Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 19 (2017), Issue 5 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Editorial

Using mHealth to Predict Noncommunicable Diseases: A Public Health Opportunity for Low- and Middle-Income Countries (e129)
Ellen Rosskam, Adnan Hyder. . . . 6

Guest Editorial

Leveraging Mobile Phones for Monitoring Risks for Noncommunicable Diseases in the Future (e137)
Jennifer Ellis. . . . 10

Original Papers

Fifteen Challenges in Establishing a Multidisciplinary Research Program on eHealth Research in a University Setting: A Case Study (e173)
Helena Grönqvist, Erik Olsson, Birgitta Johansson, Claes Held, Jonas Sjöström, Annika Lindahl Norberg, Emma Hovén, Robbert Sanderman, Theo van Achterberg, Louise von Essen. . . . 13

Self-Management Support Using a Digital Health System Compared With Usual Care for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial (e144)
Andrew Farmer, Veronika Williams, Carmelo Velardo, Syed Shah, Ly-Mee Yu, Heather Rutter, Louise Jones, Nicola Williams, Carl Heneghan, Jonathan Price, Maxine Hardinge, Lionel Tarassenko. . . . 23

Embodied Conversational Agents in Clinical Psychology: A Scoping Review (e151)
Simon Provoost, Ho Lau, Jeroen Ruwaard, Heleen Riper. . . . 38

What Matters in Weight Loss? An In-Depth Analysis of Self-Monitoring (e160)
Stefanie Painter, Rezwan Ahmed, James Hill, Robert Kushner, Richard Lindquist, Scott Brunning, Amy Margulies. . . . 56

Web-Based Cognitive Behavior Therapy for Depression in People With Diabetes Mellitus: A Randomized Controlled Trial (e157)
Jill Newby, Lisa Robins, Kay Wilhelm, Jessica Smith, Therese Fletcher, Inika Gillis, Trevor Ma, Adam Finch, Lesley Campbell, Gavin Andrews.
7 0

Digital Support Interventions for the Self-Management of Low Back Pain: A Systematic Review (e179)
Barbara Nicholl, Louise Sandal, Mette Stochkendahl, Marianne McCallum, Nithya Suresh, Ottar Vasseljen, Jan Hartvigsen, Paul Mork, Per Kjaer, Karen Søgaard, Frances Mair. . . . 87

XSL•FO
RenderX

Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study (e187)
Silje Lie, Bjørg Karlsen, Ellen Oord, Marit Graue, Bjørg Oftedal. 108

High Level of Integration in Integrated Disease Management Leads to Higher Usage in the e-Vita Study: Self-Management of Chronic Obstructive Pulmonary Disease With Web-Based Platforms in a Parallel Cohort Design (e185)
Esther Talboom-Kamp, Noortje Verdijk, Marise Kasteleyn, Lara Harmans, Irvin Talboom, Mattijs Numans, Niels Chavannes. 119

Evaluation of Pollen Apps Forecasts: The Need for Quality Control in an eHealth Service (e152)
Katharina Bastl, Uwe Berger, Maximilian Kmenta. 134

Activity Recognition for Persons With Stroke Using Mobile Phone Technology: Toward Improved Performance in a Home Setting (e184)
Megan O'Brien, Nicholas Shawen, Chaithanya Mummidisetty, Saninder Kaur, Xiao Bo, Christian Poellabauer, Konrad Kording, Arun Jayaraman.
1 5 6

Harnessing Facebook for Smoking Reduction and Cessation Interventions: Facebook User Engagement and Social Support Predict Smoking Reduction (e168)
Sunny Kim, Lisa Marsch, Mary Brunette, Jesse Dallery. 170

Evaluating the Social Media Performance of Hospitals in Spain: A Longitudinal and Comparative Study (e181)
Antonio Martinez-Millana, Carlos Fernandez-Llatas, Ignacio Basagoiti Bilbao, Manuel Traver Salcedo, Vicente Traver Salcedo. 184

Shifting Practices Toward Recovery-Oriented Care Through an E-Recovery Portal in Community Mental Health Care: A Mixed-Methods Exploratory Study (e145)
Deede Gammon, Monica Strand, Lillian Eng, Elin Børøsund, Cecilie Varsi, Cornelia Ruland. 197

Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates (e156)
Koustuv Saha, Ingmar Weber, Michael Birnbaum, Munmun De Choudhury. 212

Automated Text Messaging as an Adjunct to Cognitive Behavioral Therapy for Depression: A Clinical Trial (e148)
Adrian Aguilera, Emma Bruehlman-Senecal, Orianna Demasi, Patricia Avila. 225

Efficacy of a Web-Based Guided Recommendation Service for a Curated List of Readily Available Mental Health and Well-Being Mobile Apps for Young People: Randomized Controlled Trial (e141)
Niranjan Bidargaddi, Peter Musiat, Megan Winsall, Gillian Vogl, Victoria Blake, Stephen Quinn, Simone Orlowski, Gaston Antezana, Geoffrey Schrader. 239

Attachment Style and Internet Addiction: An Online Survey (e170)
Christiane Eichenberg, Markus Schott, Oliver Decker, Brigitte Sindelar. 253

Public Response to Obamacare on Twitter (e167)
Matthew Davis, Kai Zheng, Yang Liu, Helen Levy. 294

Prioritizing Measures of Digital Patient Engagement: A Delphi Expert Panel Study (e182)
Lynn Garvin, Steven Simon. 304

Health-Related Internet Use in People With Cancer: Results From a Cross-Sectional Study in Two Outpatient Clinics in Sweden (e163)
Susanne Mattsson, Erik Olsson, Birgitta Johansson, Maria Carlsson. 326

XSL•FO
RenderX

Diversity in Older Adults' Use of the Internet: Identifying Subgroups Through Latent Class Analysis (e180)
Leonieke van Boekel, Sebastiaan Peek, Katrien Luijkx. . . . 336

Patient-Centered Decision Support: Formative Usability Evaluation of Integrated Clinical Decision Support With a Patient Decision Aid for Minor Head Injury in the Emergency Department (e174)
Edward Melnick, Erik Hess, George Guo, Maggie Breslin, Kevin Lopez, Anthony Pavlo, Fuad Abujarad, Seth Powsner, Lori Post. . . . 347

The Virtual Care Climate Questionnaire: Development and Validation of a Questionnaire Measuring Perceived Support for Autonomy in a Virtual Care Setting (e155)
Eline Smit, Alexandra Dima, Stephanie Immerzeel, Bas van den Putte, Geoffrey Williams. . . . 359

The Development, Validation, and User Evaluation of Foodbook24: A Web-Based Dietary Assessment Tool Developed for the Irish Adult Population (e158)
Claire Timon, Richard Blain, Breige McNulty, Laura Kehoe, Katie Evans, Janette Walton, Albert Flynn, Eileen Gibney. . . . 375

Design and Evaluation of a Computer-Based 24-Hour Physical Activity Recall (cpar24) Instrument (e186)
Simone Kohler, Gundula Behrens, Matthias Olden, Sebastian Baumeister, Alexander Horsch, Beate Fischer, Michael Leitzmann. . . . 394

Effects of Improving Primary Health Care Workers' Knowledge About Public Health Services in Rural China: A Comparative Study of Blended Learning and Pure E-Learning (e116)
Xingxin Zhan, Zhixia Zhang, Fang Sun, Qian Liu, Weijun Peng, Heng Zhang, Weirong Yan. . . . 411

Effects of Using Child Personas in the Development of a Digital Peer Support Service for Childhood Cancer Survivors (e161)
Pontus Wärnestål, Petra Svedberg, Susanne Lindberg, Jens Nygren. . . . 422

Development of Trust in an Online Breast Cancer Forum: A Qualitative Study (e175)
Melanie Lovatt, Peter Bath, Julie Ellis. . . . 435

Enhancing Comparative Effectiveness Research With Automated Pediatric Pneumonia Detection in a Multi-Institutional Clinical Repository: A PHIS+ Pilot Study (e162)
Stephane Meystre, Ramkiran Gouripeddi, Joel Tieder, Jeffrey Simmons, Rajendu Srivastava, Samir Shah. . . . 448

Pharma Websites and "Professionals-Only" Information: The Implications for Patient Trust and Autonomy (e178)
Mark Graber, Eliyakim Hershkop, Rachel Graber. . . . 459

Adherence to AHA Guidelines When Adapted for Augmented Reality Glasses for Assisted Pediatric Cardiopulmonary Resuscitation: A Randomized Controlled Trial (e183)
Johan Siebert, Frederic Ehrler, Alain Gervaix, Kevin Haddad, Laurence Lacroix, Philippe Schrurs, Ayhan Sahin, Christian Lovis, Sergio Manzano. . . . 465

Identifying and Understanding the Health Information Experiences and Preferences of Caregivers of Individuals With Either Traumatic Brain Injury, Spinal Cord Injury, or Burn Injury: A Qualitative Investigation (e159)
Nathan Coffey, James Cassese, Xinsheng Cai, Steven Garfinkel, Drasti Patel, Rebecca Jones, Dahlia Shaewitz, Ali Weinstein. . . . 478

Effects of eHealth Literacy on General Practitioner Consultations: A Mediation Analysis (e166)
Peter Schulz, Mary Fitzpatrick, Alexandra Hess, Lynn Sudbury-Riley, Uwe Hartung. . . . 488

Physician Rating Websites: What Aspects Are Important to Identify a Good Doctor, and Are Patients Capable of Assessing Them? A Mixed-Methods Approach Including Physicians' and Health Care Consumers' Perspectives (e127)
Fabia Rothenfluh, Peter Schulz. . . . 500

XSL•FO
RenderX

Mobile App for Treatment of Stress Urinary Incontinence: A Cost-Effectiveness Analysis (e154)
Malin Sjöström, Lars Lindholm, Eva Samuelsson. . . . 513

Virtual Visits and Patient-Centered Care: Results of a Patient Survey and Observational Study (e177)
Kimberlyn McGrail, Megan Ahuja, Chad Leaver. . . . 554

Using a Medical Intranet of Things System to Prevent Bed Falls in an Acute Care Hospital: A Pilot Study (e150)
Henri Balaguera, Diana Wise, Chun Ng, Han-Wen Tso, Wan-Lin Chiang, Aimee Hutchinson, Tracy Galvin, Lee Hilborne, Cathy Hoffman, Chi-Cheng Huang, C Wang. . . . 566

A Smart Toy to Enhance the Decision-Making Process at Children's Psychomotor Delay Screenings: A Pilot Study (e171)
María Gutiérrez García, María Martín Ruiz, Diego Rivera, Laura Vadillo, Miguel Valero Duboy. . . . 577

The Development of an Interactive Voice Response Survey for Noncommunicable Disease Risk Factor Estimation: Technical Assessment and Cognitive Testing (e112)
Dustin Gibson, Brooke Farrenkopf, Amanda Pereira, Alain Labrique, George Pariyo. . . . 632

Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review (e139)
Dustin Gibson, Amanda Pereira, Brooke Farrenkopf, Alain Labrique, George Pariyo, Adnan Hyder. . . . 641

Building the Evidence Base for Remote Data Collection in Low- and Middle-Income Countries: Comparing Reliability and Accuracy Across Survey Modalities (e140)
Abigail Greenleaf, Dustin Gibson, Christelle Khattar, Alain Labrique, George Pariyo. . . . 653

Health Evaluation and Referral Assistant: A Randomized Controlled Trial of a Web-Based Screening, Brief Intervention, and Referral to Treatment System to Reduce Risky Alcohol Use Among Emergency Department Patients (e119)
Brianna Haskins, Rachel Davis-Martin, Beau Abar, Brigitte Baumann, Tina Harralson, Edwin Boudreaux. . . . 665

Computer Administered Safety Planning for Individuals at Risk for Suicide: Development and Usability Testing (e149)
Edwin Boudreaux, Gregory Brown, Barbara Stanley, Rajani Sadasivam, Carlos Camargo Jr, Ivan Miller. . . . 677

Reviews

Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review (e176)
Rebecca Grist, Joanna Porter, Paul Stallard. . . . 142

Toward the Design of Evidence-Based Mental Health Information Systems for People With Depression: A Systematic Literature Review and Meta-Analysis (e191)
Fabian Wahle, Lea Bollhalder, Tobias Kowatsch, Elgar Fleisch. . . . 262

Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer (e172)
Peter Hanlon, Luke Daines, Christine Campbell, Brian McKinstry, David Weller, Hilary Pinnock. . . . 310

XSL•FO
RenderX

Viewpoints

Accelerating Digital Mental Health Research From Early Design and Creation to Successful Implementation and Sustainment (e153)
David Mohr, Aaron Lyon, Emily Lattie, Madhu Reddy, Stephen Schueller. . . . 524

More Than a Text Message: Dismantling Digital Triggers to Curate Behavior Change in Patient-Centered Health Interventions (e147)
Frederick Muench, Amit Baumel. . . . 540

Noncommunicable Disease Risk Factors and Mobile Phones: A Proposed Research Agenda (e133)
Adnan Hyder, Adaeze Wosu, Dustin Gibson, Alain Labrique, Joseph Ali, George Pariyo. . . . 591

Moving the Agenda on Noncommunicable Diseases: Policy Implications of Mobile Phone Surveys in Low and Middle-Income Countries (e115)
George Pariyo, Adaeze Wosu, Dustin Gibson, Alain Labrique, Joseph Ali, Adnan Hyder. . . . 601

Ethics Considerations in Global Mobile Phone-Based Surveys of Noncommunicable Diseases: A Conceptual Exploration (e110)
Joseph Ali, Alain Labrique, Kara Gionfriddo, George Pariyo, Dustin Gibson, Bridget Pratt, Molly Deutsch-Feldman, Adnan Hyder. . . . 612

Policy Proposal

Health Surveys Using Mobile Phones in Developing Countries: Automated Active Strata Monitoring and Other Statistical Considerations for Improving Precision and Reducing Biases (e121)
Alain Labrique, Emily Blynn, Saifuddin Ahmed, Dustin Gibson, George Pariyo, Adnan Hyder. . . . 623

XSL•FO
RenderX

Editorial

Using mHealth to Predict Noncommunicable Diseases: A Public Health Opportunity for Low- and Middle-Income Countries

Ellen Rosskam[1,2], MPH, PhD; Adnan A Hyder[3,4], MPH, MD, PhD

[1]ER Global Consult, La Croix-de-Rozon, Switzerland

[2]Visiting Professor, School of Health Sciences, University of Massachusetts, Lowell, MA, United States

[3]Bloomberg School of Public Health, Department of International Health, Health Systems Program, Johns Hopkins University, Baltimore, MD, United States

[4]Bloomberg School of Public Health, International Injury Research Unit, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Ellen Rosskam, MPH, PhD
ER Global Consult
34D Route du Prieur, Landecy
La Croix-de-Rozon, 1257
Switzerland
Phone: 41 223476846
Fax: 41 00000000
Email: ellenrosskam@gmail.com

Related Article:

Comment in: http://www.jmir.org/2017/5/e137/

(J Med Internet Res 2017;19(5):e129) doi:10.2196/jmir.7593

KEYWORDS

mHealth; low- and middle-income countries; noncommunicable diseases; research agenda; population health surveys

Nearly 70% of the 56 million deaths that took place globally in 2012 were due to noncommunicable diseases (NCDs), in particular, cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes, and nearly two-thirds of all NCD deaths took place in low- and middle-income countries (LMICs) [1]. If effective steps are not taken to curb the epidemic, deaths due to NCDs are projected to rise exponentially in the coming decade [2]. Key risk factors responsible for a majority of NCDs include tobacco use, unhealthy diet, sedentary lifestyle, and excessive use of alcohol. With targeted action, these behavioral risk factors have demonstrated potential to be modified [3] to reduce NCDs and improve population health. Reducing NCDs, particularly in the world's poorest countries, can lead to increases in equity and socioeconomic development while reducing poverty due to ill health and promoting sustainable development and social justice.

Key to global efforts to prevent and control NCDs is national surveillance. A promising approach increasingly being explored for public health surveillance involves mobile phones. A nascent yet emergent field, mHealth, describes medical and public health activities that leverage the global proliferation of cellular networks and mobile phone ownership or access to improve population health outcomes. There are nearly 7.5 billion wireless phone subscriptions globally, with the majority (78%) in LMICs [4]. Global connectivity to cellular networks can make large proportions of a population accessible through their mobile phones. In response to the increasing NCD disease burden, the intersecting need for NCD data in LMICs and the near-universal population access to mobile phones in a growing number of countries presents an opportunity for public health.

This special Theme Issue of JMIR offers a step forward in documenting what is known about surveillance of risk factors for NCDs in LMICs using mobile phone surveys (MPS). The evidence illustrates that the state-of-the-art is sufficient to roll out population-level surveys in LMICs using mobile phone platforms while paying careful attention to issues such as ethics, methodology, and turning results into practice. The results offer guidance for policy and practice.

The article, "Noncommunicable Disease Risk Factors and Mobile Phones: A Proposed Research Agenda," proposes a research and development agenda for NCD risk factors and MPS [5]. The goal of the proposed agenda is to help standardize operating procedures for MPS, which will allow for comparisons of NCD risk factors within and across sites and over time. The potential is explored for MPS to collect such data, review key research issues, and introduce a multicountry effort that seeks to partly respond to this public health challenge. It is hoped that

XSL•FO
RenderX

the proposed research agenda will catalyze a global dialogue and action to enhance the use of MPS for NCDs and potentially other public health risk factor surveillance.

Limited evidence exists on the comparative effectiveness of MPS modalities in LMICs although a variety of options are available. "Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review" reviews the current landscape of MPS being used for population-level data collection in LMICs, specifically through the use of short message service, interactive voice response (IVR), and computer-assisted telephone interview survey modalities [6]. From the articles identified of MPS use to collect population estimates across a range of topics, results reveal that the state of MPS to collect population-level estimates of health and other indicators is a nascent field, indicating the need for more research.

The methodological approach used to test the use of MPS for NCDs is described in "Evaluation of Mechanisms to Improve Performance of Mobile Phone Surveys: A Research Protocol" [7]. Using microtrials, a set of future studies that will help enhance the efficiency and technical effectiveness of MPS is proposed for LMICs. The authors assess the effect of factors such as incentive timing and structure, survey introduction characteristics, different sampling frames, and survey modality on key survey metrics such as survey response, completion, and attrition rates.

Further investigating the literature, "Building the Evidence Base for Remote Data Collection in Low- and Middle-Income Countries: Comparing Reliability and Accuracy Across Survey Modalities" reviews findings that compare a mode of remote data collection to at least one other mode [8]. The synthesis examines MPS mode effects on the reliability and accuracy of results. Findings show, for example, that remote data collection consistently elicited higher reports of socially nondesirable behaviors compared to in-person data collection. The review reveals the need for additional studies that compare reliability and construct validity across survey modalities.

IVR has the potential to expand current surveillance coverage and data collection. Two rounds of IVR pilot testing in Baltimore, Maryland, revealed that most participants felt this type of survey would lead to more honest, accurate responses than face-to-face questionnaires, especially for sensitive topics. In the pilot tests, participants indicated a clear comprehension of the IVR-administered questionnaire and that the IVR platform was user-friendly. Described in "The Development of an Interactive Voice Response Survey for Noncommunicable Disease Risk Factor Estimation: Technical Assessment and Cognitive Testing," the authors conclude that formative research and cognitive testing of the questionnaire are needed for deployment in LMICs [9].

The near-ubiquitous ownership of phones in LMICs, high population mobility, and low cost demand a reexamination of statistical recommendations for MPS, especially when surveys are automated. In "Health surveys using mobile phones in developing countries: automated active strata monitoring and other statistical considerations for improving precision and reducing biases," methods are proposed to reduce estimate bias and to adjust for selectivity due to mobile ownership [10]. The authors describe using automated active strata monitoring (AASM) to improve representativeness of the sample distribution to that of the source population. They conclude that although some statistical challenges remain, MPS represents a promising emerging means for population-level data collection in LMICs.

The increasing use of MPS in LMICs brings forth a cluster of ethical challenges. The existing literature regarding the ethics of mobile or digital health, however, mainly focuses on the use of technologies in high-income countries and does not consider the specific ethical issues associated with the conduct of MPS for NCD risk factor surveillance in LMICs. In "Ethics Considerations in Global Mobile Phone-Based Surveys of Noncommunicable Diseases: A Conceptual Exploration," the authors explored central ethics issues in this domain, including identifying the nature of the activity, stakeholder engagement, appropriate design, anticipating and managing potential harms and benefits, consent, reaching intended respondents, data ownership, access and use, and ensuring LMIC sustainability [11]. The authors call for future work to develop a broad conceptual framework for the ethical, legal, and societal issues associated with MPS for NCD risk factors. They further point to the need for guidance documents to identify key issues, outline pros and cons of options available to stakeholders for each issue, review additional points to consider, and provide references to resources relevant to each issue. In order to begin to address the various needs, the researchers hope to establish a global working group inclusive of experts in ethics, mHealth survey implementation, regulatory oversight and policy, public health, social science, and MPS platform development.

The article, "Moving the Agenda on Noncommunicable Diseases: Policy Implications of Mobile Phone Surveys in Low- and Middle-Income Countries," presents the special challenges for policy makers [12]. The article discusses potential benefits of MPS for developing, implementing, and evaluating NCD prevention and control policies. It includes an overview of major global commitments to NCD prevention and control as well as an exploration of how countries can translate these commitments into policy action at the national level. Potential benefits of MPS are discussed, including cost benefits of MPS for informing NCD policy actions compared to using traditional household surveys, timeliness of assessments to feed into policy and planning cycles, tracking progress of interventions, timely course correction for suboptimal or noneffective interventions, and assessing fairness in financial contribution and financial risk protection for those affected by NCDs in the spirit of universal health coverage, inter alia. The authors demonstrate how MPS can become a powerful tool for collecting population-based data to inform policies that address key public health challenges such as NCDs. Further research in real-life settings will help to provide additional realistic world experiences.

This special issue of JMIR offers a step forward in benchmarking what is known and what is possible to know using MPS for data collection and surveillance systems. These results offer guidance for research expectations and opportunities to understand and curb the rise of NCDs in LMICs. Additional next steps are foreseen to continue documenting empirical

XSL•FO
RenderX

experiences of MPS use in LMICs to collect risk factor data on NCDs, engaging with global bodies toward the development of a research agenda, establishing a global working group of experts to address the ethical issues surrounding MPS use in LMICs, and working with international and national level policy-makers to create a comparative framework for turning results into policy and practice.

Conflicts of Interest

None declared.

References

1. Global Status Report of Noncommunicable Diseases 2014. Geneva: World Health Organization; 2014.
2. Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020. Geneva: World Health Organization; 2013.
3. Global Status Report on Noncommunicable Diseases 2010: Description of the Global Burden of NCDs, their Risk Factors, and Determinants. Geneva: World Health Organization; 2011.
4. International Telecommunication Union. Statistics URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2016-12-12] [WebCite Cache ID 6mgkxF0WI]
5. Hyder A, Wosu A, Gibson D, Labrique A, Ali J, Pariyo G. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246]
6. Gibson D, Pereira A, Farrenkopf B, Pariyo G, Labrique A, Hyder A. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. J Med Internet Res 2017;19(5):e139 [FREE Full text] [doi: 10.2196/jmir.7428]
7. Gibson D, Pariyo G, Wosu A, Greenleaf A, Ali J, Ahmed S, et al. Evaluation of mechanisms to improve performance of mobile phone surveys: a research protocol. JMIR Res Protoc 2017;6(5) [FREE Full text] [doi: 10.2196/resprot.7534]
8. Greenleaf A, Gibson DG, Khattar C, Labrique AB, Pariyo GW. Building the evidence base for remote data collection in low- and middle-income countries: comparing reliability and accuracy. J Med Internet Res 2017;19(5):e140 [FREE Full text] [doi: 10.2196/jmir.7331]
9. Gibson DG, Farrenkopf B, Pereira A, Labrique AB, Pariyo GW. The development of an interactive voice response survey for noncommunicable disease risk factor estimation: technical assessment and cognitive testing. J Med Internet Res 2017;19(5):e112 [FREE Full text] [doi: 10.2196/jmir.7340]
10. Labrique AB, Blynn E, Ahmed S, Gibson DG, Pariyo GW, Hyder A. Health surveys using mobile phones in developing countries: automated active strata monitoring and other statistical considerations for improving precision and reducing biases. J Med Internet Res 2017;19(5):e121 [FREE Full text] [doi: 10.2196/jmir.7329]
11. Ali J, Labrique AB, Gionfriddo K, Pariyo GW, Gibson DG, Pratt B, et al. Ethics considerations in global mobile phone-based surveys of noncommunicable diseases: a conceptual exploration. J Med Internet Res 2017;19(5):e110 [FREE Full text] [doi: 10.2196/jmir.7326]
12. Pariyo GW, Wosu AC, Gibson DG, Labrique AB, Ali J, Hyder A. Moving the agenda on noncommunicable diseases: policy implications of mobile phone surveys in low- and middle-income countries. J Med Internet Res 2017;19(5):e115 [FREE Full text] [doi: 10.2196/jmir.7302]

Abbreviations

IVR: interactive voice response
LMIC: low- and middle-income country
MPS: mobile phone surveys
NCD: noncommunicable disease
CATI: computer-assisted telephone interview
SMS: short message service

XSL•FO
RenderX

Edited by G Eysenbach; submitted 26.02.17; peer-reviewed by G Humphrey; comments to author 02.04.17; revised version received 11.04.17; accepted 12.04.17; published 05.05.17.

