Saf Ergon 2008;14(3):275-283. [doi: 10.1080/10803548.2008.11076767] [Medline: 18954537]
79. Pagani M, Lombardi F, Guzzetti S, Rimoldi O, Furlan RA, Pizzinelli PA, et al. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. Circ Res 1986;59(2):178-193. [doi: 10.1161/01.res.59.2.178]
80. Bigger JT, Fleiss JL, Steinman RC, Rolnitzky LM, Kleiger RE, Rottman JN. Frequency domain measures of heart period variability and mortality after myocardial infarction. Circulation 1992;85(1):164-171. [Medline: 1728446]

81. Camntech. Actiheart-overview URL: https://www.camntech.com/products/actiheart/actiheart-overview [accessed 2016-09-12] [WebCite Cache ID 6lyJqZ3jT]
82. Leroy G. Designing User Studies in Informatics. New York: Springer Science & Business Media; 2011.
83. Britton WB, Lindahl JR, Cahn BR, Davis JH, Goldman RE. Awakening is not a metaphor: the effects of Buddhist meditation practices on basic wakefulness. Ann N Y Acad Sci 2014;1307(1):64-81. [doi: 10.1111/nyas.12279]

Abbreviations

ANS: autonomic nervous system
MBSR: mindfulness-based stress reduction
HR: heart rate
HRV: heart rate variability

Edited by R Calvo, J Torous; submitted 17.11.16; peer-reviewed by S Stoyanov, C Linda, C De Berigny; comments to author 06.02.17; revised version received 24.02.17; accepted 03.03.17; published 14.06.17.

Please cite as:
Zhu B, Hedman A, Feng S, Li H, Osika W
Designing, Prototyping and Evaluating Digital Mindfulness Applications: A Case Study of Mindful Breathing for Stress Reduction
J Med Internet Res 2017;19(6):e197
URL: http://www.jmir.org/2017/6/e197/
doi:10.2196/jmir.6955
PMID:28615157

Letter to the Editor

The Research Topic Defines “Noise” in Social Media Data – a Response from the Authors

Yoonsang Kim[1], PhD; Jidong Huang[2], PhD; Sherry Emery[1], PhD

[1]NORC at the University of Chicago, Health Media Collaboratory, Chicago, IL, United States
[2]Georgia State University, Health Management and Policy, Atlanta, Georgia

Corresponding Author:
Yoonsang Kim, PhD
NORC at the University of Chicago
Health Media Collaboratory
55 E Monroe St, 30th Floor
Chicago, IL, 60603
United States
Phone: 1 312 357 3878
Fax: 1 312 759 4000
Email: kim-yoonsang@norc.org

Related Articles:

Comment on: http://jmir.org/2016/8/e219/

Comment on: http://jmir.org/2016/2/e41/

(J Med Internet Res 2017;19(6):e165) doi:10.2196/jmir.6824

KEYWORDS

automated tweets, noise, social media data

We provide a response to Allem and Ferrara [1], who recently commented on our article, “Garbage in Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection,” which was published in JMIR in February 2016 [2]. In their comment, published in JMIR in August 2016, entitled “The importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors,” Allem and Ferrara discuss the importance of removing bias in social media data. They claim that automated tweets are noise that injects bias into the data, and thus should be removed before applying the framework we proposed [1]. We believe they misunderstood our intent. In addition, their discussion misinterprets the key messages of our article; the implication of their comments, which suggests that automated tweets are garbage, is highly misleading. A formal response is provided here to articulate accurately the main focus of our article and present a different view about the “noise” in social media data.

The objective of our paper was “to develop and apply a framework of social media data collection and quality assessment, and to propose a reporting standard,” as stated in the abstract. The e-cigarette-related tweet data were used as “a real-world example” to demonstrate how to apply this framework to develop a search filter, and how to estimate the measures of data quality under different conditions. The objective of our paper was not to understand e-cigarette-related attitudes and behaviors expressed on Twitter.

The definition of the “noise” in social media data by Allem and Ferrara, as any tweets produced from an account identified as a social bot, is narrow and oversimplifying, and may even be misleading in some cases. Organic and commercial tweets are not isolated in the Twittersphere. Many organic tweets are retweets or replies to commercial tweets, of which a large number is generated by bots. Whether automated contents generated by bots should be considered as noise depends on the research topic at hand. Although it may be important to remove bot tweets and focus solely on organic contents for certain research topics, it is equally important to measure the amount of these bot tweets and the content of (mis)information in these tweets for many other research topics [3]. For example, a study that examines the commercial advertising on e-cigarette should include the tweets generated by bots. The automated social media messages are not unique to the topic of e-cigarettes. For many other research topics, including other tobacco products, pharmaceutical products, dietary supplements, etc., automatically-generated marketing content is common. In fact, one of the studies that Allem and Ferrara cited to justify removing automated tweets discussed the value of "understanding the effect of promotionally marketing vaporization products" on social media using "cyborgs to mimic organic users" because of their importance to public health and policy [4]. This underscores the importance of being able to identify and quantify such automated messages in order to understand their impact on the marketplace and individual attitudes, beliefs and behaviors.

Allem and Ferrara also briefly discussed the inherent bias in social media data due to the fact that social media users are not a representative sample of the general population. However, this itself does not limit the value of social media data, and it can be used as an advantage to study hard-to-reach populations such as young adults, and ethnic, racial, and sexual minorities. Social media can serve as a good alternative or complementary data source to understand behavior and intentions among these understudied and hard-to-reach groups.