Please cite as:
Rosskam E, Hyder AA
Using mHealth to Predict Noncommunicable Diseases: A Public Health Opportunity for Low- and Middle-Income Countries
J Med Internet Res 2017;19(5):e129
URL: http://www.jmir.org/2017/5/e129/
doi:10.2196/jmir.7593
PMID:28476727

©Ellen Rosskam, Adnan A Hyder. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Guest Editorial

Leveraging Mobile Phones for Monitoring Risks for Noncommunicable Diseases in the Future

Jennifer A Ellis[1], PhD

Bloomberg Philanthropies, New York, NY, United States

Corresponding Author:
Jennifer A Ellis, PhD
Bloomberg Philanthropies
25 E 78th St.
New York, NY,
United States
Phone: 1 212 205 0129
Fax: 1 000 000 0000
Email: Jennifer@bloomberg.org

Related Article:

Comment on: http://www.jmir.org/2017/5/e129/

(J Med Internet Res 2017;19(5):e137) doi:10.2196/jmir.7622

KEYWORDS

mHealth; low- and middle-income countries; noncommunicable diseases; health systems strengthening

> *One of the biggest contributors to preventable deaths isn't a health problem but a record-keeping problem—and it is one that can be solved. [1]*

Noncommunicable diseases (NCDs) account for two-thirds of deaths globally. The World Health Organization estimates that 67% of 56 million deaths that took place globally in 2012 were due to NCDs, in particular cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes [2]. Three-quarters of these deaths occurred in low- and middle-income countries (LMICs) and nearly half of deaths occurred in persons younger than 70 years [2]. The World Health Organization projects a rise in the number of NCD deaths from 36 million in 2008 to 55 million by 2030, of which 15.4 million are projected to be in people younger than 70 years, if effective steps are not taken to curb the pandemic [3,4].

Four key risk factors which have been stated to be responsible for a majority of NCDs are tobacco use, sodium intake, sedentary lifestyle, and excessive use of alcohol—all behavioral and largely modifiable [5]. These factors, as well as loss of employment due to NCD-related disability and the long duration and complexity of NCD treatment, pose additional challenges to poverty reduction and sustainable development [6].

In 2015, Bloomberg Philanthropies and the Australian Government launched an ambitious new effort to redress the serious need for improved data about why people die. The Data for Health Initiative [7] is a US $100 million program that aims to provide better health data, accelerate global progress, and provide technical resources for more than 1 billion people in more than 20 LMICs in Africa, Asia, and Latin America. Reliable data are essential to make effective health policies, measure the success of public health interventions, and prioritize research. To reach the Initiative's goal, Bloomberg Philanthropies has brought together a group of leading partners that includes the Johns Hopkins University Bloomberg School of Public Health, the CDC Foundation, Vital Strategies, the University of Melbourne, and the World Health Organization.

> *The day-to-day things that are killing people in large numbers are largely non communicable diseases and injuries... Yet governments, donors, and global health leadership are not responding... [8].*

Serious gaps exist between health care spending and need—particularly in LMICs. In order to understand where investment has the most potential, we need detailed and dependable information about what is killing people. Illnesses like cancer and diabetes are becoming less fatal in higher income countries thanks to investment in prevention and treatment. Without the same in LMICs, chronic diseases are projected to cost US $21.3 trillion over the next two decades—and countless lives [8].

Many LMICs seek cost-effective methods to obtain timely and quality NCD risk factor data that can be used to inform resource allocation, policy development, and assist in the evaluation of NCD trends over time. Mobile phone technology is omnipresent in many LMICs and presents an untapped opportunity for using population-level health surveys. The Data for Health Initiative is working to assess, harness, and roll out the use of mobile

XSL•FO
RenderX

phone technology in LMICs as a cost-effective method for the rapid collection of quality NCD risk factor data. Countries engaged in the Initiative expressed interest in improving their public health data and present particularly great potential for gains. There are, however, critical gaps in understanding the ways by which mobile phone surveys could be a cutting-edge aid in the collection of NCD data in LMICs.

This special Theme Issue of JMIR addresses these critical gaps through an examination of the relevant extant literature, analysis of the myriad ethical issues surrounding the use of mobile phones to collect personal health data in LMICs, methodological challenges, conceptual challenges of turning data into international- and national-level policies, and the efforts needed for using mobile phone technology to create a research agenda for NCDs, public health interventions, and strengthening health systems in LMIC contexts.

Current methods of data collection in LMICs rely on the use of household surveys to monitor disease burden in countries, prioritize resource allocation, and evaluate public health policies. In LMICs, household surveys typically rely on face-to-face interviews conducted at the respondent's household. Maintaining up-to-date data based on household surveys is difficult, however, since household surveys bear high costs, particularly in personnel and transportation [9], which means such surveys cannot be conducted too frequently. Additionally, household surveys require considerable amounts of time for data collection, data management, and data analysis which impedes the speed at which data become publically available. Mobile phone technology presents a potentially viable option for more frequent surveillance of population health, one that will permit more timely evaluation of implemented public health policies and response to public health emergencies.

To address the high costs and time requirements associated with household surveys, higher-income countries have developed and employed telephone and mobile phone surveys to collect population-level estimates of health [9]. The global spread of mobile phone ownership and access provides new opportunities to leverage mobile health technologies and communication channels to revolutionize current methods of data collection in LMICs. Mobile phone surveys involve interviewing respondents over their own personal mobile phone through the use of short message service, interactive voice response, and computer-assisted telephone interviews. This JMIR Theme Issue explores and compares the challenges and benefits of each of these survey modalities as well as the current landscape of mobile phone survey technology being used for population-level data collection in LMICs.

References

1. Bloomberg MR, Bishop J. Understanding death, extending life. Lancet 2015 Oct 17;386(10003):e18-e19. [doi: 10.1016/S0140-6736(15)00400-6] [Medline: 26434355]
2. Global Status Report of Noncommunicable Diseases—2014. Geneva: World Health Organization; 2014.
3. Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020. Geneva: World Health Organization; 2013.
4. Alwan A, Maclean DR, Riley LM, d'Espaignet ET, Mathers CD, Stevens GA, et al. Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. Lancet 2010 Nov 27;376(9755):1861-1868. [doi: 10.1016/S0140-6736(10)61853-3] [Medline: 21074258]
5. Global status report on noncommunicable diseases 2010: Description of the global burden of NCDs, their RF. Global Status Report on Noncommunicable Diseases 2010: Description of the Global Burden of NCDs, their Risk Factors and Determinants. Geneva: World Health Organization; 2011.
6. Bloom D, Cafiero E, Jane-Llopis E, Abrahams-Gessel S, Bloom L, Fathima S, et al. The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum; 2011.
7. Smith J. The Atlantic. Mystery killers: the unknown causes of mortality around the globe URL: http://www.theatlantic.com/magazine/archive/2015/03/mystery-killers/385001/ [accessed 2017-01-08] [WebCite Cache ID 6nN062ZGP]
8. McKay B. The Wall Street Journal. 2015 Mar 22. Michael Bloomberg Backs Health-Data Push URL: http://www.wsj.com/articles/michael-bloomberg-backs-health-data-push-1427052799 [accessed 2017-01-06] [WebCite Cache ID 6nN0Q94Yo]
9. Bloomberg Philanthropies. Data for Health Initiative. URL: https://www.bloomberg.org/program/public-health/data-health/ [accessed 2017-04-17] [WebCite Cache ID 6pnN3g3Zm]

Abbreviations

NCD: noncommunicable disease
LMIC: low- and middle-income country

XSL•FO
RenderX

Edited by G Eysenbach; submitted 01.03.17; this is a non–peer-reviewed article;accepted 20.03.17; published 05.05.17.

Please cite as:
Ellis JA
Leveraging Mobile Phones for Monitoring Risks for Noncommunicable Diseases in the Future
J Med Internet Res 2017;19(5):e137
URL: http://www.jmir.org/2017/5/e137/
doi:10.2196/jmir.7622
PMID:28476721

©Jennifer A Ellis. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Fifteen Challenges in Establishing a Multidisciplinary Research Program on eHealth Research in a University Setting: A Case Study

Helena Grönqvist[1*], PhD; Erik Martin Gustaf Olsson[1*], Lic Psych, PhD; Birgitta Johansson[1,2], RN, PhD; Claes Held[1,3,4], MD, PhD; Jonas Sjöström[1,5], PhD; Annika Lindahl Norberg[1,6], Lic Psych, PhD; Emma Hovén[1], PhD; Robbert Sanderman[7,8], Lic Psych, PhD; Theo van Achterberg[9], RN, PhD; Louise von Essen[1], Lic Psych, PhD

[1]U-CARE, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[2]Department of Immunology, Genetics and Pathology, Experimental and Clinical Oncology, Uppsala University, Uppsala, Sweden
[3]Department of Medical Sciences, Cardiology, Uppsala University, Uppsala, Sweden
[4]Uppsala Clinical Research Center, Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden
[5]Department of Informatics and Media, Campus Gotland, Uppsala University, Uppsala, Sweden
[6]Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden
[7]Department of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands
[8]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands
[9]Academic Centre for Nursing and Midwifery, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium
[*]these authors contributed equally

Corresponding Author:
Erik Martin Gustaf Olsson, Lic Psych, PhD
U-CARE
Department of Women's and Children's Health
Uppsala University
Box 572
Uppsala, 75123
Sweden
Phone: 46 184716663
Fax: 46 184716675
Email: erik.olsson@kbh.uu.se

Abstract

Background: U-CARE is a multidisciplinary eHealth research program that involves the disciplines of caring science, clinical psychology, health economics, information systems, and medical science. It was set up from scratch in a university setting in 2010, funded by a governmental initiative. While establishing the research program, many challenges were faced. Systematic documentation of experiences from establishing new research environments is scarce.

Objective: The aim of this paper was to describe the challenges of establishing a publicly funded multidisciplinary eHealth research environment.

Methods: Researchers involved in developing the research program U-CARE identified challenges in the formal documentation and by reflecting on their experience of developing the program. The authors discussed the content and organization of challenges into themes until consensus was reached.

Results: The authors identified 15 major challenges, some general to establishing a new research environment and some specific for multidisciplinary eHealth programs. The challenges were organized into 6 themes: Organization, Communication, Implementation, Legislation, Software development, and Multidisciplinarity.

Conclusions: Several challenges were faced during the development of the program and several accomplishments were made. By sharing our experience, we hope to help other research groups embarking on a similar journey to be prepared for some of the challenges they are likely to face on their way.

(J Med Internet Res 2017;19(5):e173) doi:10.2196/jmir.7310

XSL•FO
RenderX

KEYWORDS

organization and administration; eHealth; interdisciplinary studies

Introduction

Background

Publicly funded multidisciplinary eHealth research environments face challenges seldom described systematically in the literature [1]. This paper aims to start to fill this gap. In the following, we describe both foreseen and unforeseen challenges that arose when a relatively large research program studying eHealth solutions for people suffering emotionally with serious somatic illnesses was set up. Some of the challenges discussed below are general for creating research environments, whereas others are more specific for multidisciplinary eHealth projects.

Societies aim to support the research that best meets their needs at the time. In Sweden, based on a proposal from the government (proposition 2008/09:50), a plan to support large strategic research environments was launched from the main public research funding agencies. The overall aim was to strengthen Sweden's position as a research nation and thereby increase its scientific competitiveness in a globalized world. It also explicitly promoted multidisciplinary research on eHealth. The Uppsala University Psychosocial Care Program (U-CARE) was one of the 43 programs that were funded for 5 years based on this proposition. Thereafter, the funding of this program has been prolonged by 1 year at a time. U-CARE was, and still is, a multidisciplinary program involving the academic disciplines of caring science, clinical psychology, health economics, information systems, and medical sciences. U-CARE's main vision was, and still is, to increase cost-effective access to participatory mental health care in connection with somatic illness by using welfare technology. An online platform, the U-CARE-portal, was developed in house to support provision of self-care, care, and psychological treatment. It was designed to serve as a research backbone for the U-CARE program with built-in features such as stratified randomization, flexible data collection, logging of patient and therapist behaviors, automatic reminders, and providing an overview on study progress (eg, number of included and randomized participants).

The background for the governmental support for U-CARE was the rapid development in information and communication technology and the role of Internet, which has influenced how we get information and communicate about health. Interactive eHealth programs for behavioral change have gradually become more available and have attained research-based support of its efficacy in several areas [2]. However, there are also several new difficulties emerging as these new tools for health care become common. For example, eHealth solutions offered by public and commercial actors are now becoming so common and diverse that the quality is difficult to manage, and the shift from face-to-face to Web-based communication in health interventions may result in that even the appropriate research and evaluation methodology has to be reconsidered [3,4]. Implementation, legal matters, ethics, and integrity issues as well as technical and practical concerns are other areas where eHealth faces new and different challenges than the traditional health care does [5,6]. There is certainly a need for sustainable, multidisciplinary research environment in this area, as the potential benefits from using eHealth on a large scale for a society are many and large [7].

U-CARE's Project Goals

Five goals to be achieved by 2014 were specified in the original grant application for the U-CARE program. As written in the 2009 application, the goals were to (1) establish an internationally competitive, innovative psychosocial research platform that will be applied within pediatric oncology, adult oncology, and cardiology; (2) build a high-quality research-based, transdisciplinary education within the field of psychosocial health care and establish a National Graduate School in Innovative Psychosocial Health Care Research; (3) provide stimulating and challenging career opportunities for young researchers; (4) attract major external funding from the EU Framework Programs and Swedish research foundations; and (5) establish a Centre of Excellence, the Uppsala Care Centre (U-CARE Centre) for strong transdisciplinary research and research-based education within the field of psychosocial health care.

The original application described eHealth as a promising modality to be explored with the above goals in mind. Three randomized controlled trials (RCTs) evaluating Internet-supported interventions in the areas of pediatric oncology, adult oncology, and cardiology were already outlined in the application.

Objective

The aim of this article was to describe challenges when establishing a publicly funded multidisciplinary eHealth research environment. The challenges encountered were (1) general challenges for developing new research programs such as the political environment, communication strategies and research relevance; and (2) challenges specific to eHealth and multidisciplinary programs such as software development and dealing with research cultures from several disciplines.

Methods

Design

The case for this study is the aforementioned U-CARE program. Nota bene, this paper is not an evaluation of the project's main aims. Instead, it is a qualitative description of the challenges faced during the development process. Goal attainment for the project will only be briefly described to serve as a contextual frame of reference for the challenges described.

Data Collection and Analysis

All senior and junior researchers working with the U-CARE program until 2014, as well as two of the project's scientific advisors were invited to work with this paper. All besides one accepted the invitation and are consequently coauthors of this paper. Data were extracted from meeting protocols from the study coordination group, the program executive committee,

XSL•FO
RenderX

the yearly meetings with the international Scientific Advisory Board (SAB), evaluation reports to funding agencies and the University administration, and from reflecting on own experiences. First, all coauthors working with the U-CARE program individually identified challenges in the documentation and wrote a summary of the key issues. Thereafter, the authors met and made a preliminary categorization of the identified challenges. At a second meeting, content, boundaries, and categorization of preliminary challenges were discussed and consensus was reached. Subsequently, the two first authors scrutinized the key issues, separating challenges general to developing research programs and challenges specific to multidisciplinary eHealth.

Results and Discussion

The Challenges

In this section, we describe the identified challenges and, in some cases, how they were handled. First, 3 themes with general challenges faced when setting up the program are presented. Thereafter, challenges more specific to eHealth and the multidisciplinary nature of the program are described, organized into 3 themes. A summary of the challenges can be found in Table 1.

Organization

The U-CARE program was organized at the Disciplinary Domain of Medicine and Pharmacy at Uppsala University, above department level and was hosted by a relatively small existing research group at the Department of Public Health and Caring Sciences. It also included coworkers from other research groups, departments, and universities than from the host research group. A completely new organization was required for the strategic environment according to instructions from the university management matching the requirements from the funding agency while the program was part of the department's organization. This meant that the coworkers had obligations both to their respective departments as employees, PhD students, teachers, or researchers, and to the new overarching organization. Sometimes, there were dual instructions. For example in U-CARE, the executive committee is responsible for the sanctioning of the PhD student's study plans and for reviewing their progress. This resulted in redundancy as the departments already had detailed routines for this. An overflow of administrative duties may reduce academic output for example in terms of publications.

Figure 1 illustrates the organization of the U-CARE program. The steering committee, executive committee, and SAB included members with expertise in caring science, clinical psychology, information systems, and health economics to facilitate the multidisciplinary work in the U-CARE program. The management team consisted of the Program Director, Program Coordinator, Information Technology Coordinator, and Research Coordinator. Work packages were organized around the RCTs that were already outlined in the application. These work packages as well as the associated studies (see below) were represented in the study coordination group meetings, which aimed at communicating the progress of each study and enabling synergy effects.

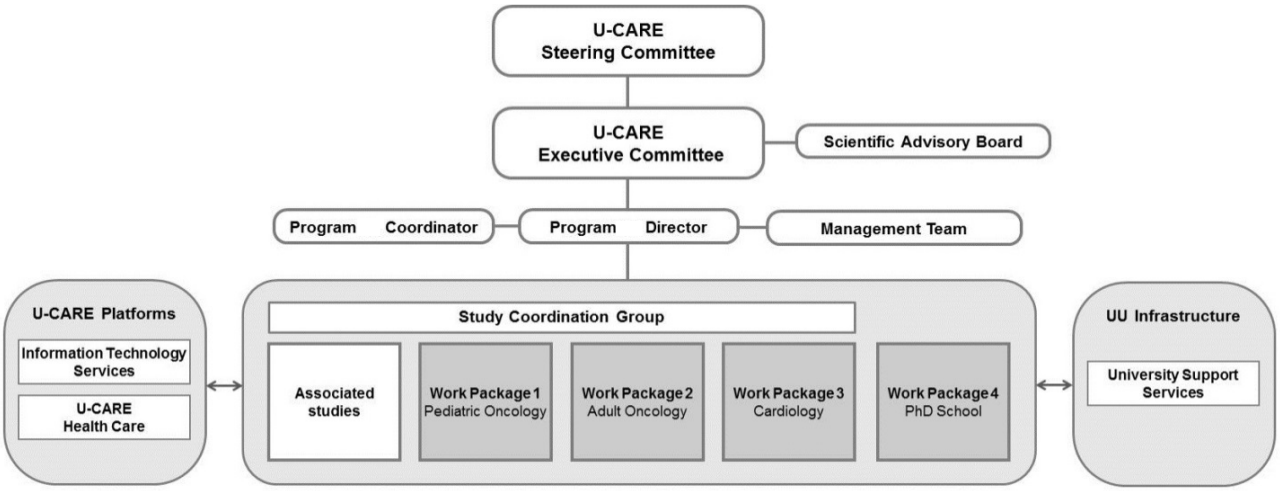
Studies initiated by other research groups were associated to U-CARE in order to benefit from the U-CARE portal. A more business-like practice emerged over time, where the associated studies became clients that purchased software as a service from U-CARE. The organization thus had to develop new knowledge and setup processes to act effectively as a service provider.

Table 1. Challenges in establishing a multidisciplinary research program on eHealth research.

General or specific to multidisciplinary eHealth projects	Themes	Challenges
General	Organization	1. The appropriate organization challenge
		2. The strategic support challenge
		3. The responsive organization challenge
		4. The continuity of productivity challenge
	Communication	5. The internal communication and documentation challenge
		6. The external communication challenge
	Implementation	7. The material sharing challenge
		8. The stakeholder involvement challenge
		9. The public involvement challenge
Specific	Legislation	10. The professor privilege challenge
		11. The competing legislation challenge
	Software development	12. The mutual understanding of software requirements challenge
		13. The software development documentation challenge
	Multidisciplinarity	14. The discipline openness challenge
		15. The challenge of a shared theoretical framework

XSL•FO
RenderX

Figure 1. Organization of U-CARE.



The appropriate organization challenge (#1) was to make the U-CARE program fit requirements from both the host university department and the funding agency as efficient as possible and with the least possible redundancy. The identification of the redundant or otherwise noneffective processes and routines was the key to necessary changes or adjustments. Our recommendation is that clarifications of responsibilities to the program and the host department should be made as soon as possible. One intention of the strategic research area initiative was to provide long-term funding to increase stability of the programs and relieve them from some of the demands normally put when applying for short-term project funding. Furthermore, the administration of a large program is demanding and time-consuming especially in the upstart phase. To facilitate the process, a program coordinator was hired at inception to support the program director. As the organization grew, responsibilities were delegated and duties were distributed within the group. Establishing a new research program is a long-term investment and takes time and effort to build the most efficient organization and administrative routines.

The strategic support challenge (#2) was to attract support from the head of the department, the vice-chancellor, and officials at university management level. A research group, receiving a large multidisciplinary grant, needs support on a strategic level to establish career opportunities as well as provide appropriate infrastructure aligned with the program. Such support is necessary to establish a successful and sustainable program that does not come easy. When this was being written, the program has moved to another host department in an attempt to improve the organizational conditions for the program.

A large program needs to be organized rigorously to meet the high demands on quality and transparency from both the funder and the scientific society. The responsive organization challenge (#3) was to make the organization flexible and responsive to new ideas and a changing environment while keeping scientific rigor and quality. This challenge was met in a very general way by putting efforts into never loosing focus on the overarching goals of the project, even when administrative short-term goals were calling for attention. This is not a challenge that was solved once and for all. As is the case with many challenges, facing them is a continuing process.

The continuity of productivity challenge (#4) was to keep an even pace of production that is, publishing articles while starting up large, time-consuming studies. The U-CARE portal was developed from scratch and all the large projects started up in parallel. The way we handled this challenge was by starting additional short-term projects to get a workable mix. This meant that the original RCTs were complemented with other types of studies. For example, brief intervention studies with student participants, qualitative interview-based studies, feasibility and pilot studies, registry-based studies, and studies describing the project development processes including trial protocols were accomplished (eg, [8-14]). It is important for a research organization's output to have nonsynchronized and timely overlapping projects.

Communication

Internal and external communication is a key factor for efficiency and goal attainment. Accordingly, much time and effort was devoted to establishing a clear intra- as well as an interorganizational communication strategy. The strategy described different information systems and communication channels for different purposes. The strategy serves as the blueprint for communication with the public, stakeholders, and between colleagues. The project's external and internal website, emails, shared server folders, the U-CARE portal itself, Web-based meeting systems, cloud technologies, and face-to-face meetings were important communication channels. Although this is an eHealth program, personal meetings ranging from regular working group meetings to international scientific conferences have been very important. The internal communication and documentation challenge (#5) was to implement an efficient strategy for internal communication aiming at making communication, information sharing, and documentation part of employees' daily routines and to ensure that all ongoing research and educational activities were documented properly. This documentation was the basis for the yearly reports of progress and productivity to funding sources. To meet these challenges, a combination of a general strategy building commitment to the project and a more specific strategy with frequent reminders and easy-to-use-templates to fill in was used.

XSL•FO
RenderX

The external communication challenge (#6) was to establish efficient external communication. One key task was to set up the U-CARE website. It allowed for effective and transparent external information sharing. The website served several of the defined communication purposes as it provided continuously updated and easily accessible information regarding main deliverables, ongoing studies, the organization structure of U-CARE, meeting protocols from the decision-making bodies, and policies, and so on. News was also distributed using Twitter with links to the U-CARE website for more information. In line with the university's third assignment, that is, to inform citizens, efforts were made to present the work to the public as well as politicians and health care officials. This was done by contacting the media, by participating in meetings arranged by nonprofit and patient organizations, and by inviting to a seminar at a popular societal and political yearly national event in Sweden (Almedalen) where many different stakeholders gathered.

Implementation

An important goal from the start for U-CARE was that any successful intervention should be implementable in regular care. Implementation of an intervention must depend on its efficacy and effectiveness, which is currently being evaluated in the RCTs. Thus, the implementation process cannot start until it is known whether an intervention is worth implementing, as the ideal in health care is to implement only evidence-based interventions. When it comes to eHealth, this creates a special difficulty as evidence takes a long time whereas the technological development is fast. At the time when a sufficient level of evidence is reached, the technology might be outdated. This may be a future challenge for U-CARE.

The idea of implementability has influenced the design of the U-CARE program from the start. For example, the recruitment of participants to the studies is done consecutively in clinical settings when possible, which is actually rare in the Internet therapy literature [15]. In addition, the material created for the portal was licensed to open access as per the most liberal Creative Commons license at the time, CC BY, where anyone is allowed to use and alter the material as long as it is attributed to the original author. It may seem uncontroversial but there are some potential negative side effects from a business perspective. The material sharing challenge (#7) is how to share the material for free without creating unfair competition or to limit the commercial interest. This challenge has not yet specifically been met in other ways than that it is acknowledged as a potential risk. A similar challenge applies to the use of open source code. In this case, we chose not to use open source, at least for now.

Newly developed eHealth technology has often had difficulties in reaching and being implemented in health care practices. This has been analyzed by, among others, van Gemert-Pijnen and colleagues [16]. They found that eHealth development often disregards the relationship among technology, human characteristics, and the socioeconomic environment. This led them to develop and suggest the CEHRES roadmap for development of eHealth technology, which emphasizes the interaction with the different stakeholders [16]. Although the CEHRES roadmap was not published when the U-CARE program started, several stakeholders were identified and approached early. The stakeholder involvement challenge (#8) was to identify and involve appropriate stakeholders. Researchers and clinicians such as psychologists and nurses have had a large input on the development of the program and the interventions. Other stakeholders who have not yet been involved are the future service providers, such as commercial or public clinics, who may deliver the interventions in regular care. For future implementation, it is also crucial to inform and engage a larger group of stakeholders such as politicians and health care officials early. Their engagement and knowledge is important for collaboration, implementation, and ultimately, value for the society.

The importance of interacting with the end-users in every part of the research process has become clear over the evolution of the program. Patient and public involvement (PPI) in research is recognized as an important strategy to ensure relevance and legitimacy of research activities and findings [17]. The involvement of patients in U-CARE aimed to ensure the clinical relevance and legitimacy of the interventions, the outcome measures, and the user interface of the U-CARE portal. In addition, after the interventions were prepared, patients have remained a valuable source of knowledge to improve the research projects. Involving patients in research activities was a novel approach to most researchers in U-CARE, and this work has led to an awareness of the complexity and costs of patient involvement as well as the benefits [18]. The public involvement challenge (#9) was to allocate time and the resources to do this work. It is not reasonable to expect contributions over several years from patient representatives without any compensation. Such resources were not planned for at an early stage and although PPI has been applied during the course of the program, greater efforts could have been made to take advantage of it. Ideally, a program of this size should have an assigned person with expertise in PPI to plan, educate, and allocate resources. Efforts were also made to change the researchers' view of patients from study objects to research partners. To enhance this work, it would be a good idea to educate both patient representatives and researchers on how to collaborate in research.

Legislation

A number of legislative challenges have arisen during the program, mainly concerning research, health care, and copyright.

The Swedish copyright legislation differs from other countries' when it comes to innovations conducted by a university employee. According to the unique Swedish professor's privilege, teachers and researchers have the rights to the intellectual property that originate from research, development, and innovation activities. This created the professor privilege challenge (#10), as it means that the university as an employer has neither rights nor direct incentives to manage these kinds of innovations [19]. It is up to the respective owner of the intellectual property, which in U-CARE's case includes many persons, to take care of how the end product should be handled (eg, advertised and maintained). With the intention to benefit as many as possible, an important ambition of the U-CARE-program was to share material with others. This has

XSL•FO
RenderX

required contracts with U-CARE staff, to license the material according to the aforementioned liberal CC BY-license. This has been done to remove some of the legislative barriers and facilitate future implementation.

At the onset of the program, the focus was on legal requirements regarding research ethics and personal integrity, which U-CARE researchers had previous experience with. The setting for the U-CARE program is academic, not clinical. When setting up the studies and routines, it became evident that no caregiver system was in place, which is required for psychological treatment in Sweden. After several discussions with the university management and National Board of Health and Welfare, it was decided to create a new caregiver organization, independent from the hospital or other clinical entities, at Uppsala University specifically for the research within U-CARE, a solution that hitherto has been very uncommon. As a result, the legal office at the university was obliged to support this new activity.

Legislation and practice around health care and research are not always in concordance. The competing legislation challenge (#11) has been to navigate these two legal systems. In research, for example, data should be as transparent as possible to other researchers (eg, Open data); whereas in health care, patients' integrity is of central importance. Another example is that in research the ideal is that the therapists are as blind as possible to the participants, for example, the baseline measures, whereas in medical practice it is essential to know details about the patient and to keep a medical record. This challenge is also a consequence of the double role that is common in clinical research, where the researchers are also therapists. Relating to this there is also an eHealth legislation lag. eHealth and related research has developed fast and the legislation is often not updated on this area, at least not in Sweden. This requires adaption to existing but not always appropriate Health and Medical Services legislation. For example, the distinction between research material and medical record information is not always clear. In two of the ongoing U-CARE studies, only participants who scored above a certain threshold level of anxiety and depression are offered to take part in an intervention after randomization. In this situation, it is not clear who are U-CARE's patients from a health care perspective and for whom a medical record is mandatory—For all the persons screened; for those randomized to treatment; or for all scoring above the threshold? On top of this, there are several new data safety issues that call for updated regulation, for example, cloud technologies.

Software Development

At the inception of U-CARE, the market was reviewed for software that could support the interventions and research for the planned studies in an appropriate way. No such software was found on the Swedish market. Hence, a strategic decision was made to collaborate with the information systems discipline in the process of developing the software program. Information systems researchers were engaged to participate both as researchers and as software developers. The decision meant that no external software consultants were engaged to help build the software. The information systems research focus was initially explorative and its research aim was somewhat unclear, but it became more focused with time.

The research in U-CARE largely depended on the U-CARE portal, which was adapted to both research and treatment needs. During the development phase, all parties underestimated the complexity of the software to be designed. Complexity was caused by a series of requirements, including (but not limited to) that the software:

- should be flexible so that it would suit the diverging needs of the different studies
- should be easy-to-use for people with diverging computer skills, provide a rich user experience, and feel attractive and modern
- should stimulate interaction between users, primarily among participants and between therapists and participants
- had to comply with current security standards and privacy regulations according to Swedish legislation
- could monitor study characteristics and provide updated study progress reports on recruitment and other parameters

The increasing complexity caused by a continuous flow of new requirements and applications from the clinical trial researchers led to considerable postponement of the launching of the U-CARE portal, which in turn affected the starting point of the planned studies.

When professionals with different backgrounds, for example, with regard to profession and academic discipline work together, there are communication challenges. In relation to software development, this is a well-known phenomenon. We had to make sure that the expressed and perceived expectations with respect to the system were in agreement. Here, we call this the mutual understanding of software requirements challenge (#12). It is hard to make demands within a field, for example, programming of software functions, in which you are not oriented. In the same way, it is difficult to respond to a demand that comes from an environment, for example, the clinical, that is unfamiliar. This challenge was met by adopting an iterative, agile, software development process and by weekly meetings between the developers or information systems researchers and the psychologists or clinical researchers [20]. The agile approach emphasizes continuous interaction between stakeholders and adaptation to changing customer requirements, in contrast to traditional plan-driven approaches to the, so-called, software development life cycle.

An additional software development documentation challenge (#13) was the need for continuous documentation of the development of the portal. New functionalities and complex configurations were added continuously as the portal developed, and with increasing time pressure, documentation was to some extent down prioritized. The lack of such documentation increases the risk to become dependent on certain individual's inherent knowledge of the code. Routines for documentation that were set up were not always adhered to. These routines were successively improved. It is of uttermost importance to have well-functioning and easy-to-use routines and management systems to maintain them.

XSL•FO
RenderX

Multidisciplinary Research

The development of multidisciplinary collaborations was included in U-CARE's fundamentals. For example, cost-benefit analyses were included in the original RCTs from the very start. This led to cooperation with researchers from health economics. As aforementioned, it was also decided early in the process that the development of the Internet portal was to be included as an integral part of the U-CARE program in contrast to using an existing portal or to work with IT-developers as consultants. Consequently, the technological development process became a research topic in itself mainly from an information systems perspective involving researchers from the informatics discipline. It was a general expectation that disciplines working together would be a success factor for dealing with complex research questions, for example, at cross-borders between the individual, the society, and a technological environment. It was also anticipated that down the road toward true interdisciplinarity, several obstacles would have to be dealt with [21]. At the outset, little was known about each other's cultures, structures, terminology, and science. Hence, it can be described as a discipline openness challenge (#14) to adjust to, be open to each other's qualities and expectations, and allocate resources in order to integrate our knowledge to reach new results. It proved to entail a greater intellectual challenge than was imagined. The initial tendency was to accumulate rather than integrate (eg, the interests of several research areas resulted in a larger number of questionnaires), with the risk of not satisfying any of the involved disciplines. Alternatively, making one discipline primary with priority to outline the research design would lead to others being secondary, marginalized, and unsatisfied.

The multidisciplinary ambition also resulted in higher demands on coordination and communication as the group became more diverse regarding professional and educational background as well as in research objectives. Collaboration was hampered by organizational differences with different systems in different departments, particularly for PhD students who had their project in one discipline or department and the main supervisor in another. Different cultures regarding PhD studies also made the integration more complicated. For an individual PhD student, the primary goal is to meet the requirements of the own discipline. Differences between disciplines regarding the various aspects of publication (type of publication, number of authors, author order, lack of interdisciplinary journals, etc) was another example of a cultural difference that was not fully anticipated beforehand.

Discussions regarding theoretical basis and research objectives have been ongoing and differences in scientific philosophy have been revealed along the way. The expectation that different scientific approaches are enriching has been confirmed. However, too large differences may also obstruct communication. The starting point for U-CARE was societal and clinical relevance rather than theoretical curiosity. The idea of theoretical framing and potential theoretical contributions emerged over time. However, given the large number of ongoing studies and the various backgrounds of the involved researchers, different subgroups relate their work to theory in different ways, without an agreed-upon shared theoretical framework. To bridge this gap was the challenge of a shared theoretical framework (#15). Several candidate theories to frame the studies were proposed including, but not limited to, learning theory, which is the base of cognitive behavioral therapy, and socio-materiality [22,23]. In addition, pragmatism has served as a philosophical starting point to guide and frame strategic research choices. We also applied more nontheoretical frameworks, for example, the British Medical Research Council's framework for development of complex interventions [24]. An implication of the lack of coordination in this respect may be a reduced long-term opportunity for theoretical contributions related to the design and application of technology for online psychosocial support. Moreover, the lack of an agreed-upon labeling of ongoing work may also lead to less-efficient communication of results to both academia and practice. Although letting each researcher frame results in a way that makes sense in his or her specific field may be effective and flexible, coordination is needed to make a true multidisciplinary contribution.

U-CARE's Accomplishments

So, was U-CARE a successful program? By the end of 2014, the following had been accomplished regarding the goals set in 2009 (see U-CARE's project goals above).

Regarding establishing a research platform, the most significant output since the inception of the U-CARE program has been building an Internet portal, that is, the U-CARE portal, a generic and flexible Internet portal that supports collection of data and provision of psychosocial care and psychological treatment. Web-based psychosocial and psychological self-help programs have been developed for pediatric oncology, adult oncology, and cardiology [8,9,11,14]. Within the U-CARE program an infrastructure, that is, logistics, structures, and policies, for delivering care and psychological treatment online has been created. A number of research groups have been associated to the U-CARE program in order to use the U-CARE portal and the self-help programs developed within U-CARE. The associated groups contribute with Web-based open access self-help material for groups such as pregnant women with fear of giving birth; women who experience post-traumatic stress after a difficult delivery or abortion; persons with risk for recurrence of depression; children who experience pain due to mucositis, and so on. By the end of 2014, 6 studies had started in the portal, all targeting Swedish participants. In the first goal, (1) international competitiveness and innovation were stressed on, but these aspects are difficult to evaluate factually at the moment. (2) A research-based multidisciplinary education within the field of psychosocial health care and a national graduate school in innovative psychosocial health care research has been organized. Eight PhD students started the graduate school in 2011 or 2012. Three multidisciplinary courses on eHealth-related topics have run between 2012 and 2014. (3) Regarding career opportunities, the following figures indicate a positive result. In 2009, the research group where U-CARE is hosted consisted of 6 persons. By the end of 2010, the group already consisted of 25 persons, and the size of the research group has been relatively stable throughout 2014. In addition, approximately 25 persons with positions at other departments at Uppsala University and 10 persons at other Swedish universities are involved in the U-CARE program. In total, 13

XSL•FO
RenderX

PhD students are involved in the U-CARE program. Four post-doctoral positions have become permanent researcher positions. (4) When it comes to attracting major external funding from major Swedish research foundations, the U-CARE program has been successful. In total, 19.354.900 SEK (approximately €2.134.000) have been received in addition to the original funding for research activities between 2011 and 2014. At the moment, 29.930.000 SEK (approximately €3.300.000) has been granted for 2015-2018. However, no funding has been applied for and thus not been received from the EU framework program. (5) Originally, there was a plan for a center of excellence organized independently of the existing university departments. However, these plans were changed already at the inception because of organizational reasons. U-CARE has since inception throughout 2016 been hosted at the Department of Public Health and Caring Sciences, group: Clinical Psychology in Healthcare, Uppsala University.

In May 2015, the Swedish Research Council published a comprehensive evaluation of all the 43 programs in the Strategic Research Area-Initiative. All programs were evaluated by an expert panel regarding, among other things, research output, collaborations, and management. The conclusion was that U-CARE has developed satisfactorily in many aspects, especially when it comes to management, association with other national research groups, and for effectively starting up a new organization within a new field of research. The main challenges for the future pointed out by the experts were research output, that is, scientific publications, international collaborations, impact on business, and contribution to undergraduate education [25].

Conclusions

Establishing a publicly funded multidisciplinary eHealth research environment is a challenging endeavor. Although several challenges were met during the development of the U-CARE program, several accomplishments were also made. Some of which could not have been made if it weren't for this new long-term research environment. By sharing our experience, we hope to help other research groups embarking on a similar journey to be prepared for some of the challenges they are likely to meet on their way. We do anticipate future challenges within these themes that are not described in this paper. This involves implementation in care if the interventions prove to be efficient, continuous adjustments to new technologies, and legislation. By reflecting on our previous challenges, we are better armed to face the challenges of the future.

For some challenges, we provide a solution that worked for us. For other challenges, we have suggestions that we have not fully implemented. We recommend our readers to reflect on how to tackle the challenges and anticipate that being prepared will help to tackle these challenges in a more efficient way.

Acknowledgments

We thank our colleagues working in, or close to, the U-CARE program. This work was supported by The Swedish Research Council (2009-1093).

Conflicts of Interest

None declared.

References

1. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-mental health care in Australia: a systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]
2. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565. [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
3. 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]
4. 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 [FREE Full text] [doi: 10.1080/15228830802094429]
5. Ahern DK. Challenges and opportunities of eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S75-S82. [doi: 10.1016/j.amepre.2007.01.016] [Medline: 17466822]
6. Cuijpers P, Donker T, van SA, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
7. Sjöström J, von EL, Grönqvist H. The origin and impact of ideals in ehealth research: experiences from the U-CARE research environment. JMIR Res Protoc 2014 May 23;3(2):e28 [FREE Full text] [doi: 10.2196/resprot.3202] [Medline: 24860071]

XSL•FO
RenderX

8. Norlund F, Olsson EM, Burell G, Wallin E, Held C. Treatment of depression and anxiety with internet-based cognitive behavior therapy in patients with a recent myocardial infarction (U-CARE Heart): study protocol for a randomized controlled trial. Trials 2015 Apr 11;16:154 [FREE Full text] [doi: 10.1186/s13063-015-0689-y] [Medline: 25873137]
9. Mattsson S, Alfonsson S, Carlsson M, Nygren P, Olsson E, Johansson B. U-CARE: Internet-based stepped care with interactive support and cognitive behavioral therapy for reduction of anxiety and depressive symptoms in cancer--a clinical trial protocol. BMC Cancer 2013;13:414 [FREE Full text] [doi: 10.1186/1471-2407-13-414] [Medline: 24024826]
10. Wallert J, Madison G, Held C, Olsson E. Cognitive ability, lifestyle risk factors, and two-year survival in first myocardial infarction men: a Swedish National Registry study. Int J Cardiol 2016 Dec 28;231:13-17 [FREE Full text] [doi: 10.1016/j.ijcard.2016.12.144] [Medline: 28062133]
11. Ander M, Wikman A, Ljótsson B, Grönqvist H, Ljungman G, Woodford J, et al. Guided internet-administered self-help to reduce symptoms of anxiety and depression among adolescents and young adults diagnosed with cancer during adolescence (U-CARE: YoungCan): a study protocol for a feasibility trial. BMJ Open 2017 Jan 27;7(1):e013906 [FREE Full text] [doi: 10.1136/bmjopen-2016-013906] [Medline: 28132011]
12. Alfonsson S, Olsson E, Hursti T. Motivation and treatment credibility predicts dropout, treatment adherence, and clinical outcomes in an internet-based cognitive behavioral relaxation program: a randomized controlled trial. J Med Internet Res 2016;18(3):e52 [FREE Full text] [doi: 10.2196/jmir.5352] [Medline: 26957354]
13. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
14. Hauffman A, Alfonsson S, Mattsson S, Forslund M, Bill-Axelson A, Nygren P, et al. The development of a nurse-led internet-based learning and self-care program for cancer patients with symptoms of anxiety and depression-A Part of U-CARE. Cancer Nurs 2016 May 24:a Epub ahead of print [FREE Full text] [doi: 10.1097/NCC.0000000000000402] [Medline: 27223884]
15. Internet-based psychological treatment for anxiety and mood disorders. Stockholm: Swedish Council on Health Technology Assessment; 2013. URL: http://www.sbu.se/en/publications/sbu-assesses/internet-based-psychological-treatments-for-anxiety-and-mood-disorders/ [accessed 2017-04-27] [WebCite Cache ID 6q1nErLVh]
16. van Gemert-Pijnen JE, Nijland N, van LM, 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]
17. Staley K. Exploring Impact: Public involvement in NHS, public health and social care research. Eastleigh: INVOLVE; 2009.
18. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, Dabrh Abd Moain Abu, et al. Patient engagement in research: a systematic review. BMC Health Serv Res 2014;14:89 [FREE Full text] [doi: 10.1186/1472-6963-14-89] [Medline: 24568690]
19. Farnstrand Damsgaard E, Thursby M. University entrepreneurship and professor privilege. Industrial and Corporate Change 2013 Jan 24;22(1):183-218 [FREE Full text] [doi: 10.1093/icc/dts047]
20. Dybå T, Dingsøyr T. Empirical studies of agile software development: a systematic review. Inform. Software Tech 2008 Aug;50(9-10):833-859 [FREE Full text] [doi: 10.1016/j.infsof.2008.01.006]
21. Roy E, Morzillo A, Seijo F, Reddy S, Rhemtulla J, Milder J. The elusive pursuit of interdisciplinarity at the human--environment interface. Bioscience 2013 Sep 01;63(9):745-753 [FREE Full text] [doi: 10.1525/bio.2013.63.9.10]
22. Dobson K. Handbook of Cognitive-Behavioral Therapies, Third Edition. New York: The Guilford Press; 2010.
23. Leonardi P. Theoretical foundations for the study of sociomateriality. Information and Organization 2013 Apr;23(2):59-76 [FREE Full text] [doi: 10.1016/j.infoandorg.2013.02.002]
24. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]
25. Swedish Research Council. Evaluation of the Strategic Research Area Initiative 2010-2014. Stockholm, Sweden: Swedish Research Council; 2015.

Abbreviations

PPI: patient and public involvement
RCT: randomized controlled trial
SAB: Scientific Advisory Board
U-CARE: Uppsala University Psychosocial Care Program

XSL•FO
RenderX

Edited by G Eysenbach; submitted 19.01.17; peer-reviewed by H Durrani, K Anderson, JA Ramos-Quiroga; comments to author 09.02.17; revised version received 22.03.17; accepted 08.04.17; published 23.05.17.

Please cite as:
Grönqvist H, Olsson EMG, Johansson B, Held C, Sjöström J, Lindahl Norberg A, Hovén E, Sanderman R, van Achterberg T, von Essen L
Fifteen Challenges in Establishing a Multidisciplinary Research Program on eHealth Research in a University Setting: A Case Study
J Med Internet Res 2017;19(5):e173
URL: http://www.jmir.org/2017/5/e173/
doi:10.2196/jmir.7310
PMID:28536090

©Helena Grönqvist, Erik Martin Gustaf Olsson, Birgitta Johansson, Claes Held, Jonas Sjöström, Annika Lindahl Norberg, Emma Hovén, Robbert Sanderman, Theo van Achterberg, Louise von Essen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Self-Management Support Using a Digital Health System Compared With Usual Care for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial

Andrew Farmer[1], DM (Oxon); Veronika Williams[1], PhD; Carmelo Velardo[2], PhD; Syed Ahmar Shah[2], DPhil (Oxon); Ly-Mee Yu[1], DPhil (Oxon); Heather Rutter[1], DipHE; Louise Jones[1], DPhil (Oxon); Nicola Williams[1], MSc; Carl Heneghan[1], DPhil (Oxon); Jonathan Price[3], DPhil (Oxon); Maxine Hardinge[4], FRCP; Lionel Tarassenko[2], DPhil (Oxon)

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom

[3]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[4]Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

Corresponding Author:
Andrew Farmer, DM (Oxon)
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865 289280
Fax: 44 1865 289200
Email: andrew.farmer@phc.ox.ac.uk

Abstract

Background: We conducted a randomized controlled trial of a digital health system supporting clinical care through monitoring and self-management support in community-based patients with moderate to very severe chronic obstructive pulmonary disease (COPD).

Objective: The aim of this study was to determine the efficacy of a fully automated Internet-linked, tablet computer-based system of monitoring and self-management support (EDGE, sElf-management anD support proGrammE) in improving quality of life and clinical outcomes.

Methods: We compared daily use of EDGE with usual care for 12 months. The primary outcome was COPD-specific health status measured with the St George's Respiratory Questionnaire for COPD (SGRQ-C).

Results: A total of 166 patients were randomized (110 EDGE, 56 usual care). All patients were included in an intention to treat analysis. The estimated difference in SGRQ-C at 12 months (EDGE−usual care) was −1.7 with a 95% CI of −6.6 to 3.2 (P=.49). The relative risk of hospital admission for EDGE was 0.83 (0.56-1.24, P=.37) compared with usual care. Generic health status (EQ-5D, EuroQol 5-Dimension Questionnaire) between the groups differed significantly with better health status for the EDGE group (0.076, 95% CI 0.008-0.14, P=.03). The median number of visits to general practitioners for EDGE versus usual care were 4 versus 5.5 (P=.06) and to practice nurses were 1.5 versus 2.5 (P=.03), respectively.

Conclusions: The EDGE clinical trial does not provide evidence for an effect on COPD-specific health status in comparison with usual care, despite uptake of the intervention. However, there appears to be an overall benefit in generic health status; and the effect sizes for improved depression score, reductions in hospital admissions, and general practice visits warrants further evaluation and could make an important contribution to supporting people with COPD.

Trial registration: International Standard Randomized Controlled Trial Number (ISRCTN): 40367841; http://www.isrctn.com/ISRCTN40367841 (Archived by WebCite at http://www.webcitation.org/6pmfIJ9KK)

(J Med Internet Res 2017;19(5):e144) doi:10.2196/jmir.7116

XSL•FO
RenderX

KEYWORDS

pulmonary disease, chronic obstructive; telehealth; self-care; randomized controlled trial

Introduction

Background

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality [1], with global costs of US $141 billion. In the United Kingdom, the total annual estimated cost of COPD to the National Health Service (NHS) is over £800 million, over half attributable to hospital-based care [2]. The impact of COPD on the health-related quality of life of patients is well-established [3,4].

Correct use of long-acting beta-agonists alone or in combination with anti-inflammatory inhaled steroids can reduce the overall rate of exacerbations by 20% [5]. Early recognition of and intervention during an exacerbation can also reduce admission risk [6]. Training and support for patients in the self-management of their condition, for example, through outpatient pulmonary rehabilitation [7], improves quality of life and can reduce unplanned hospital admissions [8-10]. There is an evidence base for use of psychological therapies such as cognitive behavioral therapy [11], although a wider range of interventions designed to promote self-management do not appear to work consistently, possibly because of nonadherence to the self-management program or to action plans put in place [12,13], or because not all the components required to improve self-management are included.

There is a need to identify different forms of self-management that can improve outcomes and to develop and optimize ways of delivering available interventions to maximize effectiveness and safety. Delivery of interventions at a widescale and continuing the intervention over time are likely to be important factors addressing feasibility and maintaining effectiveness. Use of converging computer and communication technologies in the form of digital health interventions offers a means of helping patients monitor their condition, providing support in interpreting data for self-management, and supplying a means of delivering individually tailored education and treatment plans.

Digital Health Intervention

Digital health interventions in COPD are increasingly employed, and particularly health apps have been found to show potential in improving symptoms management through self-management support [14]. Current digital health interventions and apps include a range of components, and mainly differ in the combination of tools used (patient education materials, exercise support, self-care plans, remote monitoring of symptoms, and clinical parameters) and input from clinicians (interpretation of patient data and feedback to patient through phone calls, text chats, or video calls) [15,16].

Understanding the heterogeneous nature of COPD events with a variable time course of symptom onset and recovery is particularly important to underpin interventions that may work for individual patients [17].

Systematic reviews of digital health interventions till date in COPD provide evidence to support continuing research [18], but recent large-scale evaluations have not shown convincing evidence of their effectiveness [19,20]. Three large trials of telehealth solutions for COPD have identified little or no benefit [19,21,22]. Limitations of current systems have included low compliance rates with the technology resulting from having to engage with a dedicated telehealth box; lack of flexibility in the technical specification; artifacts in data collection leading to high rates of false alerts; and limited personalization and support for self-management in the telehealth solution. Therefore, solutions need to be straightforward and easy-to-use by patients, be straightforward to implement with low-cost widely available technology, and utilize individualized predictive algorithms to address the variability of the condition [23-25].

Evaluation of digital health interventions require multiple perspectives to be considered within the evaluation [26]. We have recently carried out a cohort study in which we have shown that a tablet computer-based system for supporting patients with COPD is acceptable to them and feasible to use [23,27].

We therefore set out to determine the efficacy of an Internet-linked, tablet computer-based system of monitoring, and self-management support (EDGE, sElf-management anD support proGrammE) in improving quality of life and clinical outcomes when used by patients with moderate to very severe COPD.