Removing automated contents and applying other approaches to remove noise can be considered in the stage of developing search filters if it is deemed appropriate for the research topics in study. However, it is not a necessary component to be considered for all research using social media data. This point underscores the main thesis of our paper: that clear disclosure about data cleaning and processing (e.g. whether bot tweets are included or not) is important.

Acknowledgments

This study was supported by the FDA Center for Tobacco Products under Award Number P50CA179546. The content is solely the responsibility of the authors and does not necessarily represent the official views of the FDA.

Conflicts of Interest

None declared.

References

1. Allem JP, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. J Med Internet Res 2016 Aug 09;18(8):e219 [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]
2. Kim Y, Huang J, Emery S. Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. J Med Internet Res 2016 Feb 26;18(2):e41 [FREE Full text] [doi: 10.2196/jmir.4738] [Medline: 26920122]
3. Gruzd A. Who are We Modelling: Bots or Humans? In: Proceedings of the 25th International Conference Companion on World Wide Web. Geneva: International World Wide Web Conferences Steering Committee; 2016:551.
4. Clark EM, Jones CA, Williams JR, Kurti AN, Norotsky MC, Danforth CM, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. PLoS One 2016 Jul;11(7):e0157304 [FREE Full text] [doi: 10.1371/journal.pone.0157304] [Medline: 27410031]

Edited by G Eysenbach; submitted 11.11.16; this is a non–peer-reviewed article;accepted 08.05.17; published 02.06.17.

Please cite as:
Kim Y, Huang J, Emery S
The Research Topic Defines "Noise" in Social Media Data – a Response from the Authors
J Med Internet Res 2017;19(6):e165
URL: http://www.jmir.org/2017/6/e165/
doi:10.2196/jmir.6824
PMID:28576756

Corrigenda and Addenda

Metadata Correction: Importance-Performance Analysis of Personal Health Records in Taiwan: A Web-Based Survey

Hsiao-Hsien Rau[1], PhD; Yi-Syuan Wu[2,3*], Master; Chi-Ming Chu[4*], PhD; Fu-Chung Wang[5*], PhD; Min-Huei Hsu[1*], MD, PhD; Chi-Wen Chang[6,7*], PhD; Kang-Hua Chen[6*], PhD; Yen-Liang Lee[1,8*], PhD; Senyeong Kao[4*], PhD; Yu-Lung Chiu[4*], PhD; Hsyien-Chia Wen[9*], PhD; Anis Fuad[1,10*], MSc; Chien-Yeh Hsu[1,11,12*], PhD; Hung-Wen Chiu[1*], PhD

[1]Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan
[2]Administration Department, Healthconn Corp, Taipei, Taiwan
[3]Institute of Medical Informatics, Department of Computer Science of Information Engineering, National Cheng Kung University, Tainan, Taiwan
[4]School Of Public Health, National Defense Medical Center, Taiepi, Taiwan
[5]National Health Insurance Administration, Ministry of Health and Welfare, Taiepi, Taiwan
[6]School of Nursing, College of Medicine, Chang Gung University, Linkou, Taiwan
[7]Division of Endocrinology, Department of Pediatrics, Linkou Chang Gung Memorial Hospital, Linkou, Taiwan
[8]Internet of Things Laboratory, Chunghwa Telecom Laboratories, TaoYuen, Taiwan
[9]School of Health Care Administration, Taipei Medical University, Taipei, Taiwan
[10]Department of Biostatistics, Epidemiology and Population Health, Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia
[11]Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan
[12]Master Program in Global Health and Development, Taipei Medical University, Taipei, Taiwan
*these authors contributed equally

Corresponding Author:
Hung-Wen Chiu, PhD
Graduate Institute of Biomedical Informatics
Taipei Medical University
15F, No.172-3, Sec 2, Keelung Rd, Da’an Dist
Taipei, 106
Taiwan
Phone: 886 266382736 ext 1511
Email: hwchiu@tmu.edu.tw

Related Article:

Correction of: http://www.jmir.org/2017/4/e131

(J Med Internet Res 2017;19(6):e209) doi:10.2196/jmir.7944

The authors of “Importance-Performance Analysis of Personal Health Records in Taiwan: A Web-Based Survey” (J Med Internet Res 2017;19(4):e131) overlooked omissions in the metadata. Two of the article's authors should have been designated as having made an equal contribution. The authors are Dr. Yu-Lung Chiu and Dr. Hsyien-Chia Wen. Their contributions to this paper were significant, and we apologize for the omission in the original article.

This correction has been made in the online version of the paper on the JMIR website on June 9, 2017, together with the publication of this corrigendum.

A correction notice has been sent to PubMed, and the publication was resubmitted to Pubmed Central and other full-text repositories.

Edited by G Eysenbach; submitted 30.04.17; this is a non–peer-reviewed article;accepted 29.05.17; published 09.06.17.

Please cite as:
Rau HH, Wu YS, Chu CM, Wang FC, Hsu MH, Chang CW, Chen KH, Lee YL, Kao S, Chiu YL, Wen HC, Fuad A, Hsu CY, Chiu HW
Metadata Correction: Importance-Performance Analysis of Personal Health Records in Taiwan: A Web-Based Survey
J Med Internet Res 2017;19(6):e209
URL: http://www.jmir.org/2017/6/e209/
doi:10.2196/jmir.7944
PMID:30578232

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