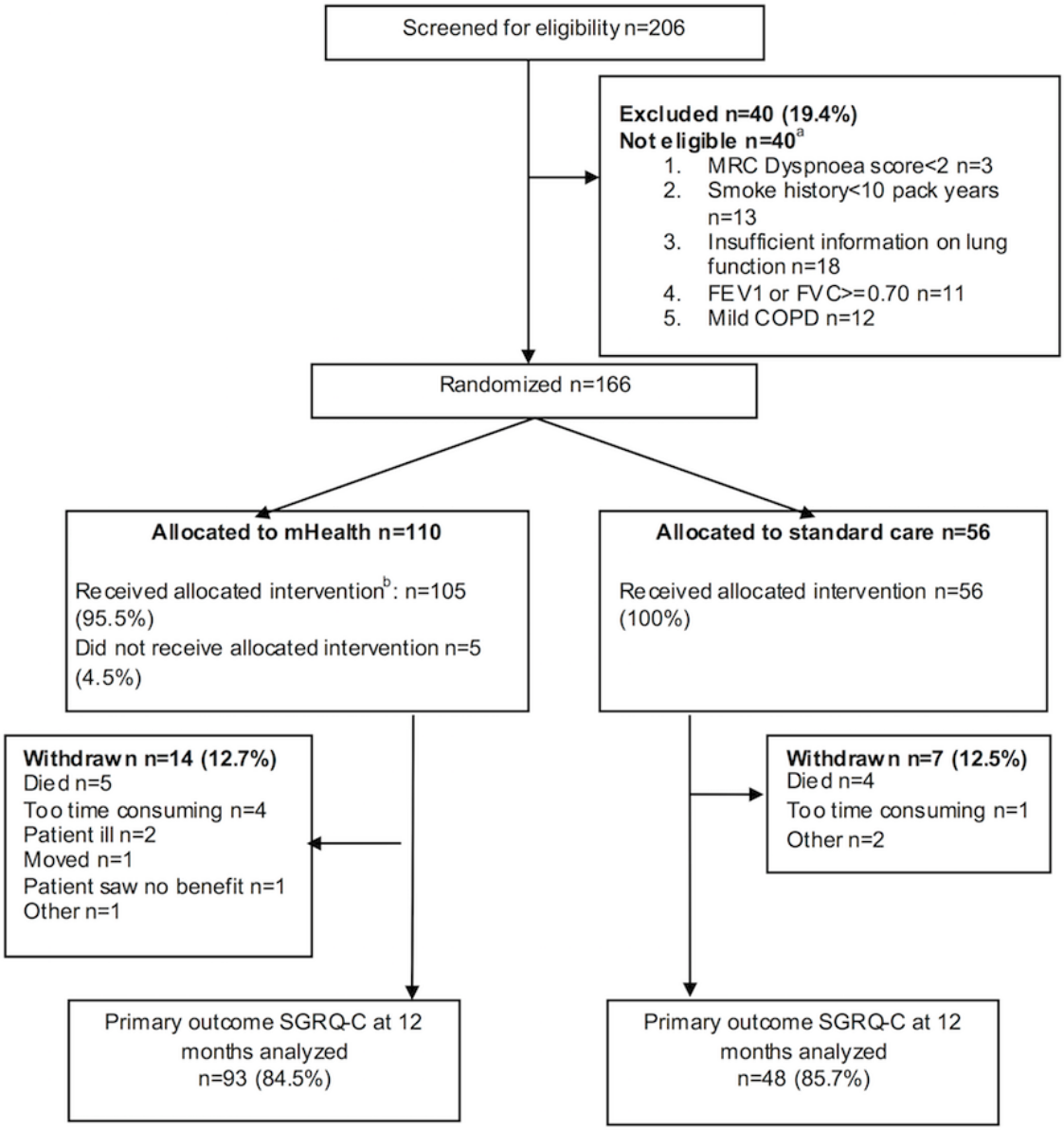
Methods

Trial Design

EDGE for COPD is a multicenter, randomized controlled trial of 12-month duration [28]. Patients were individually randomized to receive either a system of care (the EDGE intervention) delivered via a digital health, Internet-linked platform implemented on a low-cost tablet computer (the EDGE platform) providing monitoring and self-management support or standardized usual care in a 2:1 allocation ratio (Figure 1). A 2:1 allocation ratio was chosen to maximize the information available about the use of the system across the population of individuals with COPD [29]. There were no changes to the design or methods of the study after recruitment commenced. An embedded qualitative study was used to carry out a process evaluation and to explore the experience of using the system.

XSL•FO
RenderX

Figure 1. Participant flow legend. a: not mutually exclusive. b: defined as using the intervention for at least 30 days, during which time it was used for at least 3 out of 7 days per week. MRC: Medical Research Council; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; COPD: chronic obstructive pulmonary disease, SGRQ-C: St George's Respiratory Questionnaire for COPD.



Participants

Eligibility Criteria for Participants

Eligible patients were aged ≥40 years with a confirmed diagnosis of COPD defined as a forced expiratory volume in 1 s (FEV1), post-bronchodilation of <70% [2], and a predicted ratio of FEV1 to forced vital capacity of <0.70. Eligible patients had a smoking-pack history >10 pack-years and a Medical Research Council dyspnea score of ≥2. Further trial eligibility criteria are reported in the trial protocol [28].

Setting

Patients were recruited from a variety of settings encompassing primary and secondary care as well as community services. Patients attending respiratory hospital outpatient clinics and pulmonary rehabilitation courses in the adjacent counties of Oxfordshire and Berkshire, UK, were invited to participate. In addition, eligible patients were identified from primary care clinics and from those recently (within the preceding 2 weeks) discharged from hospital following a COPD-related admission.

Trial Interventions

Intervention Development and Specification

On the basis of open-source app software, the EDGE platform was designed to be integrated in clinical care, by a team of clinicians and engineers working with patients. The platform was refined in a 6-month cohort study of a group of patients with COPD, who were selected using eligibility criteria matching those of the trial [27].

The EDGE intervention was designed to include tools to help patients identify exacerbations and to monitor their condition; to help support good compliance with inhaled medication; and to support psychological well-being. It incorporates a daily

XSL•FO
RenderX

symptom diary consisting of standard questions about symptoms based on previous trial protocols [30,31]. Questions include general well-being, cough, breathlessness, sputum (quantity produced and color), and use of medications. A 30-s period of data acquisition using a Bluetooth-enabled pulse oximeter with finger probe (Nonin, PureSAT, 9560BT) manufactured by Nonin Medical Inc of Plymouth, Minnesota, USA, allows daily collection of heart rate and oxygen saturation data. Mood screening questionnaires [32-34] were presented each month for completion; further details are given in Multimedia Appendix 1.

The EDGE platform also includes a number of software modules, including videos tailored to the patient's entries in the symptom diary or answers to the mood-screening questionnaires. These videos provide additional self-management support, and are listed in Multimedia Appendix 1. These include inhaler techniques, pulmonary rehabilitation exercises, and self-management techniques for breathlessness. All of the multimedia components were developed by the local clinical respiratory team and are based on current pulmonary rehabilitation interventions, for which there is good evidence of effectiveness in improving quality of life for patients with COPD [35]. As soon as the patient finishes using the app, data is securely transmitted to a server hosted behind NHS firewalls.

Use of the EDGE Platform

Participants allocated to receive the EDGE platform-based intervention were provided with an Android tablet computer (Samsung Galaxy Tab) running the app software and a Bluetooth-enabled oximeter probe.

Participants were briefly instructed on the use of the EDGE platform by the research nurse and given a brief information booklet detailing its use. Patients were informed that the EDGE platform was not a replacement for their usual clinical care, and that in the event of deterioration in their health they should contact their general practitioner or community respiratory nurse as usual. The intention of this approach was to establish the safe use of new technology, whereas not intervening in a way that might expose individuals to unintended harms [36].

In an initial 6-week period of use, EDGE users completed the symptom diary and recorded their oxygen saturation and heart rate with the pulse oximeter on a daily basis. Following this initial run-in period, the distributions of values for the oxygen saturation, heart rate, and symptom scores were calculated for the run-in period for each participant. The 97th centile was computed for the distribution of heart rate and symptom score, and the 3rd centile for the distribution of oxygen saturation; these were then used as the threshold for the participant safety alert for that parameter. The option to modify thresholds following a hospital admission was available. Participants continued to input their symptom data and clinical recordings daily throughout the duration of the trial.

One of three respiratory clinicians (nurse, physiotherapist, or doctor) reviewed a summary of the oxygen saturation, heart rate, and symptom diary module data twice weekly to ensure that data transmission was taking place and to deal with safety alerts. Data were assigned priority according to the number of "alerts" in the most recent two weeks since the last review. In this context, alerts were generated when the vital sign value (pulse rate or oxygen saturation) went above (or below) the safety threshold or the overall symptom score went above the safety threshold.

If data were not received or there were safety alerts, the participant record was accessed for review. If, on reviewing the data, there was judged to be a clinically important change in the data, then the patient was contacted either via message or telephone. A clinically important parameter was defined as either heart rate or symptom score moving above the 97th centile or oxygen saturation falling below the 3rd centile, as defined by baseline observation and persisting for at least 2 days. If depression or anxiety scores equaled or exceeded a threshold of 10, then the patient's general practitioner was informed by letter. The intervention is summarized in Multimedia Appendix 2.

The Standardized Usual Care Intervention

Participants allocated to receive standardized usual care were provided with all the information given to those allocated to use the EDGE system, but without the use of a tablet computer or the facility for daily monitoring of symptoms and physiological variables. Participants were provided with leaflets based on those currently produced by the Oxfordshire Community Respiratory service. Further details are given in Multimedia Appendix 3.

Primary Outcome Measure

The primary outcome was the change in St George's Respiratory Questionnaire for COPD (SGRQ-C) [37], which was used to assess COPD-specific health status from baseline to 1 year in patients with moderate to very severe COPD.

Secondary Outcome Measures

The following secondary endpoints were used to evaluate the impact of the intervention in comparison with usual care: (1) impact on hospital admissions (number of admissions and days in hospital) and deaths; (2) the number of recorded exacerbations defined as episodes in which antibiotics or oral steroids were prescribed or in which the patients were seen in the accident and emergency department or admitted to hospital in the presence of an acute change in respiratory symptoms defined as the presence of at least two symptoms, one of which should be major (major symptoms: change in sputum, more breathless, chest tight; minor symptom: unwell, tired, temperature, a cold) or a report of a patient taking more salbutamol, either blue inhaler or by nebulizer, for at least 48 h (Multimedia Appendix 1); (3) time to first exacerbation; (4) beliefs about respiratory medicine use measured with the Beliefs about Medicines Questionnaire [38]; (5) self-reported medication use measured with the Medication Adherence Report Schedule [38]; (6) self-reported smoking cessation; (7) mood measured with the Standard Checklist 20-item Questionnaire (SCL-20) for depression [39] and the Standard Checklist 10-item Anxiety Measure (SCL-10A) [40]; and (8) a comprehensive measure of health status using the EuroQol 5-Dimension Questionnaire (EQ-5D) [41]. The SCL20 and SCL-10A are derived from a 90-question standard measure and have been used extensively

XSL•FO
RenderX

to measure mood and anxiety as outcome measures in studies with people who have long-term conditions. The EQ-5D includes 5 questions asking about mobility, self-care, usual activities, pain and discomfort, and anxiety or depression. Each question has a 3-level response, and the responses are used to estimate a preference weight for that health status, presented as a single index value.

Details of number and duration of hospital admissions were measured by self-report and confirmed where possible by review of hospital discharge letters and central hospital admissions data. Records of deaths were obtained from general practices and further details were obtained, where necessary, from hospital records. Details of exacerbations of COPD were recorded on a record form held by all participants [28].

Sample Size

The sample size calculations were based on the number of patients required to demonstrate a mean difference of 6.6 on the St George's Respiratory Questionnaire between the two allocated groups, at 12-months from randomization (equivalent to 7.3 on SGRQ-C) [37]. We estimated the SD at 12.7 based on a study using the SGRQ-C [42]. For a power of 90% and significance level of .05 (2-sided), with 2:1 allocation between intervention and usual care and allowing for 10% loss to follow-up, we required 165 patients.

We also had 98% power to identify the difference in admissions to hospital at 3 months based on effect sizes of previous intensive interventions with this group of patients [7], and 52% power to detect the difference in admissions at 12 months based on a systematic review of interventions in COPD [43]. In both cases, a 5% loss to follow-up was assumed.

Randomization

Participants were randomized with an allocation ratio of 2:1 intervention to usual care using Sortition V.1.2 [28]. The research nurse carried out randomization by accessing Sortition using a Web-browser on a tablet computer at the assessment visit only after completion of consent procedures and baseline measurements, including completion of the SGRQ-C.

Trial Procedures

Recruitment

Potentially eligible patients were sent an invitation to participate in the trial. The invitation included a patient information booklet, a reply slip, and prepaid envelope. Patients who were interested in participating were asked to return their reply slips by post to the research team. The research nurse then contacted the patient by telephone to arrange an initial assessment visit. At this visit, eligibility was confirmed, written informed consent was obtained, and baseline data were collected for those consenting to participate.

All participants were assessed at baseline by a health care professional and had finished self-completed measures before randomization and intervention allocation. The use of medication by participants was recorded at the baseline and follow-up assessment visits. Information collected included type, dose, and frequency of COPD medication (tablets and inhalers) as well as a list of other medication taken. A detailed smoking history was taken at the baseline assessment visit; self-reported smoking status was recorded at subsequent assessments. All participants had either a written or tablet computer held action plan (uploaded as Multimedia Appendix 3) for use with changing symptoms, and were encouraged to ensure that, in line with standard practice, they had reserve supplies of antibiotics and steroids.

Patient Follow-Up and Retention

Patients remained in the trial for 12 months with assessments at a baseline visit, 3, 6, and 12 months. The primary outcome measure was collected at baseline, and 6 and 12 months after randomization. Secondary outcome measures were collected at baseline, 3, 6, and 12 months.

Postal reminders were sent before all follow-up assessment visit dates. The 3-month assessment was a telephone contact with patients. For patients allocated to standardized usual care, a reminder was posted before the assessment date. The 6 and 12 months visits were carried out either at home or at clinic.

Self-completed outcome measures were completed without guidance by the research team and before any further assessment or discussion of clinical care. Research and clinical teams were trained in the potential for measures to be biased by their interactions with participants. A record of all contacts with trial participants was kept to examine potential for interactions with patients not specified in the trial protocol.

All patients had the right to withdraw from the trial at any point, without providing a reason. Those patients who withdrew from receiving the intervention were asked if they would be willing to provide follow-up information within the trial at the 6- and 12-month assessment points. If patients declined, no further information was collected.

Statistical Methods

The principal comparisons were performed on an intention-to-treat basis. The trial results are presented as comparative summary statistics (difference in response rates or means) with 95% CI. A linear mixed-effects model was used to analyze SGRQ-C including randomized group (intervention or control), time point (6 months and 12 months), baseline SGRQ-C score, age (analyzed continuously), gender, current smoking status (yes or no), severity of COPD (according to NICE clinical guidelines or GOLD classification, moderate or severe or very severe), and site as fixed effects and a patient-specific random intercept. Treatment-time interaction was included in the model to assess the treatment effect at 12 months. Higher scores reflect better quality of life compared with lower scores. Binary outcomes were analyzed using log-binomial regression, adjusting for covariates as described above. The threshold for statistical significance was less than P=.05 with no adjustments made for multiple testing among secondary outcomes.

The intervention effect was assessed by analysis of subgroups defined by severity of COPD, smoking status, hospital admission in the previous year, attending a pulmonary rehabilitation course in the previous year, and the presence or absence of live-in

XSL•FO
RenderX

support. A full detailed statistical analysis plan was prepared before the final analysis by a trial statistician.

The EDGE COPD trial was carried out in conformance with the principles of the current version of the Declaration of Helsinki and the other regulations in force. The trial is registered at ISRCTN 40367841.

Results

Trial Progress and Baseline Characteristics of Participants

The CONSORT flow chart is presented in Figure 1. The reasons for patient withdrawal are also detailed in Figure 1. The first patient was randomized on June 26, 2013 and follow-up was completed on July 27, 2015.

Baseline characteristics of trial participants are summarized in Table 1 and in Multimedia Appendix 1. There were no relevant differences in characteristics between those assigned to EDGE and to usual care.

A total of 14 patients, who were allocated to receive the EDGE system, withdrew from the trial of whom 5 died: 7 withdrew from the usual care group with 4 deaths (Figure 1).

Use of the System

Out of the 110 patients who were part of the intervention arm of the study, 100 patients were in the study for at least 180 days. Compliance with use of the system was a mean (SD) of 5.9 (1.1) days per week of use across all patients (range 1.4-7.0). Among the 100 patients, only 2 patients had a compliance of less than 3 times per week. The video clips offered to intervention group participants were used with varying frequency. The videos relating to mood and breathing exercises were reviewed more frequently than others (Multimedia Appendix 1). In total, 90% (99/110) of participants viewed at least one video. The mean (SD) number of videos viewed by participants was 5 (3.5) and the mean number of times a video was viewed was 22.5 (19.9). The mean (SD) number of accesses made by the nurse on the website to the data for each participant in the intervention group was 33.4 (SD 15.4, range 11-79), that is, an average of 2.78 accesses per patient per month.

Primary Outcome

Quality of life as measured with the SGRQ-C improved in patients allocated to both the EDGE system and to usual care from baseline to 6 months, and again to 12 months. The estimated difference in SGRQ-C at 12 months (EDGE system−usual care) was −1.7 with a 95% CI of −6.6 to 3.2 (P=.49). SGRQ-C scores and changes over time are summarized in Table 2. Data were available on 84.5% (93/110) of patients in the EDGE system group, and from 85.7% (48/56) patients in the usual care group.

XSL•FO
RenderX

Table 1. Baseline characteristics of participants.

Characteristics	EDGE[a] intervention n=110	Standard Care n=56
Clinical data		
Male, n (%)	68 (61.8)	34 (60.7)
Age, mean (SD[b])	69.8 (9.1)	69.8 (10.6)
BMI[c], mean (SD)	28.6 (7.1)	29.1 (7.8)
FEV1[d], mean (SD)	47.4 (15.6)	50.1 (16.9)
FEV1 or FVC[e], mean (SD)	47.6 (11.3)	49.8 (11.5)
Number of COPD[f]medications, median (IQR[g])	5 (3-6)	5 (4-6)
Number of other medications, median (IQR)	4 (2-7)	5 (2.5-8)
Smoking history		
Current, n (%)	23 (20.9)	13 (23.2)
Ex-smoker (<2 years), n (%)	17 (15.5)	8 (14.3)
Ex-smoker (≥2 years), n (%)	70 (63.6)	35 (62.5)
COPD severity		
Moderate, n (%)	41 (37.3)	23 (41.1)
Severe or very severe, n (%)	69 (62.7)	33 (58.9)
MRC[h] dyspnoea score, N (%)		
2	17 (15.5)	10 (17.9)
3	74 (67.3)	39 (69.6)
4	19 (17.3)	7 (12.5)
Comorbid conditions including high blood pressure, osteoporosis, high cholesterol, diabetes, heart disease, and depression), N (%)	89 (80.9)	47 (83.9)
Patient reported outcome measures		
SGRQ-C (St George's Respiratory questionnaire for COPD patients), mean (SD)	56.4 (19.7)	55.5 (16.2)
SCL-10A[i], median (IQR)	0.3 (0.1-0.9)	0.3 (0-0.5)
SCL-20[j], median (IQR)	0.53 (0.3-1.15)	0.68 (0.3-1.1)
BMQ (Beliefs about Medicines Questionnaire), mean (SD)	24.6 (4.8)	25.3 (5.7)
MARS (Medicines Adherence Report Scale), mean (SD)	23.4 (2.3)	22.5 (3.8)
EQ-5D[k]Index, mean (SD)	0.62 (0.24)	0.63 (0.24)
Deprivation score[l], mean (SD)	22,440 (7951.9)	22,777 (7261.5)

[a]EDGE: sElf-management anD support proGrammE.

[b]SD: standard deviation.

[c]BMI: body mass index.

[d]FEV1: forced expiratory volume in 1 s.

[e]FVC: forced vital capacity.

[f]COPD: chronic obstructive pulmonary disease.

[g]IQR: interquartile range (25th, 75th percentiles).

[h]MRC: Medical Research Council.

[i]SCL-10A: Standard Checklist 10-item Anxiety Measure.

[j]SCL-20: Standard Checklist 20-item Questionnaire.

[k]EQ-5D: EuroQol 5-Dimension Questionnaire.

[l]On the basis of postcode with deprivation rankings accessed from the UK Office of National Statistics.

XSL•FO
RenderX

Table 2. Primary outcome—St George's Respiratory Questionnaire for chronic obstructive pulmonary disease.

Primary Outcome	EDGE[a]intervention N=110	Standard care N=56
SGRQ-C[b]		
Baseline, mean (SD[c])	56.4 (19.7)	55.5 (16.2)
6 months, mean (SD)	55.7 (20.2) N=98	54.3 (21.8) N=51
12 months, mean (SD)	56.9 (19.5) N=93	56.8 (20.9) N=48
Difference between groups[d]		
6 months, mean (95% CI; *P* value)	0.99 (−3.81 to 5.78; .69)	
12 months, mean (95% CI; *P* value)	−1.74 (−6.65 to 3.16; .49)	

[a]EDGE: sElf-management anD support proGrammE.

[b]SGRQ-C: St George's Respiratory Questionnaire for COPD.

[c]SD: standard deviation.

[d]From mixed effects model including randomized group (intervention or control), time point (6 months and 12 months), baseline SGRQ-C score, age (analyzed continuously), gender, current smoking status (yes or no), severity of COPD (according to NICE clinical guidelines or GOLD classification, moderate or severe or very severe), and site as fixed effects and a patient-specific random intercept. Higher scores reflect better quality of life compared with lower scores.

Secondary Outcomes

Secondary outcomes are summarized in Table 3. Deaths did not differ between groups. Numbers of exacerbations did not differ overall between groups. The relative risk of hospital admission for EDGE was 0.83 (0.56-1.24, *P*=.37) compared with usual care. There was a significant difference in overall health status measured with the 5-Level EuroQol 5-Dimension Questionnaire (EQ-5D-5L) between groups 0.076 (0.009-0.14, *P*=.03), with better health status for the digital health group. There were fewer visits to the GP practice nurses 1.5 (digital health) versus 2.5 (usual care), *P*=.03 in comparison to the usual care group. The difference did not reach statistical difference for the median number of visits to general practitioners: 4.0 (digital health) versus 5.5 (usual care), *P*=.06. There was no difference in Beliefs about Medicines Questionnaire (BMQ) or self-reported medicines adherence (MARS, Medication Adherence Report Scale), and no differences in self-reported smoking cessation. Depression measured with the SCL-20 decreased in the EDGE group and increased in the standard care, but the difference in change between groups was not statistically significant. Anxiety measured with the SCL-10A was unchanged in the EDGE group, whereas increased in the usual care group, but again the difference in change between groups was not statistically significant.

XSL•FO
RenderX

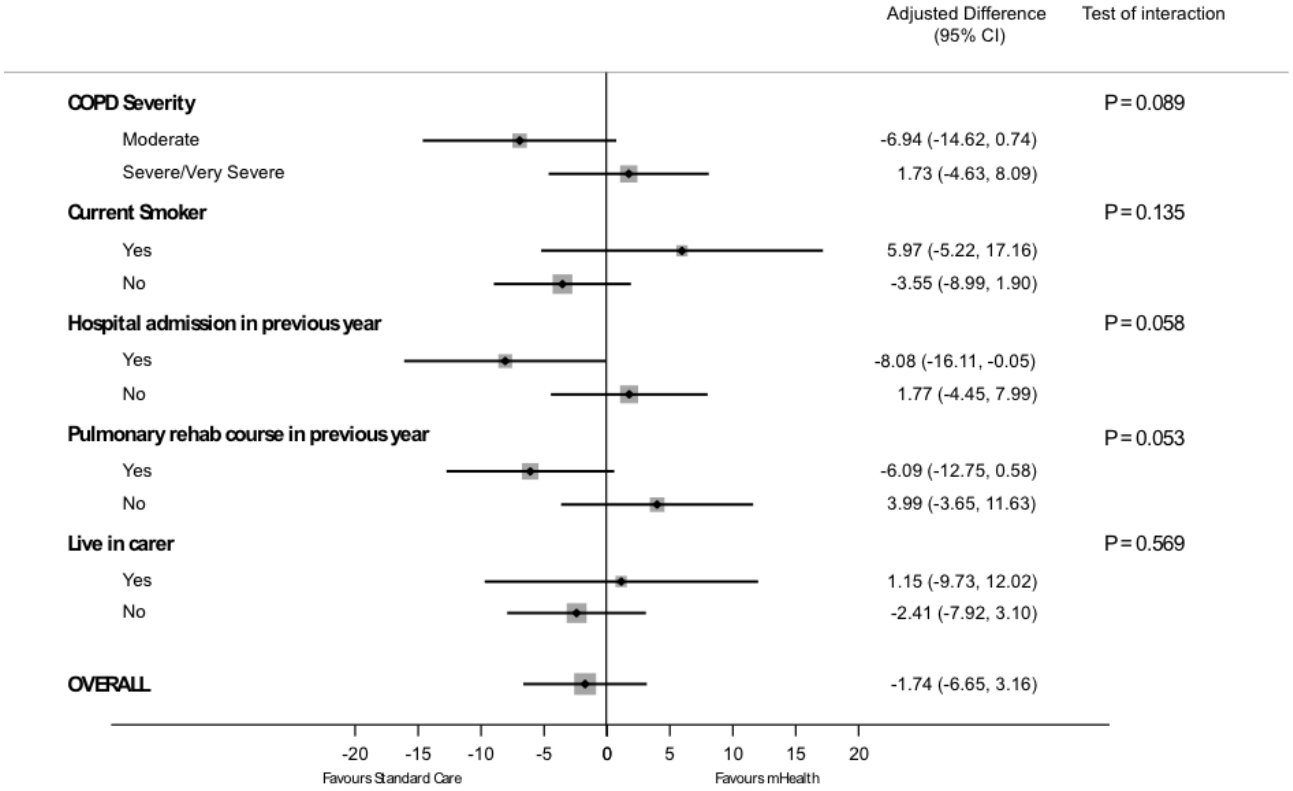
Table 3. Secondary outcomes baseline to 12 months.

Secondary outcomes	Overall effect comparing EDGE[a] and usual care			
	EDGE intervention	Standard care	Adjusted treatment effect[b] (95% CI)	*P* value
Death rates, n (%)	6 (5.5)	4 (7.1)	1.7[c] (−11.0 to 4.9)	.73
Number with at least one admission, n (%)	38 (34.6)	23 (41.1)	0.83[d] (0.56-1.24)	.37
Number of exacerbations, median (IQR[e])	1 (0-2)	1 (0-3)	1.21[f] (0.79-1.84)	.39
Change in BMQ[g], mean (SD[h])	−1.26 (5.33)	−0.80 (4.69)	−0.64[i] (−2.28 to 1.00)	.44
Medication Adherence Report Schedule, mean (SD)	0.17 (2.47)	0.33 (3.65)	0.10[j] (−0.57 to 0.77)	.77
Smoking cessation, n (%)	76 (81.7)	41 (85.4)	1.40[k] (0.24-7.96)	.71
Change in EQ-5D-5L[l], mean (SD)	0.01 (0.21)	−0.08 (0.19)	0.076[i] (0.009-0.14)	.03
Change in SCL-20[m], mean (SD)	−0.04 (0.46)	0.14 (0.56)	−0.13[i] (−0.29 to 0.03)	.10
Change in SCL-10A[n], mean (SD)	0.03 (0.59)	0.13 (0.43)	0.002[i] (−0.19 to 0.19)	.98
Change in lung function, mean (SD)	−0.78 (10.23)	−1.40 (5.67)	0.32[i] (−5.58 to 6.22)	.91
Number of GP[o] contacts (surgery), median (IQR)	4 (2-7)	5.5 (2-10)	-	.06
Number of nurse contacts (surgery), median (IQR)	1.5 (1-3)	2.5 (1-7)	-	.03

[a]EDGE: sElf-management anD support proGrammE.
[b]Adjusted for baseline values and minimization factors.
[c]Difference in proportion.
[d]Relative risk.
[e]IQR: interquartile range.
[f]Incidence rate ratio.
[g]BMQ: Beliefs about Medicines Questionnaire.
[h]SD: standard deviation.
[i]Difference in change in mean.
[j]Difference in mean.
[k]Odds ratio.
[l]EQ-5D-5L: 5-Level EuroQol 5-Dimension Questionnaire.
[m]SCL-20: Standard Checklist 20-item Questionnaire.
[n]SCL-10A: Standard Checklist 10-item Anxiety Measure.
[o]GP: general practitioner.

XSL•FO
RenderX

Figure 2. Subgroup analyses for primary outcome.



Subgroup Analysis of Primary Outcome

Prespecified subgroup analyses for the change in SGRQ-C between intervention groups are summarized in Figure 2. Severity of COPD, current smoker, and no previous hospital admissions in the last year appeared to favor the EDGE system, although interaction terms did not reach statistical significance.

Safety Data

There was no difference in the rates of adverse events or serious adverse events between groups, and none of these events was deemed to be trial or intervention-related. Primary care clinicians were contacted on the same day that a clinically important safety alert was identified.

Costs

The cost of each tablet computer used was £319 and the cost of the pulse oximeter probe was £399. Reviewing patient data on the clinician website using the prioritization algorithms took a median (IQR, interquartile range) of 1.2 h per session (2.9), with a median number of participants accessed for each session of 16 (54). The cost of a hospital admission was estimated as £2900, a GP appointment as £36, and a practice nurse appointment as £11 [44]. The respiratory clinicians reviewing the data reported that up to one phone call a week to a general practitioner, practice or respiratory nurse was required, following patient data review.

Discussion

Principal Findings

This trial provides evidence that use of a tablet computer-based system of monitoring and self-management support does not impact, either positively or negatively on COPD-specific health status over a period of 12 months. However, the finding of a significant improvement in prespecified secondary outcomes supports a beneficial impact on broader measures of health status and number of visits to practice nurses. Although the trial did not test the intervention in a sufficiently large sample to detect a significant difference between patients allocated to intervention and usual care, the effect sizes observed for an improved depression score, reductions in hospital admissions, and reduction in visits to a general practitioner compared with usual care suggest a need for further evaluation.

This trial was carried out in a representative population of individuals with moderate to very severe COPD using minimal exclusion criteria. It employed state-of-the art trial management including use of minimization to balance groups, an unbalanced allocation to allow the possibility of examining the effect of the EDGE platform across a wide range of participants, and measures to avoid contamination. In contrast to several recently reported trials of telehealth in COPD, the equipment was affordable and did not require specialist installation. Features of the EDGE platform addressing previous concerns about telehealth systems include the use of a generic tablet computer, access through icons rather than a keyboard, use of Bluetooth sensors with error-checking algorithms to ensure high-quality data, support for patient interpretation of data, and tailored self-management content [45]. The study protocol mandated patient contact when patient-specific, rather than study-wide thresholds were reached. The prioritization algorithms used to implement personalized alerting ensured that the workload of the nurses reviewing the patient data was kept to manageable levels (a median of 1.2 h per session, with a median of 16 patients accessed per session; 2 sessions per week). This trial

XSL•FO
RenderX

addresses previous concerns about the lack of well-designed randomized controlled trials with appropriate follow-up raised in a recent review of telemedicine interventions for COPD [46].

The use of a self-reported primary outcome measure is a potential limitation, but this was completed before measurement and other data collection at the final visit. Other limitations of the trial design that could be addressed in future work include moving to the use of the updated GOLD classification to characterize participants at baseline. Although no adjustment for testing of multiple secondary outcomes was made, all were prespecified.

As with other trials [22,47,48], and in line with a systematic review of telehealth trials [49], an impact on COPD-specific health status was not observed. Although COPD-specific health status is an important aspect of COPD that has been shown to be responsive to group and one-to-one interventions [50], a digital health intervention used on its own may not replace a more intensive and personalized intervention for many patients. However, other outcomes are also important, including broader measures of health status, psychological health measures, and use of hospital services. It is possible that use of the intervention in clinical practice may lead to changes in the behavior of doctors and nurses, and the way that the health system responds to patients. These changes could further improve outcomes. Further studies to evaluate the system may therefore need to be carried out with clustering of intervention delivery by functional units, for example, individual primary care sites.

The EDGE system was used daily, to send pulse oximeter and diary data, by over 80% of the trial patients. The use of algorithms on the tablet computer to assess quality of monitoring data minimized false alerts from the system. The limited impact observed on some secondary outcomes in this trial could be mediated by regular self-monitoring and review of charted data together with use of the educational and motivational material available on the tablet computer, and, perhaps for some, the wider use of the tablet computer to communicate with family and friends. The effects observed in this trial could be further enhanced by use of better predictive algorithms guiding self- and clinician-management. Monitoring additional parameters (eg, mobility) and prompting additional measurement may also improve algorithm performance. The comprehensive data from this trial, where the external monitoring of data was restricted to ensuring patients' safety, provide a source for predictive modeling [24]. Predictive modeling from these data, along with additional parameters (including mobility) will be integrated into strategies for early treatment of exacerbations [51]. Strategies will include prompting for additional measurement to improve the accuracy of predictions, graded alerts to patients prompting clinical review or starting treatment, and alerts to clinicians to proactively contact patients.

The trial intervention uses a novel implementation of telehealth using a nonproprietary tablet computer designed for integration into day-to-day life and clinical care. In addition to being nonobtrusive, it provides, at relatively low cost in relation to previous telehealth systems, facilities for monitoring, communication, self-management support, and education delivery. The development of the system was carried out iteratively using best practice to involve patients, engineers, and clinicians in repeated testing and assessment [52]. Future iterations of the technology could extend beyond the limited implementation of the EDGE system used in this study [36], for example, in integration with electronic health records.

The underlying approach to implementing digital health within this trial was to provide a system focused around the needs of the patient, with collection of data that can be analyzed over a period of time and used to inform future management. Data from this trial will be used to evaluate the potential of patient-specific tailored alerts, as well as informing the design of multicenter trials to explore cost-effectiveness and potential for reduction in hospital admissions [46]. The dataset would also be available to test algorithms derived from other projects.

Conclusions

Although this clinical trial does not provide evidence for an effect on COPD-specific health status from the EDGE digital health system in comparison with usual care, there may be an overall benefit to patients through better overall health status. If an intervention with the effect sizes for reduced hospital admissions and primary care visits of the magnitude reported in this trial were implemented at scale, it would make an important contribution to monitoring and self-management support for people with moderate to very severe COPD.

Acknowledgments

The authors would like to acknowledge the EDGE (sElf-management anD support proGrammE) chronic obstructive pulmonary disease (COPD) Trial group consisting of, in addition to the authors of this manuscript, Dr Gari Clifford, who provided clinical and scientific input to the design of the intervention and trial and the trial engineering team including Dr Oliver Gibson, who contributed to the development of the EDGE platform. Dr Christy Toms was study manager at the beginning of the trial. Linda Heritage collected data and Stephanie Robinson provided administrative support to the team. Grace Robinson (respiratory consultant and principal investigator), Katherine Thomas (respiratory clinician), and Jacqui Webb (research nurse) led recruitment at the Royal Berkshire Hospital Department of Respiratory Medicine. We are grateful for the support and assistance of the Oxfordshire Respiratory and Home Oxygen Assessment Services.

This publication presents independent research supported from the Department of Health and Wellcome Trust through the Health Innovation Challenge (HIC) Fund commissioned by the Health Innovation Challenge Fund (HICF-1010-032), a parallel funding partnership between the Wellcome Trust and the Department of Health. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health or Wellcome Trust. The trial is sponsored by the University of

XSL•FO
RenderX

Oxford. Oversight for these activities resides with the EDGE (sElf-management anD support proGrammE) chronic obstructive pulmonary disease (COPD) Trial group. LT and AF receive funding from the Oxford National Institute for Health Research (NIHR) Biomedical Research Centre (BRC). AF is an NIHR Senior Investigator. Ethics approval was received from the South Central, Berkshire Research Ethics Committee of the UK National Research Ethics Service (Ethics Ref: 12/SC/0437). The datasets used and analyzed during this study are available from the corresponding author on reasonable request. The trial steering committee (independent members) comprised: Prof Brian McKinstry (Chair), Dr Duncan Keeley, Prof Najib Rahman. The data management committee comprised: Dr Duncan Geddes (Chair), Prof Andy Wilson, Dr Isabel Reading.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables and figures.

[PDF File (Adobe PDF File), 60KB - jmir_v19i5e144_app1.pdf]

Multimedia Appendix 2

TIDIER Checklist (Intervention description).

[PDF File (Adobe PDF File), 54KB - jmir_v19i5e144_app2.pdf]

Multimedia Appendix 3

Action plan provided to all trial participants to manage disease exacerbations.

[PDF File (Adobe PDF File), 4MB - jmir_v19i5e144_app3.pdf]

Multimedia Appendix 4

CONSORT E-Health Checklist V 1.6.1.

[PDF File (Adobe PDF File), 619KB - jmir_v19i5e144_app4.pdf]

References

1. World Health Organization. The global burden of disease update. In: The Global Burden of Disease: 2004 Update. Geneva: World Health Organization; 2004:1-133.
2. National Clinical Guideline Centre. Chronic obstructive pulmonary disease: Management of chronic obstructive pulmonary disease in adults in primary and secondary care. London: NICE; 2010.
3. Okubadejo AA, O'Shea L, Jones P. Home assessment of activities of daily living in patients with severe chronic obstructive pulmonary disease on long-term oxygen therapy. Eur Respir J 1997;10:1572-1575. [doi: 10.1183/09031936.97.10071572]
4. Ståhl E, Lindberg A, Jansson S, Rönmark E, Svensson K, Andersson F, et al. Health-related quality of life is related to COPD disease severity. Health Qual Life Outcomes 2005 Sep 09;3:56 [FREE Full text] [doi: 10.1186/1477-7525-3-56] [Medline: 16153294]
5. Wedzicha JA, Seemungal T. COPD exacerbations: defining their cause and prevention. The Lancet 2007;370:786-796.
6. Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004 Jun 15;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
7. Seymour JM, Moore L, Jolley CJ, Ward K, Creasey J, Steier JS, et al. Outpatient pulmonary rehabilitation following acute exacerbations of COPD. Thorax 2010 May;65(5):423-428. [doi: 10.1136/thx.2009.124164] [Medline: 20435864]
8. Bourbeau J, Julien M, Maltais F. Reduction of hospital utilization in patients with chronic obstructive pulmonary disease: a disease-specific self-management intervention. Arch Intern Med 2003;163:585-591.
9. Rea H, McAuley S, Stewart A, Lamont C, Roseman P, Didsbury P. A chronic disease management programme can reduce days in hospital for patients with chronic obstructive pulmonary disease. Intern Med J 2004 Nov;34(11):608-614. [doi: 10.1111/j.1445-5994.2004.00672.x] [Medline: 15546454]
10. Effing T, Monninkhof EM, van der Valk PD, van der Palen J, van Herwaarden CL, Partidge MR, et al. Self-management education for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2007 Oct 17(4):CD002990. [doi: 10.1002/14651858.CD002990.pub2] [Medline: 17943778]
11. Baraniak A, Sheffield D. The efficacy of psychologically based interventions to improve anxiety, depression and quality of life in COPD: a systematic review and meta-analysis. Patient Educ Couns 2011 Apr;83(1):29-36. [doi: 10.1016/j.pec.2010.04.010] [Medline: 20447795]

XSL•FO
RenderX

12. Bischoff EW, Hamd DH, Sedeno M. Effects of written action plan adherence on COPD exacerbation recovery. Thorax 2011;66:26-31. [doi: 10.1136/thx.2009.127621]
13. Fan VS, Gaziano JM, Lew R, Bourbeau J, Adams SG, Leatherman S, et al. A comprehensive care management program to prevent chronic obstructive pulmonary disease hospitalizations: a randomized, controlled trial. Ann Intern Med 2012 May 15;156(10):673-683. [doi: 10.7326/0003-4819-156-10-201205150-00003] [Medline: 22586006]
14. 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 May;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
15. Lundell S, Holmner Å, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015 Jan;109(1):11-26 [FREE Full text] [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]
16. McLean S, Nurmatov U, Liu JLY, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease: Cochrane Review and meta-analysis. Br J Gen Pract 2012 Nov;62(604):739-749 [FREE Full text] [doi: 10.3399/bjgp12X658269] [Medline: 23211177]
17. Aaron SD, Donaldson G, Whitmore G. Time course and pattern of COPD exacerbation onset. Thorax 2011;67:238-243.
18. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-127. [doi: 10.1258/jtt.2009.090812] [Medline: 20197355]
19. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, et al. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012 Jun 21;344:e3874. [doi: 10.1136/bmj.e3874]
20. Fairbrother P, Pinnock H, Hanley J, McCloughan L, Sheikh A, Pagliari C, et al. Continuity, but at what cost? the impact of telemonitoring COPD on continuities of care: a qualitative study. Prim Care Respir J 2012 Sep;21(3):322-328 [FREE Full text] [doi: 10.4104/pcrj.2012.00068] [Medline: 22875143]
21. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van HH, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Arch Intern Med 2012 May 28;172(10):773-779 [FREE Full text] [doi: 10.1001/archinternmed.2012.256] [Medline: 22507696]
22. 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;347:6070. [doi: 10.1136/bmj.f6070]
23. Williams V, Price J, Hardinge M, Tarassenko L, Farmer A. Using a mobile health application to support self-management in COPD: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e392-e400 [FREE Full text] [doi: 10.3399/bjgp14X680473] [Medline: 24982491]
24. Shah SA, Velardo C, Gibson OJ, Rutter H, Farmer A, Tarassenko L. Personalized alerts for patients with COPD using pulse oximetry and symptom scores. In: Conf Proc IEEE Eng Med Biol Soc. 2014 Presented at: Conf Proc IEEE Eng Med Biol Soc; 2014; Chicago p. 3164-3167 URL: http://ieeexplore.ieee.org/document/6944294/ [doi: 10.1109/EMBC.2014.6944294]
25. Hardinge M, Rutter H, Velardo C. Using a mobile health application to support self-management in COPD-development of alert thresholds derived from variability in self-reported and measured clinical variables. Am J Respir Crit Care Med 2014;189:A1396.
26. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012 Jul;2(4):e001574 [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]
27. Hardinge M, Rutter H, Velardo C, Shah SA, Williams V, Tarassenko L, et al. Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study. BMC Med Inform Decis Mak 2015 Jun 18;15:1-10 [FREE Full text] [doi: 10.1186/s12911-015-0171-5] [Medline: 26084626]
28. Farmer A, Toms C, Hardinge M, Williams V, Rutter H, Tarassenko L. Self-management support using an internet-linked tablet computer (the EDGE platform)-based intervention in chronic obstructive pulmonary disease: protocol for the EDGE-COPD randomised controlled trial. BMJ Open 2014 Jan;4(1):e004437 [FREE Full text] [doi: 10.1136/bmjopen-2013-004437] [Medline: 24401729]
29. Dumville JC, Hahn S, Miles JNV, Torgerson DJ. The use of unequal randomisation ratios in clinical trials: a review. Contemp Clin Trials 2006 Feb;27(1):1-12. [doi: 10.1016/j.cct.2005.08.003] [Medline: 16236557]
30. Wedzicha JA, Decramer M, Ficker JH, Niewoehner DE, Sandström T, Taylor AF, et al. Analysis of chronic obstructive pulmonary disease exacerbations with the dual bronchodilator QVA149 compared with glycopyrronium and tiotropium (SPARK): a randomised, double-blind, parallel-group study. Lancet Respir Med 2013 May;1(3):199-209. [doi: 10.1016/S2213-2600(13)70052-3]
31. Bucknall CE, Miller G, Lloyd SM, Cleland J, McCluskey S, Cotton M, et al. Glasgow supported self-management trial (GSuST) for patients with moderate to severe COPD: randomised controlled trial. BMJ 2012 Mar 06;344(1):e1060. [doi: 10.1136/bmj.e1060]

XSL•FO
RenderX

32. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the patient health questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122:86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]
33. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114:163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
34. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
35. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2015:CD003793. [doi: 10.1002/14651858.CD003793.pub3/full]
36. Hey SP, Kimmelman J. The risk-escalation model: a principled design strategy for early-phase trials. Kennedy Inst Ethics J 2014;24:121-139.
37. Meguro M, Barley EA, Spencer S, Jones PW. Development and validation of an improved, COPD-specific version of the St. George respiratory questionnaire. Chest 2007 Aug;132(2):456-463. [doi: 10.1378/chest.06-0702] [Medline: 17646240]
38. Horne R, Weinman J, Hankins M. The beliefs about medicines questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. Psychol Health 1999 Jan;14(1):1-24. [doi: 10.1080/08870449908407311]
39. Derogatis LRL, Lipman R, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. Psychopharmacol Bull 1973;9:13-28.
40. Derogatis LR, Lipman R, Rickels K. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behavioral Science 1974;19:1-15.
41. EuroQol Group. EuroQol: a new facility for the measurement of health related quality of life. Health Pol 2001;16:199-208.
42. Welte T, Miravitlles M, Hernandez P, Eriksson G, Peterson S, Polanowski T, et al. Efficacy and tolerability of budesonide/formoterol added to tiotropium in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2009 Oct 15;180(8):741-750. [doi: 10.1164/rccm.200904-0492OC] [Medline: 19644045]
43. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2011(7):CD007718. [doi: 10.1002/14651858.CD007718.pub2] [Medline: 21735417]
44. Curtis L, Burns A. Unit Costs of Health & Social Care 2016. Canterbury, UK: PSSRU; 2016.
45. Velardo C, Shah SA, Gibson O, Clifford G, Heneghan C, Rutter H, et al. Digital health system for personalised COPD long-term management. BMC Med Inform Decis Mak 2017 Feb 20;17(1):1-13 [FREE Full text] [doi: 10.1186/s12911-017-0414-8] [Medline: 28219430]
46. Goldstein RS, O'Hoski S. Telemedicine in COPD: time to pause. Chest 2014 May;145(5):945-949. [doi: 10.1378/chest.13-1656] [Medline: 24798834]
47. Steventon A, Bardsley M, Doll H, Tuckey E, Newman SP. Effect of telehealth on glycaemic control: analysis of patients with type 2 diabetes in the whole systems demonstrator cluster randomised trial. BMC Health Serv Res 2014 Aug 06;14:1-12 [FREE Full text] [doi: 10.1186/1472-6963-14-334] [Medline: 25100190]
48. Schou L, Ostergaard B, Rydahl-Hansen S, Rasmussen LS, Emme C, Jakobsen AS, et al. A randomised trial of telemedicine-based treatment versus conventional hospitalisation in patients with severe COPD and exacerbation - effect on self-reported outcome. J Telemed Telecare 2013 Apr 23;19:160-165. [doi: 10.1177/1357633X13483255] [Medline: 23612519]
49. Gregersen TL, Green A, Frausing E, Ringbæk T, Brøndum E, Suppli UC. Do telemedical interventions improve quality of life in patients with COPD? a systematic review. Int J Chron Obstruct Pulmon Dis 2016 Apr;11:809-822 [FREE Full text] [doi: 10.2147/COPD.S96079] [Medline: 27143872]
50. Zwerink M, Brusse-Keizer M, van der Valk PD, Zielhuis GA, Monninkhof EM, van der Palen J, et al. Self management for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2014 Mar 19(3):CD002990. [doi: 10.1002/14651858.CD002990.pub3] [Medline: 24665053]
51. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. J Med Internet Res 2017 Mar 07;19(3):e69 [FREE Full text] [doi: 10.2196/jmir.7207] [Medline: 28270380]
52. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. BMJ 2007 Mar 03;334(7591):455-459 [FREE Full text] [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]

Abbreviations

BMQ: Beliefs about Medicines Questionnaire
COPD: chronic obstructive pulmonary disease
EDGE: sElf-management anD support proGrammE
EQ-5D-5L: 5-Level EuroQol 5-Dimension Questionnaire
IQR: interquartile range

XSL•FO
RenderX

MARS: Medication Adherence Report Scale

Edited by G Eysenbach; submitted 08.12.16; peer-reviewed by H Pinnock, N Amoyal Pensak, K Phanareth; comments to author 26.01.17; revised version received 02.03.17; accepted 14.03.17; published 03.05.17.

Please cite as:
Farmer A, Williams V, Velardo C, Shah SA, Yu LM, Rutter H, Jones L, Williams N, Heneghan C, Price J, Hardinge M, Tarassenko L
Self-Management Support Using a Digital Health System Compared With Usual Care for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial
J Med Internet Res 2017;19(5):e144
URL: http://www.jmir.org/2017/5/e144/
doi:10.2196/jmir.7116
PMID:28468749

©Andrew Farmer, Veronika Williams, Carmelo Velardo, Syed Ahmar Shah, Ly-Mee Yu, Heather Rutter, Louise Jones, Nicola Williams, Carl Heneghan, Jonathan Price, Maxine Hardinge, Lionel Tarassenko. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 03.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Embodied Conversational Agents in Clinical Psychology: A Scoping Review

Simon Provoost[1,2], MSc; Ho Ming Lau[2,3], MSc, MDiv; Jeroen Ruwaard[1,2], PhD; Heleen Riper[1,2,3,4], PhD

[1]Department of Clinical, Neuro & Developmental Psychology, Section Clinical Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]EMGO Institute for Health and Care Research, VU University Medical Centre, Amsterdam, Netherlands

[3]GGZ inGeest, Amsterdam, Netherlands

[4]Telepsychiatry Unit, Faculty of Health Science, University of Southern Denmark, Odense, Denmark

Corresponding Author:
Simon Provoost, MSc
Department of Clinical, Neuro & Developmental Psychology, Section Clinical Psychology
Faculty of Behavioural and Movement Sciences
Vrije Universiteit Amsterdam
Van der Boechorststraat 1
Amsterdam, 1081 BT
Netherlands
Phone: 31 205985059
Fax: 31 205988758
Email: s.j.provoost@vu.nl

Abstract

Background: Embodied conversational agents (ECAs) are computer-generated characters that simulate key properties of human face-to-face conversation, such as verbal and nonverbal behavior. In Internet-based eHealth interventions, ECAs may be used for the delivery of automated human support factors.

Objective: We aim to provide an overview of the technological and clinical possibilities, as well as the evidence base for ECA applications in clinical psychology, to inform health professionals about the activity in this field of research.

Methods: Given the large variety of applied methodologies, types of applications, and scientific disciplines involved in ECA research, we conducted a systematic scoping review. Scoping reviews aim to map key concepts and types of evidence underlying an area of research, and answer less-specific questions than traditional systematic reviews. Systematic searches for ECA applications in the treatment of mood, anxiety, psychotic, autism spectrum, and substance use disorders were conducted in databases in the fields of psychology and computer science, as well as in interdisciplinary databases. Studies were included if they conveyed primary research findings on an ECA application that targeted one of the disorders. We mapped each study's background information, how the different disorders were addressed, how ECAs and users could interact with one another, methodological aspects, and the study's aims and outcomes.

Results: This study included N=54 publications (N=49 studies). More than half of the studies (n=26) focused on autism treatment, and ECAs were used most often for social skills training (n=23). Applications ranged from simple reinforcement of social behaviors through emotional expressions to sophisticated multimodal conversational systems. Most applications (n=43) were still in the development and piloting phase, that is, not yet ready for routine practice evaluation or application. Few studies conducted controlled research into clinical effects of ECAs, such as a reduction in symptom severity.

Conclusions: ECAs for mental disorders are emerging. State-of-the-art techniques, involving, for example, communication through natural language or nonverbal behavior, are increasingly being considered and adopted for psychotherapeutic interventions in ECA research with promising results. However, evidence on their clinical application remains scarce. At present, their value to clinical practice lies mostly in the experimental determination of critical human support factors. In the context of using ECAs as an adjunct to existing interventions with the aim of supporting users, important questions remain with regard to the personalization of ECAs' interaction with users, and the optimal timing and manner of providing support. To increase the evidence base with regard to Internet interventions, we propose an additional focus on low-tech ECA solutions that can be rapidly developed, tested, and applied in routine practice.

XSL•FO
RenderX

(J Med Internet Res 2017;19(5):e151) doi:10.2196/jmir.6553

KEYWORDS

eHealth; review; embodied conversational agent; human computer interaction; clinical psychology; mental disorders; intelligent agent; health behavior

Introduction

Background

Internet-based interventions can be effective in the treatment of various mental disorders compared with care-as-usual (eg, face-to-face treatment) and waiting list control groups [1]. Many interventions, especially those aimed at mood, anxiety, and substance use disorders, are based on cognitive behavioral therapy (CBT). These interventions can be unguided or guided, with guidance typically being provided by licensed health professionals or trained volunteers. Guided interventions are typically more clinically effective than unguided ones [2-4]. Their superiority is most likely explained by the interaction between the participant and the person providing guidance, and although concepts such as treatment adherence have been suggested as a working mechanism [5], that is, human support having a positive effect on adherence, and in turn on effectiveness, it remains unclear how exactly human support accounts for the difference. Thus, there is good reason to explore whether the gap between guided and unguided interventions can be bridged. In this paper we focus on a potential automated solution: Embodied conversational agents (ECAs).

Embodied Conversational Agents

ECAs can be defined as "more or less autonomous and intelligent software entities with an embodiment used to communicate with the user" [6]. Examples of real-world ECAs are interactive characters in video games and virtual characters that assist customers in Web stores. Conceptually, ECAs consist of three components [7]. The first is an application interface that allows users to communicate with the ECA and provide it with information. These interfaces can range from Web-based questionnaires to real-time audio and video input. Second, ECAs are endowed with computer models that give them their "mental" capacities, essentially programmed knowledge used to reason over the factual "observations" derived from the interface. Such models can range from concise decision trees in which different answers on a questionnaire lead to different responses by the ECA, to machine learning–based algorithms that classify real-time video and audio input into a user's emotional state, allowing the ECA to react empathically. Third, ECAs have an embodiment, or visual representation, which allows them to communicate with users verbally or nonverbally. Embodiments can range from virtual human characters on computer screens to robots, and communication from text messages to human communication modalities such as speech, gestures, and facial expressions. There are advantages and disadvantages to whatever implementation of the design aspects is chosen. Highly advanced ECAs, for example, those using multimodal and real-time user input such as video recordings and natural language, can be more believable than simplistic ones, but their complexity means that they require more development time, greater technological expertise, and that mistakes (eg, in interpreting semantics of natural language) become more likely. Low-tech approaches based, for example, on decision tree mechanisms or relatively simplistic graphics can be utilized to deal with these problems, but they also make for a less realistic experience. These kinds of trade-offs make finding the optimal configuration in a certain setting a nontrivial task.

Are ECAs Ready for Clinical Practice?

Working with an existing Web- and mobile-based cognitive behavioral treatment for depression [8], our interest lies with techniques that can be applied in clinical practice. Given the many design decisions that can be made with respect to an ECA's configuration, it is not immediately evident what an ECA should look like, and how it should behave in our context, namely as a bridge between guided and unguided interventions in clinical psychology. Paradigms exist that offer concrete design guidelines with respect to some (eg, [9]), or many (eg, [10]), of the aspects of ECA development. However, their empirical foundations generally rest on outcome measures such as "user satisfaction," "engagement with the ECA," or "intention to use," and the context is not necessarily that of clinical psychology. Although such measures might be indicative of clinical effectiveness, they do not necessarily translate to the clinical outcomes we aim to improve. For example, even though users might be fully satisfied with an ECA, this does not necessarily mean that the average treatment outcome (eg, a significant reduction on a clinical measure of depression) will improve. Our chances to successfully bridge the gap between guided and unguided interventions will increase if we can determine how we can apply ECA technology with respect to improving clinical outcomes.

The interdisciplinary nature of ECA research makes almost any intervention that includes an ECA inherently complex. The UK Medical Research Council's (MRC) framework for complex interventions [11] defines four phases through which such interventions move before being fully embedded in practice: development, piloting, evaluation, and implementation. The difference between the piloting and evaluation phase is crucial. Whereas interventions might still be subject to changes in the piloting phase, the evaluation phase is characterized by a focus on clinical outcomes and a more rigorous study design. Although routine practice sometimes evolves along different lines, as a golden rule, it is only once an intervention has successfully moved through the evaluation phase and can be considered effective and safe to use, that it becomes of practical value to psychologists.

Scoping Review

As a first step toward designing our own ECA, we wanted to review the relevant literature in a systematic manner to find out how ECAs had previously been used in psychotherapy, and to what extent the approaches taken were supported by evidence.

XSL•FO
RenderX

Initial exploration of the literature to determine a suitable review method revealed a large variety of ECA applications, study designs, and outcome measures, such that a traditional systematic review (more emphasis on hard evidence) or meta-analytic approach (requires comparable outcomes) appeared inappropriate. We therefore adopted the scoping review method [12]. Scoping reviews aim to map the key concepts underpinning a research area, as well as the main sources and types of evidence that are available [13]. Compared with traditional systematic reviews, scoping reviews address broader topics where many different study designs might be applicable, and do not emphasize quality assessment (eg, the power of the study or nature of control groups) of the included studies, as the research questions are less specific [12].

This scoping review aims to inform health professionals about the technological and clinical possibilities and evidence base for ECA applications in clinical psychology, and to provide an overview of the activity in this field of research.

Methods

Study Design

We adopted the Arksey and O'Malley framework for scoping reviews [12], which distinguishes five different stages: (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) charting the data, and (5) collating, summarizing, and reporting the results. We took into account recommendations about using an iterative team approach throughout stages (1) to (4) [14,15] by having regular discussions with other team members. As theoretical underpinnings to scoping reviews, as well as transparency about the process by which results are obtained, are often lacking [16], we also made an attempt to provide clear concept definitions. Stages (1) to (4) are described in this section and stage (5) in the Results section.

Identifying the Research Question

Given the generic features of human support in psychotherapy, techniques seen in the treatment of disorders other than depression might be applicable in our context as well. Hence, we broadened our scope to include other common mental health disorders known to be a target for e-Mental Health interventions, namely mood disorders, anxiety disorders, post-traumatic stress disorder (PTSD), psychotic disorders, eating disorders, autism spectrum disorders (ASDs), and substance-related disorders.

Study Identification

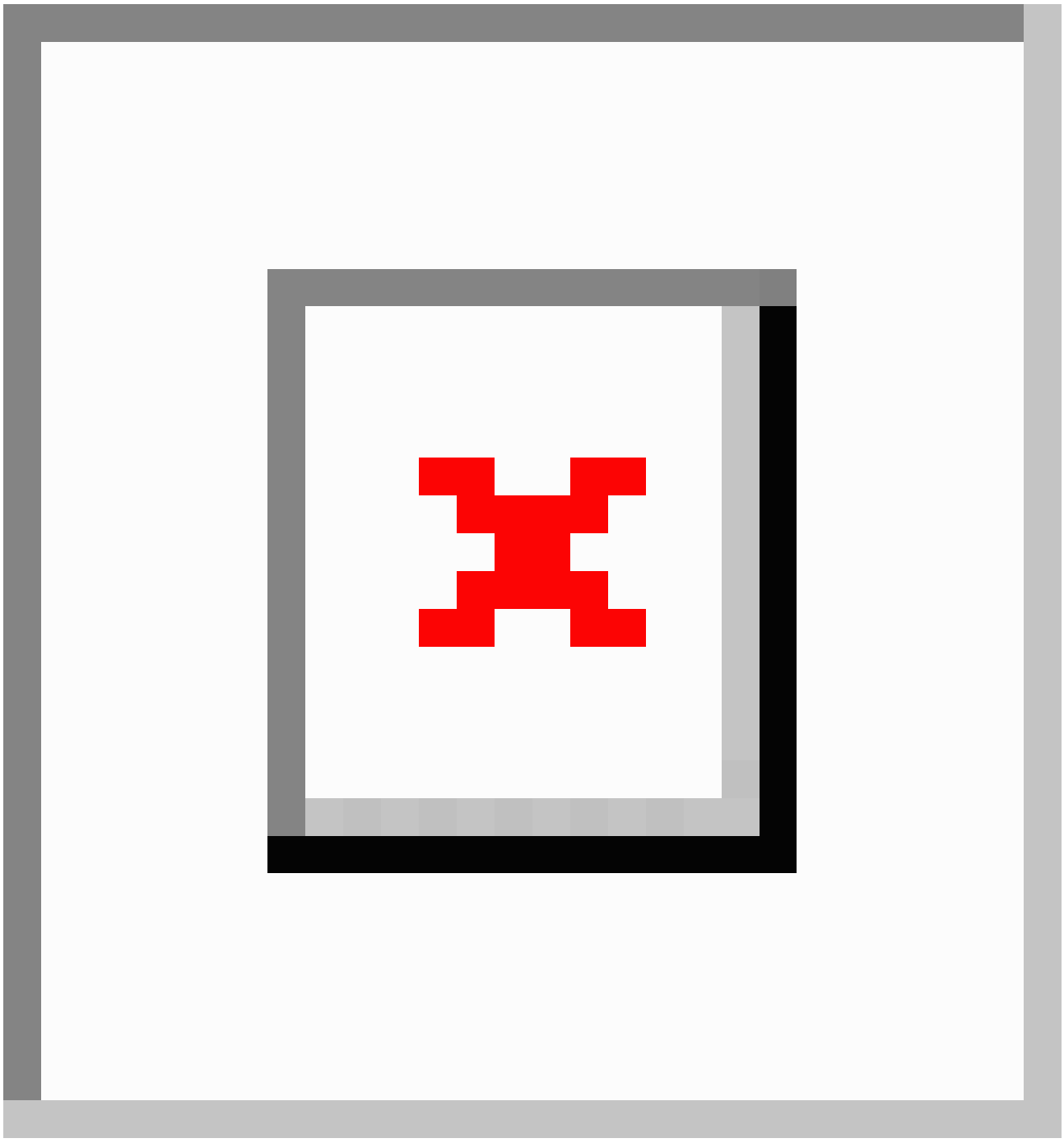
Our generic search query was as follows:

> *embodied conversational agent AND mood disorder OR anxiety disorder OR psychotic disorder OR eating disorder OR autism spectrum disorder OR substance-related disorder*

Our list of search terms for the ECA concept included those we observed to be most common, for example, "virtual agent," "virtual character," "virtual human," or "avatar." The terms for the mental disorders were based on those found in PubMed's MeSH (medical subject headings) index. We searched both psychology and computer science databases, including PubMed (psychology), ScienceDirect (interdisciplinary), WebOfScience (interdisciplinary), ACM (Association for Computing Machinery) Digital Library (computer science), and SpringerLink (artificial intelligence). The detailed search strings and search procedures are described in Multimedia Appendix 1. The final search was conducted in, and included articles published up to, July 2015. References were stored in Microsoft Excel, and duplicates were removed.

XSL•FO
RenderX

Figure 1. Examples of embodied conversational agent embodiments: top-left: emotional reinforcement with a smiley face; bottom-left: virtual psychiatric nurse; middle: SPARX's (Smart, Positive, Active, Realistic, X-factor thoughts) guide character; top-right: SimSensei Kiosk virtual counselor; bottom-right: humanoid robot KASPAR.



Study Selection

Study selection was conducted by two independent reviewers (SP & HL), who screened titles and abstracts on the sequential eligibility criteria, and then assessed the full-text versions of the remaining articles. A third reviewer (JR) was consulted in case of disagreement. We included full articles that:

(1) were written in English, (2) included an ECA in (3) an applied mental health context, (4) conveyed primary research findings, (5) targeted a mood, anxiety, psychotic, eating, autism spectrum, or substance-related disorder, and (6) described an experimental or focus group study.

Regarding criterion (2), for a software entity to be considered an ECA, it required to, first, have a virtual or physical embodiment (eg, Figure 1), second, interact with a user and, third, have a reasonable sense of agency, meaning its behavior had to be autonomous, and the software entity had to exhibit some form of reasoning. As for criterion (3), an applied mental health context implied that ECAs were used in an application that aimed to improve patient outcomes directly related to the targeted disorder, or that a reasonable argument could be made that the proposed application could eventually be used to do so.

XSL•FO
RenderX

Charting the Data

Data extraction was conducted independently by two reviewers (SP & HL). Concepts were mapped in four categories: (1) meta-information, (2) study characteristics, (3) study methodology, and (4) ECA characteristics. Precise definitions of the concepts are listed in Multimedia Appendix 2.

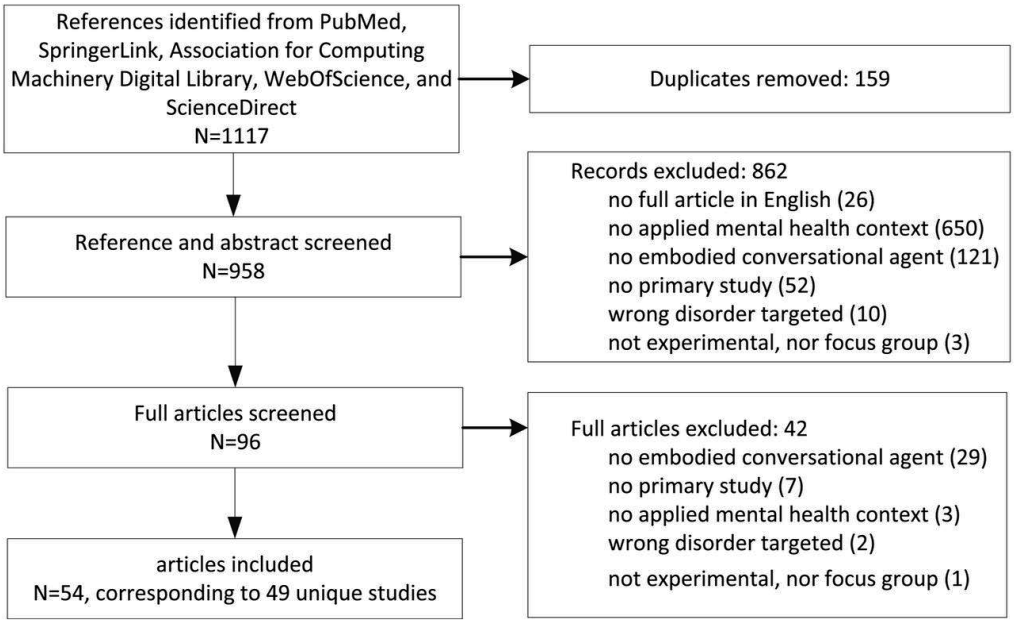
During the data collection process, several concept definitions were refined. In trying to map the studies' intended interventions and provide a taxonomy, we found that from a low level of abstraction, interventions targeted very specific behaviors or skills. Our listing grew so expansive that we considered a higher-level classification to be useful. Something similar could be said for the ECAs' social roles, which were difficult to define precisely given the large variety in applications. In our attempt to provide a useful taxonomy, we tagged all studies with their most predominant social role and intended intervention during discussions with all reviewers present. In these discussions, definitions of the different outcome types and development phases (based on the MRC framework for complex interventions [11]) were refined as well, until each of the studies could be tagged unambiguously. The resulting definitions are also listed in Multimedia Appendix 2.

Results

Main Findings

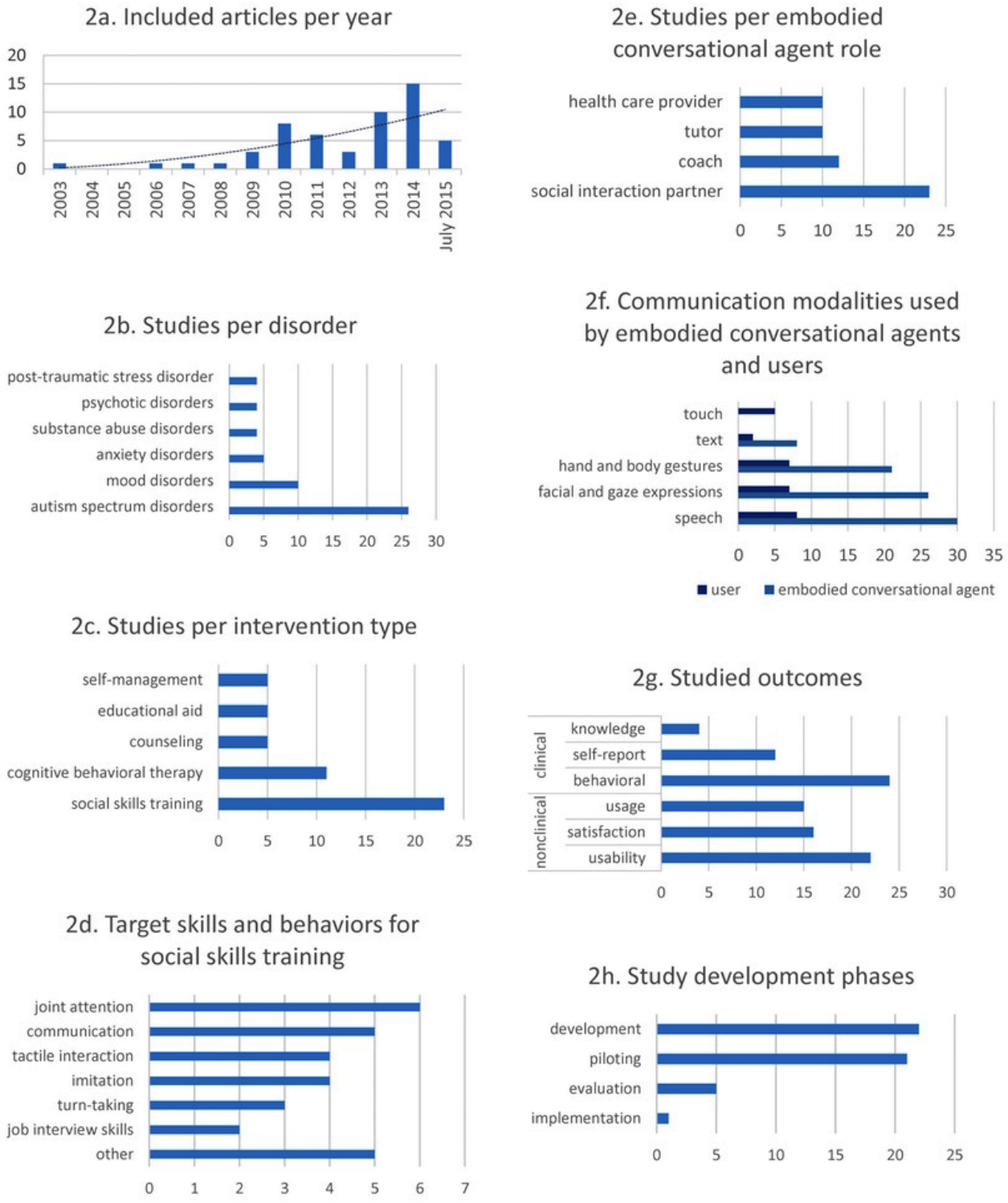
The search identified N=1117 references. After the removal of duplicates, N=958 references remained. Next, a total of N=862 references were excluded by both reviewers after screening titles and abstracts. Of the remaining N=96 references, the reviewers did not agree about the inclusion based on screening in 78 instances, primarily because it was not entirely clear from the limited information provided by title and abstract whether the ECA inclusion criterion was satisfied (N=40). After full assessment of the 96 remaining articles, disagreement on 8 articles remained and was resolved in a discussion with the third reviewer. Finally, N=54 articles were considered eligible for full review. These 54 articles corresponded to N=49 unique studies (Figure 2). Multimedia Appendices 3 and 4, respectively, list the included studies' intervention and ECA, and experimental design characteristics. Figure 3 depicts the most important results of the overall summative analysis and Figure 1 gives an illustration of some of the ECAs described in this review, more specifically emotional reinforcement with a smiley face [17] in the top-left, a virtual psychiatric nurse [18] in the bottom-left, SPARX (Smart, Positive, Active, Realistic, X-factor thoughts)'s guide character [19] in the middle, the SimSensei Kiosk virtual counselor [20] in the top-right, and humanoid robot KASPAR [21] in the bottom-right.

Figure 2. Flowchart describing study identification and selection.



XSL•FO
RenderX

Figure 3. Results of the overall summative evaluation. Note that the categories were not mutually exclusive besides the intervention types and development phases.



Autism Spectrum Disorders

Over half the studies (n=26) targeted ASDs (Table 1). They either involved a form of social skills training (n=21), aimed at a variety of target behaviors or skills (Figure 3, graph d), or were presented as an educational aid (n=5) to accommodate children with autism's special needs. Autism was the only disorder targeted with robotic applications (n=12), and most of the virtual characters appeared in serious games (n=8). Most ECAs assumed the role of a social interaction partner (n=18) or tutor (n=8), and in two studies [22,23] social interaction partners were accompanied by a coach who provided additional feedback on their performance. The relative predictability of ECA behavior compared with that of humans, the possibility to repeatedly practice certain behaviors more often than with human partners, and children with autism's fascination for technology were important reasons to explore the use of ECA technology in autism treatment.

XSL•FO
RenderX

Table 1. Summative results per disorder.

Total number of studies	n=26 ASD[a]	n=10 Depression	n=5 Anxiety	n=4 PTSD[b]	n=4 Psychotic	n=4 Substance use
Interventions						
Social skills training	21	1			2	
Educational aid	5					
CBT[c]		4	2	1		4
Counseling		3	3	2		
Self-management		2		1	2	
Platform						
Serious game	8	2	1	1		
Stand-alone Software	4	2	3	1	2	
Robotics	12					
Virtual reality	1		1		1	
Web based	1	6		2	1	4
Social role						
Social interaction partner	18	1	3		2	
Tutor	8		1		1	
Coach	2	6		3	1	2
Health care provider		5	2	1	2	2
ECA[d] human communication modalities						
Speech	18	5	3	2	4	2
Facial and gaze expressions	15	5	4	2	3	1
Hand and body gestures	14	3	3	2	2	1
Text	2	3	1	1		1
Touch						
User human communication modalities						
Speech	3	2	3	1	1	1
Facial and gaze expressions	3	1	3	1		1
Hand and body gestures	4	1	3	1		
Text	1	1		1		
Touch	5					
Personalization						
Static user model	4					
Dynamic user model	2	7	1	2	3	3
Menu-based dialog	2	4		1	4	1
Natural language dialog		2	1	2		1
Development phase						
Development	13	5	2	4		1
Piloting	11	2	3		3	2
Evaluation	2	2			1	1
Implementation		1				
Outcomes						

XSL•FO
RenderX

Total number of studies	n=26	n=10	n=5	n=4	n=4	n=4
	ASD[a]	Depression	Anxiety	PTSD[b]	Psychotic	Substance use
Usability	11	5	2	3	3	2
Satisfaction	3	8	1	2	4	2
Usage	11	1	1		2	1
Behavioral	17	2	3		3	
Self-report	2	4	3	1	2	2
Knowledge	4					
Study participants						
Mean N (SD)	14.6 (14.2)	93.8 (108.8)	107.4 (141.0)	293 (306.7)	19.8 (11.9)	380.5 (624.8)
Min N, max N	1, 49	8, 351	15, 351	10, 700	10, 37	35, 1317
Clinical sample	21	7		1	4	2
Adult sample	5	7	4	4	4	4

[a]ASD: autism spectrum disorder.

[b]PTSD: post-traumatic stress disorder.

[c]CBT: cognitive behavioral therapy.

[d]ECA: embodied conversational agent.

Psychotherapeutic Interventions

A first group of ASD studies targeted nonverbal communication skills. In *joint-attention skills* training, virtual characters [17,22,24-26] or a robot [27] would nonverbally cue targets by pointing or gazing at them, after which children were instructed to pay attention to the targets [17,22,24-27]. *Imitation skills* were taught by asking children to repeat nonverbal gestures made by robots [28-31]. Other applications focused on *tactile interaction* by letting children play with robots equipped with tactile sensors [31-34], *turn-taking* behavior through playing games with a virtual character [17] or robot [30], and *facial and emotion recognition* by reconstructing faces of dynamic photographs [22].

Applications focusing on verbal skills used ECAs to teach children *communication skills* such as general conversation [35,36] and antibullying strategies [37], and stimulated cooperation in multiplayer games [21,38,39]. Lastly, a *job interview training* application allowed users to practice with a virtual job interviewer [23]. It is in this group of ASD studies that we found the only three applications focusing on adults [23,35,39].

Four of the educational aid applications involved virtual tutors, targeting vocabulary [40], daily-life skills [41], general educational needs of children with ASDs [42], and comprehension of idioms [43]. One study used a humanoid robot to teach children about body postures and aid them in their sense of body consciousness [44].

ECA Technology

From a technological perspective, the focus in many of these studies was on letting ECAs communicate with users through human communication modalities, most notably speech (n=18), facial and gaze expressions (n=15), and hand and body gestures (n=14). A notable exception was touch (n=5), which could only be used by human users.

Few studies employed user models to personalize subsequent interactions. Static user models were mostly used to call the user by name [24,25,27,42], and only two studies employed dynamic user models to personalize subsequent interactions. In one study [26], emotion recognition was used to structure the narrative of a game, and in another [23] the interactions were structured according to the user's performance and level of rapport with the virtual job interviewer. Only two studies allowed users to enter into a dialog with the ECA, both using menu-based dialogs [23,36].

Evidence

Although the most commonly studied outcomes were behavioral (n=17), and most of the studies (n=21) were conducted with clinical samples, the sample sizes were generally very small (mean 14.6 [SD 14.2]), and in many cases the behavioral outcomes were short term and restricted to experimental settings (eg, a researcher's observation of child-robot interaction during an experiment [31]). Most studies were still in the development (n=13) and piloting (n=11) phase.

Two ASD studies had moved beyond the development and piloting phase. In the first study [23], the Web platform virtual reality (VR) job interview training was evaluated in a randomized controlled trial (N=26). Users were subjected to an interview with a job interviewer, and could ask a coach for feedback regarding their performance. Users could interact with the interviewer through a menu-based system, but they were also given the option of speaking their choices out loud to help them practice their verbal skills. The interviewer could be configured to have different personalities (from friendly and easy-going to mean and asking illegal questions), and based her responses on the user's answers and a user model that kept track of the level of rapport between the two. The interviewer used

XSL•FO
RenderX

speech output and conveyed dynamic emotional states through facial and gaze expressions. There was a significant improvement in the user's researcher-scored interview skills in a role-played interview, as well as self-confidence, compared with a control group that received no intervention. Although this result is promising, generalizability to real-world settings and the application's effectiveness compared with conventional job interview skills training programs remain unclear.

The second study used a computer program to improve joint attention, and emotion and face recognition in children with autism [22]. It involved exercises with dynamic realistic photographs of human faces, "coming to life" after successful completion. This intervention also included an animal avatar coach embodied by a realistic photograph that provided additional motivational reinforcement. Following a randomized controlled study (N=49), the intervention was found to be more effective in comparison to a control condition in which children used drawing software. Both children with high- and low-functioning autism improved in terms of emotion recognition and observed social skills, while children with high-functioning autism also improved in facial recognition. Similar to the other evaluation study, it remains unclear how the intervention would compare with conventional interventions targeting similar social skills.

Depression

A total of n=10 studies targeted depression. These studies revolved around CBT interventions (n=4), counseling (n=3), self-management skills (n=2), and social skills training (n=1). Most (n=6) of the applications were Web based, and the social roles fulfilled by ECAs were a coach (n=6) or health care provider (n=5). The anonymity provided by ECAs, their availability compared with humans, their nonjudgmental nature, and the ability for people to practice social interaction in a safe environment were important reasons to explore the use of ECAs in depression treatment.

Psychotherapeutic Interventions

The CBT-based applications targeted *symptoms of depression* in general, with a virtual coach guiding people with depression through a Web-based intervention [45], a photograph of a clinician embodying weekly feedback in a Web-based intervention [46], and a fantasy character guiding users through a serious game [19]. In [47], requirements for the virtual agent used in [45] were determined through a focus group study.

A second group of applications explored elements of counseling with a virtual agent, more specifically the elicitation of *self-disclosure*, that is, getting people to talk about their problems, by a virtual counselor in an open-ended dialogue [20], the elicitation of self-disclosure as well as the provision of relevant information by a Web-based virtual counselor among active soldiers, war veterans, and their families [48], and *diagnosis* by a virtual therapist guiding users through a Web-based version of the Beck Depression Inventory questionnaire [49].

Self-management skills were targeted by an application that supported hospitalized patients during their *discharge procedure* [50], and by a serious game in which people with depression could practice *communicating about their health* with a virtual doctor [51].

The last study concerned the same job interview training application used in [23], this time targeting people with other psychiatric disabilities, including depression [52].

ECA Technology

Looking at the use of human communication modalities, the most technologically advanced developments take place in counseling interventions from studies conducted by the Institute for Creative Technologies, associated with the University of Southern California (USC-ICT). Over the years, they have developed an extensive framework that allows users to communicate with ECAs in a natural manner through verbal and nonverbal behavior. In one study [48], users could communicate by textual natural language. Using speech and synchronized nonverbal behaviors, the ECA was able to take initiative in the conversation and probe for information related to depression and PTSD. An even more advanced approach in terms of user input was taken in another study [20], in which users' speech and nonverbal behavior was taken as input, and used to engage them in open-ended dialogues aimed at self-disclosure about psychological problems. The ECA was endowed with a set of fixed utterances and interview questions, applied back-channel behaviors (eg, saying "uhuh" and nodding while listening) and empathic responses to build rapport with the user, and used continuation prompts (eg, a new question) to keep the conversation going.

Another, technologically less advanced, approach that emphasized longer-term user modeling was taken by the Relational Agents Group of Northeastern University. During the past decade, this research group has developed a framework for so-called relational agents that apply a variety of techniques (eg, daily small-talk, empathic displays, referencing to previous interactions) to develop a long-lasting relationship with the user through menu-based dialogs over multiple interactions [50].

A last set of studies focused on modeling users' emotional state in real time based on their interaction with the application itself, for example, based on their answers to depression questionnaires [45,49].

Evidence

The outcome studied the most (n=8) was user satisfaction. Although the studies focusing on more advanced technologies such as natural interaction and empathy modeling were still in the development (n=5) and piloting (n=2) phase, n=3 studies moved beyond that.

The sole study around an implementation question [19] concerned a focus group study (N=16) of the acceptability of SPARX, a gamified CBT intervention developed in New Zealand, in which players can, for example, use a staff to shoot physically manifested negative thoughts. ECAs as such are not a predominant theme in the SPARX game, but regular mention of a guide character is made. Players choose their own avatar that provides instructions throughout the game in dialogs with the user. Australian participants indicated that it was important that the guide's gender could be customized, did not mind its

XSL•FO
RenderX

foreign accent, and liked the idea of being able to socialize with it. Being a focus group study, it remains unclear whether these results would hold in an experimental setting.

The first evaluation study (N=134) used the photograph of a clinician as an embodiment to deliver automated motivational support in a computerized acceptance and commitment therapy [46]. Users could not interact with the ECA directly, and personalized support occurred on a predefined schedule through a user model based on the user's actions in the intervention. Participants receiving automated ECA feedback were found not to be significantly less involved than those receiving real human support. Although this result was very interesting in the sense that ECA support was compared with real human support, the ECA itself made little use of state-of-the-art ECA technology, and therefore gives us little to go on in terms of ECA design. The second evaluation study concerned another randomized controlled study (N=37) using the job interview training application [52] that was also used for ASDs [23]. Again, the users' researcher-scored interview skills in a role-played interview, as well as users' self-confidence, improved significantly compared with a control group that received no intervention.

The development and piloting studies provided us with some initial evidence that practicing health communication with virtual health care providers in a serious game can be efficacious [51], that ECAs in a CBT-based intervention should have a coaching role, be configurable, adaptable, trustworthy, guiding rather than directive, and capable of empathic expressions without reflecting negative ones back to the user [47], that ECAs endowed with empathy are more highly valued than those without it [45,49], that people do not experience less rapport when interacting with an ECA than when interacting with a human [20], that people appreciate the anonymous nature of interacting with an ECA [48], and that people with depression experience a stronger working alliance with a virtual nurse guiding a hospital check-out procedure than do the nondepressed [50].

Anxiety Disorders

There were N=5 studies that targeted anxiety disorders, either with CBT (n=2) or counseling (n=3) interventions. ECAs assumed the role of a social interaction partner (n=3), a health care provider (n=2), and a tutor (n=1), and most of them were implemented in stand-alone software (n=3). Reasons to use ECAs for anxiety disorders were similar to those mentioned for depression.

Psychotherapeutic Interventions

The counseling studies experimented with various techniques to elicit self-disclosure in counseling sessions with a virtual agent. While we already discussed one study [20] for depression, two other studies focused solely on eliciting personal information from people with anxiety in the context of finding a new roommate [53] and counseling [54]. In the CBT-based applications, virtual animals helped children to conquer performance anxiety in a serious game [55], and a virtual character evoked anxiety in a VR environment [56].

ECA Technology

Most innovative here are the counseling studies, again conducted by USC-ICT. Although all three ECAs are based on the same framework, those described in [53] and [54] differ from [20] in that a so-called "Wizard of Oz" paradigm was applied to control the ECAs' verbal behavior, that is, it was controlled by the researchers. Whereas this violated our definition of agency in terms of verbal behavior, the so-called rapport agents' nonverbal behavior was completely automated. Interpreting the phonetic aspects of a user's speech input, as well as video recordings of his or her nonverbal behavior, they were able to display appropriate nonverbal behaviors themselves.

Evidence

The applications we considered were still in the development (n=2) and piloting (n=3) phase, and there was no predominant outcome measure used. Although some studies worked with large sample sizes (N=351 in [20], and N=108 in [53]), none of the studies experimented with clinical samples.

In these studies, most relevant to our purpose were the findings that people with elevated levels of social anxiety may find it easier to disclose personal information to an ECA than to a human [53], that human backstories may be more effective in this respect than (true) computer backstories [54], and that highly anxious people approached a character in a VR environment more slowly, and kept more distance, than less-anxious ones [56].

Post-Traumatic Stress Disorder

A total of n=4 studies targeted PTSD. Besides two studies on counseling interventions that also targeted depression [20,48], one study proposed a virtual coach in a CBT-based Web-based platform [57], and one concerned a virtual guide in a serious gaming healing environment [58]. The studies involved ECAs in the role of a coach (n=3) and a health care provider (n=1).

Psychotherapeutic Interventions

In the first study [58], a fantasy character acted as an engaging information repository in a virtual healing environment for returning soldiers that stimulated social comradery, healing activities, and personal exploration. The other study involved a focus group of experts in trauma treatment, in which design requirements were gathered for a virtual agent supporting a Web-based exposure therapy-based application [57].

ECA Technology

From a technological perspective, the most interesting developments took place in the two studies we already discussed. The guide in [58] was implemented as a virtual character in a private space built in the Second Life virtual worlds platform, but the details about its design remained unclear.

Evidence

All studies (n=4) were still in the development phase. Even though [58] had an impressive sample size (N=700), the focus was still on usability of the healing environment itself. Some examples of suggested guidelines for the virtual coach from the focus group study (N=10) [57] were that it should acknowledge

XSL•FO
RenderX

patients' feelings, remind them of their goals when they indicate they wish to quit, be factual in complimenting, and never express negative emotion.

Psychotic Disorders

The n=4 studies involving psychotic disorders revolved around social skills training (n=2) and self-management (n=2). Aside from the Web-based job interview training application that also targeted depression [52], two applications were implemented in stand-alone software, and one in a VR environment. Interestingly, this set of studies was the only one to consider ECAs in all four social roles. Important reasons to explore the use of ECAs in the treatment of psychotic disorders were that social skills could be practiced in a safe environment, and that ECAs can always be available to provide support or information.

Psychotherapeutic Interventions

Two studies applied the Relational Agent framework described in the section on depression [50], and used an ECA to host a system that provided general lifestyle support with an emphasis on promoting medication adherence for people suffering from schizophrenia over a 1-month period [18,59]. In the other study, people with schizophrenia could practice *conversational skills* with virtual characters in a VR social situation [60].

ECA Technology

The self-management interventions that were based on the Relational Agent framework used a similar set of techniques as in [50] to develop a long-term relationship between a virtual psychiatric nurse and the user. The conversations in [60] followed a branching tree model approach, which allowed users to communicate through a multiple choice menu. A virtual coach could help users in case they ended up in negative situations. All of the studies allowed users to interact with ECAs through menu-based dialogs.

Evidence

Although the studies that have not already been discussed under the various disorders were all in the piloting phase (n=3), and sample sizes were fairly small (mean 19.8 [SD 11.9]), they all studied clinical populations. Whereas usability (n=3) and satisfaction (n=4) outcomes were studied most often, with positive results, there is some initial evidence that the Relational Agent applications helped people with schizophrenia to adhere to their medication intake [18,59], and that VR social situations evoke similar negative symptoms in people with schizophrenia to what would be expected from real-world situations [60].

Substance Use

All ECA applications (n=4) targeting substance use were Web based and included CBT elements. ECAs assumed the role of a coach (n=2) and a health care provider (n=2). The main reason to use ECAs in the context of substance use is that they are more available than supportive humans would be, thereby increasing accessibility.

Psychotherapeutic Interventions

In [61], a "makeover host" was used to deliver and highlight personalized content in an intervention called REALU2, targeting healthy lifestyle behavior with an emphasis on smoking cessation. Smoking cessation was also the topic of [62], which investigated the acceptability of a proposed virtual agent for an intervention based on motivational interviewing. Motivational interviewing was used in a brief intervention targeting problematic drinking behavior in [63,64].

ECA Technology

The makeover host in [61] delivered personalized messages based on how users interacted with the intervention, but the details of its design remained unclear. The motivational interview intervention described in [63] allowed users to interact with a virtual counselor through a menu-based system, and their facial expressions were recorded to deduce their emotional state. The combination of user input and emotional state allowed the ECA to guide the conversation and respond empathically. In [64], efforts were made to make interaction with the same system more natural by using speech rather than menu-based input.

Evidence

This set of studies contained one study in the evaluation phase, which had the highest number of participants (N=1317 adults with a history of smoking) out of all the studies considered. In a randomized controlled trial, the ECA intervention was found to be more effective in reducing self-reported smoking than a control condition in which participants used an intervention based on general lifestyle support. Including peer support further boosted the ECA intervention's effectiveness. Because the ECA's design was not described in detail, and because the intervention with the ECA was not compared with one without it, it remains unclear how the ECA itself contributed to the results. Some evidence for the importance of empathic behavior by ECAs was provided in a randomized controlled trial on the brief motivational intervention [63]. The ECA using an empathy module performed significantly better than an ECA without it on various outcome measures, but long-term effects regarding substance use remained unclear.

Discussion

Principal Findings

This review aimed to inform health professionals about the technological possibilities and evidence base for ECA applications in clinical psychology, and to provide an overview of the activity in this field of research. Research on the use of ECAs in psychotherapy is emerging (Figure 3, graph a), and we reviewed N=49 studies of which the majority targeted ASDs (Figure 3, graph b). A general distinction could be made between applications in which ECAs were used as an adjunct to an intervention that could also have been used independently, and applications in which the interaction between the ECA and user was central. The former were mostly CBT-based programs, educational aids, and self-management interventions, whereas the latter were mostly social interaction skills training interventions and counseling interventions. Social skills training interventions were by far the most popular for ASDs, which also made them the predominant type of intervention overall (Figure 3, graph c). As a result, ECAs in the role of a social interaction partner were the most frequent (Figure 3, graph e).

XSL•FO
RenderX

The large variety in ECA applications and types of interventions (eg, Figure 3, graph d) made it a nontrivial task to provide a taxonomy of interventions and ECA social roles. Although clinical behavioral outcomes were studied most often (Figure 3, graph g), they were in many cases restricted to pre-post measurements within experiments that had relatively small sample sizes. Consequently, few studies exceeded the piloting phase (Figure 3, graph h).

ECA Technology

The balance in terms of the ability to communicate through human communication modalities such as gestures, expressions, and speech highly favors ECAs compared with human users (Figure 3, graph f). Nevertheless, work to shift this balance has been conducted in research on social interaction training and counseling interventions. The latter made a lot of use of technologically more advanced user modeling (eg, real-time emotional states), and innovative ways for humans to communicate with computer systems, for example, by interpreting natural language, phonetic aspects from speech input, and recorded nonverbal behavior. While ASD and counseling interventions have a more short-term focus, there has also been some activity in establishing longer-term relationships by using less technical user models, the input of which came from dialog-tree systems and indirect communication with ECAs through the intervention interface itself. Given the development phases of the studies in which these technologies were applied, the more high-tech solutions seem to be most suited for experimental research into the client-therapist relationship or screening for disorders such as depression or PTSD (eg, [20] or [48]). At present, the more low-tech approaches have been evaluated most thoroughly and, therefore, seem most promising for direct application in routine clinical practice.

Evidence Base

From the evaluation phase studies we learned that there is reasonable evidence that (1) ECAs can have a positive effect on user engagement and involvement, (2) they can be effective in this sense as an adjunct to already existing interventions for mental disorders, and (3) it is important for them to convey empathy when interacting with users. An important limitation of the evaluation studies we considered is the nature of the control groups. Besides the study by Kelders [46], none compared an ECA intervention to a conventional treatment used in routine practice, or compared the intervention with the ECA to the intervention without it. Additionally, the more rigorous studies either put little emphasis on the role and design of the ECAs, or targeted disorders indirectly (eg, job interviews skills). Although most of the developmental and piloting studies show promising results with respect to usability and user acceptance, we still have little hard evidence that the proposed applications are reasonable alternatives to established treatments, or that ECAs used as an adjunct to existing interventions, bridging the gap between guided and unguided interventions, make them more clinically effective.

Future Work

In Table 1, some notable blank spots can be identified. Most obvious is the scarcity of evaluation and implementation phase studies, which requires more research with larger sample sizes, suitable control groups, and clinical populations. In this respect, the emerging nature of the field is a reassuring consideration. Two other notable blank spots are Web-based ECA applications for ASDs, and the use of ECA technology in VR applications in general. Taking into account the individual studies' descriptions, there is also still quite some research to be done with regard to effective ECA configurations, by comparing different parameter settings in controlled studies.

Web-Based CBT

To revert to delivering support in Internet interventions, we know that not all people require the same amount of support, considering, for example, that people with low intrinsic motivation benefit more from human support than those who are already motivated or prefer to work on their own [5]. User models on, for example, patient motivation could contribute to an accurate timing of ECA support, for example, by providing support when motivation is low and not disturbing them when motivation is already high. Considering the working mechanisms through which human support may increase the effectiveness of Internet interventions, little work has been conducted in the area of treatment adherence, making this a pertinent target around which to focus our efforts with respect to increasing motivation.

Another important point related to extending the evidence base is illustrated by the Help4Mood project [45] and the study by Kelders [46]. The Help4Mood project attempts to integrate a Web-based CBT-based treatment with a technologically advanced ECA that communicates through speech, and that is endowed with a dynamic emotion model used to convey empathy in real time. In the study by Kelders [46], automated textual feedback in an already existing intervention was embodied by accompanying it with the photograph of a clinician. Help4Mood is technologically more challenging, but requires a long period of development and piloting. Because the ECA plays such a central role in the intervention, it is far from straightforward to simply "add" the ECA as an adjunct to an already existing and well-evaluated intervention to study its effects. Rather, the intervention would have to be built around the ECA framework, and the resulting new intervention would once again have to proceed through the development phases. This is a general issue whenever we consider adopting one of the other ECA frameworks we came across, such as the Relational Agent Group's Litebody [65], and USC-ICT Virtual Human Toolkit [66], especially if we consider that the input used in Help4Mood is still relatively straightforward to interpret compared with human speech or nonverbal behavior.

Low-Tech Approach

If we want to investigate how to improve interventions that have already been set out in the field, a "low-tech" approach similar to [46] can be advocated because it (1) saves development time that can be used to design and set up larger studies, (2) forces us to think about the core attributes that can make the ECA

XSL•FO
RenderX

effective, and (3) makes it easier to judge whether it is safe to use the ECA in a clinical setting with real patients. Given the sparse evidence on the clinical effectiveness of ECAs thus far, this approach and its three advantages may be just what we need to study how we can effectively use ECAs in existing Internet interventions: it will be easier to (1) conduct studies that move beyond the piloting phase, (2) identify the core attributes that make ECAs effective in Internet interventions such that ECA design can be more focused and less time-consuming, and (3) conduct experiments with clinical populations such that we can study ECAs' effects on clinical outcomes.

Limitations

Our definition of ECAs had three components. With respect to the embodiment and interaction capabilities, we took a liberal stance, but our requirement of agency was rather conservative (autonomous behavior and reasoning). This excluded a fair number of studies (29 during the screening of full articles) from our review that some might consider to be relevant. Regarding the criterion of autonomous behavior, we excluded quite some studies using what is often called a "Wizard of Oz" paradigm, in which the ECA's behavior is not controlled by a software entity, but by a human operator instead. Examples are a study in which an ECA representing the hallucinated voices of people with schizophrenia spoke the transformed utterances of a therapist [67], and one in which a robot aimed at improving the mood of hospitalized children suffering from cancer was under the control of a researcher [68]. An example of a study that was excluded because the embodied characters lacked reasoning capabilities, that is, the ECA would act the same regardless of user input, was [69].

Although ECA research is almost inherently interdisciplinary, we refrained from going too deep into the technological aspects. This was because our target audience consisted of health professionals with a generally less technical background and we wanted to focus on opening up the ECA domain for them as well as providing them with an overview of the available evidence for application in routine clinical practice. For this reason, we refrained from a highly technical discussion of, for example, verbal and nonverbal ECA capabilities. However, it has to be noted that, depending on how one would like to use ECAs in future work, many more detailed questions could be investigated surrounding ECA design aspects, such as the required capabilities for, and their impact on, specific disorders or types of ECA interventions. With respect to our search strategy, we looked specifically for articles that mention ECAs. As exemplified by the sole included article on SPARX, which is actually supported by more research than reviewed here (eg, [70]), there is a possibility that we missed out on articles describing, for example, serious games or VR environments in which ECAs are used, but not specifically mentioned.

Another limitation relates to the bibliographic databases we considered. Computer science research publications are more dispersed than those of psychology research and computer science databases are less suited to systematic searches. Although our interdisciplinary approach was already broader than what is usual in psychology research, there is a possibility that we might have missed relevant research in, for example, the IEEE (Institute of Electrical and Electronics Engineers) Xplore digital library or Google Scholar. We refrained from using IEEE Xplore digital library due to practical constraints and Google Scholar because its search algorithm is often updated and personalized, which makes it difficult to replicate search results. Additionally, due to practical constraints, we did not conduct searches in the gray literature or manual searches through cross-referencing, nor did we conduct a follow-up search after the original one.

While the idea of applying ECAs in psychotherapy is far from new (eg, [71,72]), to our knowledge this is the first review to specifically consider psychotherapeutic applications of ECAs in a systematic manner. There are several areas of research that are closely related to ours. One of these is social robotics research, which often focuses on providing company to elderly people (eg, [73]) or people suffering from dementia (eg, [74]). While this area of research could be seen as an attempt to prevent the psychological consequences that might ensue from loneliness, we did not consider the focus to be on psychopathology. Robotic applications targeting autism have previously been reviewed (eg, [75]), but without the constraints implied by our ECA concept. Consequently, many of these robotic applications do not adhere to our criterion of agency, that is, they do not act autonomously or intelligently. Besides research on robotics, there is a large corpus of literature on the application of virtual agents in other highly relevant domains from which we can draw inspiration. Although this review focuses solely on psychotherapeutic applications, there seems to be little reason not to consider, for example, motivational [9], pedagogical [76], or lifestyle-support agents [77].

Conclusions

Research into the psychotherapeutic application of ECAs is emerging. We identified 49 studies, with over half of them focusing on autism. The field is characterized by a large variety in all its aspects, for example, type of intervention, target behavior, platform, ECA embodiment, communication modalities, ECA "mental" states, and study design. While there are several studies surpassing the development and piloting phases, as might be expected in a relatively new field, evidence about the clinical effectiveness of ECA applications remains sparse. Technologically advanced ECA applications are very interesting and show promising results, but their complex nature makes it difficult for now to prove that they are effective and safe to use in clinical practice. Therefore, at present, clinical practice seems well served by an additional focus on a more low-tech approach based on the elementary principles that make ECAs effective, that can progress through the development and piloting phases at a faster pace, and that can therefore more easily be proven to be safe and effective for routine clinical practice.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strings.

[PDF File (Adobe PDF File), 11KB - jmir_v19i5e151_app1.pdf]

Multimedia Appendix 2

Concept definitions.

[PDF File (Adobe PDF File), 88KB - jmir_v19i5e151_app2.pdf]

Multimedia Appendix 3

Overview of the intervention aims and embodied conversational agent (ECA) characteristics.

[PDF File (Adobe PDF File), 97KB - jmir_v19i5e151_app3.pdf]

Multimedia Appendix 4

Overview of study design characteristics by disorder.

[PDF File (Adobe PDF File), 163KB - jmir_v19i5e151_app4.pdf]

References

1. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
2. 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]
3. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
4. 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] [Medline: 22853793]
5. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
6. Isbister K, Doyle P. The blind men and the elephant revisited evaluating interdisciplinary ECA research. In: Ruttkay Z, Pelachaud C, editors. From brows to trust evaluating embodied conversational agents. Dordrecht, Netherlands: Springer; 2004:3-26.
7. Ruttkay Z, Dormann C, Noot H. Embodied conversational agents on a common ground. In: Ruttkay Z, Pelachaud C, editors. From brows to trust evaluating embodied conversational agents. Dordrecht, Netherlands: Springer; 2004:27-66.
8. Warmerdam L, Riper H, Klein M, van den Ven P, Rocha A, Ricardo HM, et al. Innovative ICT solutions to improve treatment outcomes for depression: the ICT4Depression project. Stud Health Technol Inform 2012;181:339-343. [Medline: 22954884]
9. Baylor AL. The design of motivational agents and avatars. Education Tech Research Dev 2011 Feb 27;59(2):291-300. [doi: 10.1007/s11423-011-9196-3]
10. van Vugt HC, Hoorn JF, Konijn EA. Interactive engagement with embodied agents: an empirically validated framework. Comp Anim Virtual Worlds 2009 Jun;20(2-3):195-204. [doi: 10.1002/cav.312]
11. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
13. Mays N, Roberts E. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, editors. Studying the organisation and delivery of health services: research methods. London: Routledge; 2001.
14. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

XSL•FO
RenderX

15. Daudt HM, Van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
16. Davis K, Drey N, Gould D. What are scoping studies? a review of the nursing literature. Int J Nurs Stud 2009 Oct;46(10):1386-1400. [doi: 10.1016/j.ijnurstu.2009.02.010] [Medline: 19328488]
17. Agarwal R, Sampath H, Indurkhya B. A usability study on natural interaction devices with ASD children. In: Proc 7th Int Conf Univers Access Human-Computer Interact User Context Divers - Vol 2. Heidelberg, Berlin, Germany: Springer; 2013 Presented at: 7th Int Conf Univers Access Human-Computer Interact User Context Divers; July 21-26, 2013; Las Vegas, NV p. 21-26. [doi: 10.1007/978-3-642-39191-0_49]
18. Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: Relational agents for antipsychotic medication adherence. Interacting with Computers 2010 Jul;22(4):276-288. [doi: 10.1016/j.intcom.2010.02.001]
19. Cheek C, Bridgman H, Fleming T, Cummings E, Ellis L, Lucassen MF, et al. Views of young people in rural Australia on sparx, a fantasy world developed for New Zealand youth with depression. JMIR Serious Games 2014;2(1):e3 [FREE Full text] [doi: 10.2196/games.3183] [Medline: 25659116]
20. Devault D, Artstein R, Benn G, Dey T, Fast E, Gainer A, et al. SimSensei Kiosk: a virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 Int Conf Auton Agents Multi-Agent.: International Foundation for Autonomous Agents and Multiagent Systems; 2014 Presented at: 2014 Int Conf Auton Agents Multi-Agent; May 5-9, 2014; Paris, France p. 1061-1068.
21. Wainer J, Dautenhahn K, Robins B, Amirabdollahian F. A pilot study with a novel setup for collaborative play of the humanoid robot KASPAR with children with autism. Int J of Soc Robotics 2013 Sep 11;6(1):45-65. [doi: 10.1007/s12369-013-0195-x]
22. Hopkins IM, Gower MW, Perez TA, Smith DS, Amthor FR, Wimsatt FC, et al. Avatar assistant: improving social skills in students with an ASD through a computer-based intervention. J Autism Dev Disord 2011 Nov;41(11):1543-1555. [doi: 10.1007/s10803-011-1179-z] [Medline: 21287255]
23. Smith MJ, Ginger EJ, Wright K, Wright MA, Taylor JL, Humm LB, et al. Virtual reality job interview training in adults with autism spectrum disorder. J Autism Dev Disord 2014 Oct;44(10):2450-2463 [FREE Full text] [doi: 10.1007/s10803-014-2113-y] [Medline: 24803366]
24. Bernardini S, Porayska-Pomsta K, Smith TJ. ECHOES: an intelligent serious game for fostering social communication in children with autism. Information Sciences 2014 Apr;264:41-60. [doi: 10.1016/j.ins.2013.10.027]
25. Alcorn A, Pain H, Rajendran G, Smith T, Lemon O, Porayska-Pomsta K, et al. Social communication between virtual characterschildren with autism. In: Biswas G, Bull S, Kay J, Mitrovic A, editors. Artificial Intelligence in Education. Heidelberg, Berlin, Germany: Springer; 2011:7-14.
26. Chen J, Chen D, Wang L, Lemon O. An advanced learning environment aided by recognition of multi-modal social signals. In: Luo X, Spaniol M, Wang L, Li Q, Nejdl W, Zhang W, editors. Proceedings of the 9th Int Conf Adv Web-Based Learn - ICWL 2010. Heidelberg, Berlin, Germany: Springer; 2010:41-51.
27. Bekele E, Crittendon JA, Swanson A, Sarkar N, Warren ZE. Pilot clinical application of an adaptive robotic system for young children with autism. Autism 2014 Jul;18(5):598-608 [FREE Full text] [doi: 10.1177/1362361313479454] [Medline: 24104517]
28. Fujimoto I, Matsumoto T, De Silva SP, Kobayashi M, Higashi M. Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot. Int J of Soc Robotics 2011 Nov 17;3(4):349-357. [doi: 10.1007/s12369-011-0116-9]
29. Warren Z, Zheng Z, Das S, Young EM, Swanson A, Weitlauf A, et al. Brief report: development of a robotic intervention platform for young children with ASD. J Autism Dev Disord 2015 Dec;45(12):3870-3876. [doi: 10.1007/s10803-014-2334-0] [Medline: 25503680]
30. Boccanfuso L, O'Kane J. Adaptive robot design with hand and face tracking for use in autism therapy. In: Ge SS, Li H, Cabibihan JJ, Tan YK, editors. ICSR 2010: Social Robotics. Heidelberg, Berlin, Germany: Springer; 2010:265-274.
31. Robins B, Dautenhahn K. Tactile interactions with a humanoid robot: novel play scenario implementations with children with autism. Int J of Soc Robotics 2014 Apr 4;6(3):397-415. [doi: 10.1007/s12369-014-0228-0]
32. Amirabdollahian F, Robins B, Dautenhahn K, Ji Z. Investigating tactile event recognition in child-robot interaction for use in autism therapy. Conf Proc IEEE Eng Med Biol Soc 2011:5347-5351. [doi: 10.1109/IEMBS.2011.6091323] [Medline: 22255546]
33. Dickerson P, Robins B, Dautenhahn K. Where the action is: a conversation analytic perspective on interaction between a humanoid robot, a co-present adult and a child with an ASD. Interaction Studies 2013 Jul 22;14(2):297-316. [doi: 10.1075/is.14.2.07dic]
34. Costa S, Lehmann H, Dautenhahn K, Robins B, Soares F. Using a humanoid robot to elicit body awareness and appropriate physical interaction in children with autism. Int J of Soc Robotics 2014 Aug 24;7(2):265-278. [doi: 10.1007/s12369-014-0250-2]

XSL•FO
RenderX

35. Tanaka H, Sakti S, Neubig G, Toda T, Negoro H, Iwasaka H. Automated social skills trainer. In: Proc 20th International Conference on Intelligent User Interfaces. New York, NY: ACM Press; 2015 Presented at: 20th International Conference on Intelligent User Interfaces; March 29 - April 1, 2015; Atlanta, GA p. 17-27. [doi: 10.1145/2678025.2701368]
36. Lahiri U, Bekele E, Dohrmann E, Warren Z, Sarkar N. Design of a virtual reality based adaptive response technology for children with autism spectrum disorder. In: D'Mello S, Graesser A, Schuller B, Martin JC, editors. ACII 2011: Affective Computing and Intelligent Interaction. Heidelberg, Berlin, Germany: Springer; 2011:165-174.
37. Milne M, Powers D, Leibbrandt R. Development of a software-based social tutor for children with autism spectrum disorders. In: Proc 21st Annu Conf Aust Comput Interact Spec Interes Gr Des Open 24/7. 2009 Presented at: OZCHI '09; November 23-27, 2009; Melbourne, Australia p. 265-268. [doi: 10.1145/1738826.1738870]
38. Ribeiro P, Araujo BD, Raposo A. ComFiM: a cooperative serious game to encourage the development of communicative skills between children with autism. In: Proc 2014 Brazilian Symp Comput Games Digit Entertain.: IEEE Computer Society; 2014 Presented at: Computer Games and Digital Entertainment (SBGAMES); November 12-14, 2014; Porto Alegro, Brazil p. 148-157. [doi: 10.1109/SBGAMES.2014.19]
39. Jordan K, King M, Hellersteth S, Wirén A, Mulligan H. Feasibility of using a humanoid robot for enhancing attention and social skills in adolescents with autism spectrum disorder. Int J Rehabil Res 2013 Sep;36(3):221-227. [doi: 10.1097/MRR.0b013e32835d0b43] [Medline: 23328816]
40. Cole R, Van Vuuren S, Pellom B, Hacioglu K, Jiyong M, Movellan J, et al. Perceptive animated interfaces: first steps toward a new paradigm for human-computer interaction. Proceedings of the IEEE 2003 Sep 8;91(9):1391-1405. [doi: 10.1109/JPROC.2003.817143]
41. Bamasak O, Braik R, Al-Tayari H, Al-Harbi S, Al-Semairi G, Abu-Hnaidi M. Improving autistic children?s social skills using virtual reality. In: Marcus A, editor. Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience. DUXU 2013. Lecture Notes in Computer Science, vol 8013. Heidelberg, Berlin, Germany: Springer; 2013:342-351.
42. Konstantinidis E, Hitoglou-Antoniadou M, Luneski A, Bamidis P, Nikolaidou M. Using affective avatars and rich multimedia content for education of children with autism. In: Proc 2nd Int Conf Pervasive Technol Relat to Assist Environ. New York, USA: ACM New York; 2009 Presented at: PETRA '09; 2009 June 9-13; Corfu, Greece p. 1-6. [doi: 10.1145/1579114.1579172]
43. Shoukry L, Caserman P, Göbel S, Steinmetz R. Blitzmerker: learning idioms with a mobile game. In: Göbel S, Ma M, Baalsrud Hauge J, Oliveira M, Wiemeyer J, Wendel V, editors. Serious Games. JCSG 2015. Lecture Notes in Computer Science, vol 9090. Cham, Switzerland: Springer; 2015:77-88.
44. Palestra G, Bortone I, Cazzato D, Adamo F, Argentiero A, Agnello N. Social robots in postural education: a new approach to address body consciousness in ASD children. In: Beetz M, Johnston B, Williams MA, editors. Social Robotics. ICSR 2014. Lecture Notes in Computer Science, vol 8755. Cham, Switzerland: Springer; 2014:290-299.
45. Martínez-Miranda J, Bresó A, García-Gómez J. Look on the bright side: a model of cognitive change in virtual agents. In: Bickmore T, Marsella S, Sidner C, editors. Intelligent Virtual Agents. IVA 2014. Lecture Notes in Computer Science, vol 8637. Cham, Switzerland: Springer; 2014:285-294.
46. Kelders S. Involvement as a working mechanism for persuasive technology. In: MacTavish T, Basapur S, editors. PERSUASIVE 2015. Lecture Notes in Computer Science, vol 9072. Cham, Switzerland: Springer; 2015:3-14.
47. Pagliari C, Burton C, McKinstry B, Szentatotai A, David D, Serrano Blanco A, et al. Psychosocial implications of avatar use in supporting therapy for depression. Stud Health Technol Inform 2012;181:329-333. [Medline: 22954882]
48. Swartout W, Artstein R, Forbell E, Foutz S, Lane H, Lange B, et al. Virtual humans for learning. AI Mag 2013;34:13-30.
49. Pontier M, Siddiqui G. A virtual therapist that responds empathically to your answers. In: Lester H, Ishizuka M, editors. Intelligent Virtual Agents. IVA 2008. Lecture Notes in Computer Science, vol 5208. Springer: Proc of the 8th Int Conf on Intell Virtual Agents; 2008:417-425.
50. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
51. Pinto MD, Greenblatt AM, Hickman RL, Rice HM, Thomas TL, Clochesy JM. Assessing the critical parameters of eSMART-MH: a promising avatar-based digital therapeutic intervention to reduce depressive symptoms. Perspect Psychiatr Care 2015 Mar 19;52(3):157-168. [doi: 10.1111/ppc.12112] [Medline: 25800698]
52. Smith MJ, Ginger EJ, Wright M, Wright K, Boteler Humm L, Olsen D, et al. Virtual reality job interview training for individuals with psychiatric disabilities. J Nerv Ment Dis 2014 Sep;202(9):659-667 [FREE Full text] [doi: 10.1097/NMD.0000000000000187] [Medline: 25099298]
53. Kang S, Gratch J. Virtual humans elicit socially anxious interactants' verbal self-disclosure. Comp Anim Virtual Worlds 2010;21(3-4):473-482. [doi: 10.1002/cav.345]
54. Kang S, Gratch J. Socially anxious people reveal more personal information with virtual counselors that talk about themselves using intimate human back stories. Stud Health Technol Inform 2012;181:202-206. [Medline: 22954856]
55. Schmidt R, Eifler P, Masuch M. Playfully conquering performance anxiety. In: Schouten B, Fedtke S, Bekker T, Schijven M, Gekker A, editors. Games for Health. Wiesbaden: Springer Fachmedien; 2013:267-279.

XSL•FO
RenderX

56. Rinck M, Rörtgen T, Lange W, Dotsch R, Wigboldus DH, Becker ES. Social anxiety predicts avoidance behaviour in virtual encounters. Cogn Emot 2010 Nov;24(7):1269-1276. [doi: 10.1080/02699930903309268]
57. Tielman M, Brinkman W, Neerincx M. Design guidelines for a virtual coach for post-traumatic stress disorder patients. In: Bickmore T, Marsella S, Sidner C, editors. Intelligent Virtual Agents. IVA 2014. Lecture Notes in Computer Science, vol 8637. Cham, Switzerland: Springer; 2014:434-437.
58. Morie JF, Antonisse J, Bouchard S, Chance E. Virtual worlds as a healing modality for returning soldiers and veterans. Stud Health Technol Inform 2009;144:273-276. [Medline: 19592780]
59. Puskar K, Schlenk EA, Callan J, Bickmore T, Sereika S. Relational agents as an adjunct in schizophrenia treatment. J Psychosoc Nurs Ment Health Serv 2011 Aug;49(8):22-29. [doi: 10.3928/02793695-20110705-01] [Medline: 21766736]
60. Ku J, Han K, Lee HR, Jang HJ, Kim KU, Park SH, et al. VR-based conversation training program for patients with schizophrenia: a preliminary clinical trial. Cyberpsychol Behav 2007 Aug;10(4):567-574. [doi: 10.1089/cpb.2007.9989] [Medline: 17711366]
61. An LC, Demers MR, Kirch MA, Considine-Dunn S, Nair V, Dasgupta K, et al. A randomized trial of an avatar-hosted multiple behavior change intervention for young adult smokers. J Natl Cancer Inst Monogr 2013 Dec;2013(47):209-215 [FREE Full text] [doi: 10.1093/jncimonographs/lgt021] [Medline: 24395994]
62. Grolleman J, van DB, Nijholt A, Van Ernst A. Break the habit! Designing an e-therapy intervention using a virtual coach in aid of smoking cessation. In: IJsselsteijn WA, De Kort YA, Midden C, Eggen B, Van den Hoven E, editors. Persuasive Technology. PERSUASIVE 2006. Lecture Notes in Computer Science, vol 3962. Heidelberg, Berlin, Germany: Springer; 2006:133-141.
63. Lisetti C, Amini R, Yasavur U, Rishe N. I can help you change! an empathic virtual agent delivers behavior change health interventions. ACM Transactions on Management Information Systems (TMIS) - Special Issue on Informatics for Smart Health and Wellbeing 2013;4(4):1-28. [doi: 10.1145/2544103]
64. Yasavur U, Lisetti C, Rishe N. Let's talk! speaking virtual counselor offers you a brief intervention. J Multimodal User Interfaces 2014 Sep 5;8(4):381-398. [doi: 10.1007/s12193-014-0169-9]
65. Bickmore T, Schulman D, Shaw G. DTaskLitebody: open source, standards-based tools for building web-deployed embodied conversational agents. In: Ruttkay Z, Kipp M, Nijholt A, Vilhjálmsson HH, editors. Intelligent Virtual Agents. IVA 2009. Lecture Notes in Computer Science, vol 5773. Heidelberg, Berlin, Germany: Springer; 2009:425-431.
66. Hartholt A, Traum D, Marsella S, Shapiro A, Stratou G, Leuski A. All together now. In: Aylett R, Krenn B, Pelachaud C, Shimodaira H, editors. Intelligent Virtual Agents. IVA 2013. Lecture Notes in Computer Science, vol 8108. Heidelberg, Berlin, Germany: Springer; 2013:368-381.
67. Leff J, Williams G, Huckvale MA, Arbuthnot M, Leff AP. Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. Br J Psychiatry 2013 Jun;202:428-433 [FREE Full text] [doi: 10.1192/bjp.bp.112.124883] [Medline: 23429202]
68. Alemi M, Meghdari A, Ghanbarzadeh A, Moghadam L, Ghanbarzadeh A. Impact of a social humanoid robot as a therapy assistant in children cancer treatment. In: Beetz M, Johnston B, Williams MA, editors. Social Robotics. ICSR 2014. Lecture Notes in Computer Science, vol 8755. Cham, Switzerland: Springer; 2014:11-22.
69. Wallace S, Parsons S, Westbury A, White K, White K, Bailey A. Sense of presence and atypical social judgments in immersive virtual environments. Responses of adolescents with autism spectrum disorders. Autism 2010 May;14(3):199-213. [doi: 10.1177/1362361310363283] [Medline: 20484000]
70. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. 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]
71. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Educ Couns 2013 Aug;92(2):160-166 [FREE Full text] [doi: 10.1016/j.pec.2013.05.007] [Medline: 23809167]
72. Bickmore T, Gruber A. Relational agents in clinical psychiatry. Harv Rev Psychiatry 2010;18(2):119-130. [doi: 10.3109/10673221003707538] [Medline: 20235777]
73. Broekens J, Heerink M, Rosendal H. Assistive social robots in elderly care: a review. Gerontechnology 2009 Apr;8(2):94-103. [doi: 10.4017/gt.2009.08.02.002.00]
74. Mordoch E, Osterreicher A, Guse L, Roger K, Thompson G. Use of social commitment robots in the care of elderly people with dementia: a literature review. Maturitas 2013 Jan;74(1):14-20. [doi: 10.1016/j.maturitas.2012.10.015] [Medline: 23177981]
75. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord 2012 Dec;6(1):249-262 [FREE Full text] [doi: 10.1016/j.rasd.2011.05.006] [Medline: 22125579]
76. Baylor A, Kim Y. Pedagogical agent design: the impact of agent realism, gender, ethnicity, instructional role. In: Lester JC, Vicari RM, Paraguacu F, editors. Intelligent Tutoring Systems. ITS 2004. Lecture Notes in Computer Science, vol 3220. Heidelberg, Berlin, Germany: Springer; 2004:592-603.

XSL•FO
RenderX

77. Smith C, Cavazza M, Charlton D, Zhang L, Turunen M, Hakulinen J. Integrating planning and dialogue in a lifestyle agent. In: Prendinger H, Lester J, Ishizuka M, editors. Intelligent Virtual Agents. IVA 2008. Lecture Notes in Computer Science, vol 5208. Heidelberg, Berlin, Germany: Springer; 2008:146-153.

Abbreviations

ACM-DL: Association for Computing Machinery Digital Library
ASD: autism spectrum disorder
CBT: cognitive behavioral therapy
ECA: embodied conversational agent
MRC: UK Medical Research Council
PTSD: post-traumatic stress disorder
SPARX: Smart, Positive, Active, Realistic, X-factor thoughts
VR: virtual reality

Edited by G Eysenbach; submitted 29.08.16; peer-reviewed by S Scherer, M Lucassen, H Bridgman, B McKinstry; comments to author 26.11.16; revised version received 07.01.17; accepted 07.03.17; published 09.05.17.

Please cite as:
Provoost S, Lau HM, Ruwaard J, Riper H
Embodied Conversational Agents in Clinical Psychology: A Scoping Review
J Med Internet Res 2017;19(5):e151
URL: http://www.jmir.org/2017/5/e151/
doi:10.2196/jmir.6553
PMID:28487267

©Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, Heleen Riper. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

What Matters in Weight Loss? An In-Depth Analysis of Self-Monitoring

Stefanie Lynn Painter[1*], LD, RD, DHEd; Rezwan Ahmed[2*], PhD; James O Hill[3], PhD; Robert F Kushner[4], MD; Richard Lindquist[5], MD; Scott Brunning[2], MS; Amy Margulies[2], LDN, CDE, RD

[1]Retrofit INC, Chicago, IL, United States

[2]Retrofit, INC, Chicago, IL, United States

[3]University of Colorado, Aurora, CO, United States

[4]Northwestern University, Feinberg School of Medicine, Chicago, IL, United States

[5]Swedish Medical Center, Seattle, WA, United States

[*]these authors contributed equally

Corresponding Author:
Stefanie Lynn Painter, LD, RD, DHEd
Retrofit INC
123 N. Wacker Drive
Suite 1250
Chicago, IL, 60606
United States
Phone: 1 8007745962
Fax: 1 8007745962
Email: stefanie@retrofitme.com

Abstract

Background: Using technology to self-monitor body weight, dietary intake, and physical activity is a common practice used by consumers and health companies to increase awareness of current and desired behaviors in weight loss. Understanding how to best use the information gathered by these relatively new methods needs to be further explored.

Objective: The purpose of this study was to analyze the contribution of self-monitoring to weight loss in participants in a 6-month commercial weight-loss intervention administered by Retrofit and to specifically identify the significant contributors to weight loss that are associated with behavior and outcomes.

Methods: A retrospective analysis was performed using 2113 participants enrolled from 2011 to 2015 in a Retrofit weight-loss program. Participants were males and females aged 18 years or older with a starting body mass index of ≥25 kg/m2, who also provided a weight measurement at the sixth month of the program. Multiple regression analysis was performed using all measures of self-monitoring behaviors involving weight measurements, dietary intake, and physical activity to predict weight loss at 6 months. Each significant predictor was analyzed in depth to reveal the impact on outcome.

Results: Participants in the Retrofit Program lost a mean –5.58% (SE 0.12) of their baseline weight with 51.87% (1096/2113) of participants losing at least 5% of their baseline weight. Multiple regression model (R^2=.197, P<0.001) identified the following measures as significant predictors of weight loss at 6 months: number of weigh-ins per week (P<.001), number of steps per day (P=.02), highly active minutes per week (P<.001), number of food log days per week (P<.001), and the percentage of weeks with five or more food logs (P<.001). Weighing in at least three times per week, having a minimum of 60 highly active minutes per week, food logging at least three days per week, and having 64% (16.6/26) or more weeks with at least five food logs were associated with clinically significant weight loss for both male and female participants.

Conclusions: The self-monitoring behaviors of self-weigh-in, daily steps, high-intensity activity, and persistent food logging were significant predictors of weight loss during a 6-month intervention.

(J Med Internet Res 2017;19(5):e160) doi:10.2196/jmir.7457

KEYWORDS

behavior; body mass index; fitness trackers; self-monitoring; obesity; overweight; weight loss

XSL•FO
RenderX

Introduction

Self-monitoring is commonly used in weight-loss regimens to increase awareness of current and desired behaviors. Both consumers and health companies are incorporating self-monitoring technology through mobile phone apps, smart scales, and other wearable devices into their weight-loss programs. However, understanding how to best use the information being gathered by this relatively new technology needs more rigorous study, especially with recent controversy regarding the benefits of wearable activity trackers [1,2]. According to the Centers for Disease Control and Prevention (CDC), 36.5% of adults are classified as obese in the United States and US $147 billion is spent on obesity-related medical costs per year; therefore, determining whether and how self-monitoring contributes to weight loss is important for improving the health of the overall population [3].

Standard behavioral treatment in obesity includes dietary and physical activity counseling and self-monitoring of body weight, activity, and diet [4]. Behavioral weight-loss interventions up to 12 months have average outcomes between 5% to 10% weight loss [5-11]. Although clinically significant, the studies reviewed showed less than half of participants are successful at losing 5% or more of their weight [6,7,12-14].

Regular self-weighing or weighing in a consistent pattern over time provides awareness to specific behaviors, situations, or environments that could promote desired or undesired changes in weight. Self-weighing correlates with successful weight loss and has been shown to significantly increase weight loss success in the first 6 months of an intervention [15-18]. Specifically, a minimum of weekly self-weigh-ins has been shown to be effective; however, a higher frequency of self-weigh-ins more than once per week increases weight-loss outcomes [19-24]. Once a consistent pattern of self-weighing has been established, not weighing for more than a month increases likelihood of weight gain, as shown by Helander et al [15].

Both wearing an activity tracker and setting a step goal are associated with lower body mass index (BMI) and an increase in activity [25]. The average American gets 5117 steps per day [26]. High step averages were associated with younger, single males with higher education and lower BMI (kg/m^2) [26]. Individuals with obesity averaged 1500 fewer steps per day than normal or overweight individuals [26]. Modest weight loss has been shown with pedometer interventions [27,28]. By setting individualized physical activity goals around steps per day and active minutes per day, participants are more likely to increase and maintain physical activity postintervention [29]. More frequent self-monitoring and higher adherence are related to greater physical activity over time, which can lead to a greater decrease in weight at 6 months [30].

Dietary self-monitoring with feedback can improve clinically significant weight-loss outcomes [31-34], whereas personalized feedback can improve consistency of dietary self-monitoring [34,35]. Consistency has the greatest association between dietary self-monitoring and achieving clinically significant weight loss [31-34,36]. Self-monitoring for consecutive days is linked to greater outcomes, such as logging at least one food log entry per day has been shown to increase weight loss [31,32].

The purpose of this study was to analyze the self-monitoring behaviors of participants around weight, activity, and nutrition in a 6-month weight-loss intervention administered by Retrofit (see Multimedia Appendix 1), a personalized weight-management and Web-based disease-prevention solution. The self-monitoring behaviors were evaluated for their association with weight loss to determine the level of impact on predicting weight loss outcomes. Additionally, each high impact behavior was evaluated independently to assess the association between the behavior and weight loss to determine best practices around self-monitoring recommendations. The analysis of the significant self-monitoring behaviors focused on understanding the following questions:

1. What is the association between a participant's level of self-monitoring and weight loss?
2. What is the association between different levels of weight loss outcomes and the corresponding participant's commitment to self-monitoring?

Methods

Study Design

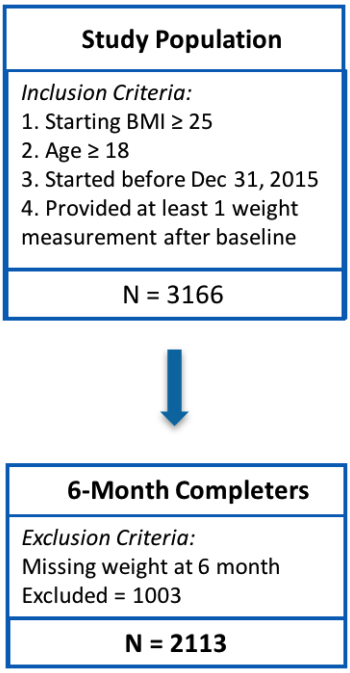
A retrospective analysis was performed to assess the effect of various self-monitoring behaviors during a 6-month weight-loss intervention using de-identified data from the Retrofit weight-loss program.

Participants

Participants in the study were paying customers of the Retrofit Program who enrolled through the direct-to-consumer website [37] or through an employer-sponsored program. Customers were considered as eligible study participants if they were age at least 18 years; had a starting BMI of 25 kg/m^2 or higher; had signed up for the program between September 27, 2011 and December 31, 2015; and provided at least one weight measurement beyond baseline measurement. Participants were considered to have completed the program if they provided a weight measurement at the sixth month of their program. A total of 3166 customers satisfied all inclusion criteria to be study participants (Figure 1). Approximately 80.35% (2544/3166) of the study participants were direct-to-consumer customers and the remaining 19.65% (622/3166) were part of an employer-sponsored program. A total of 2113 (66.74%) participants completed the 6-month program. All customers who satisfied the inclusion criteria and provided a weight at 6 months were included as participants. No customer was removed or eliminated from the population due to a lack of success on the program.

XSL•FO
RenderX

Figure 1. Study population with inclusion / exclusion criteria.



Retrofit Program

The analysis included data from six Retrofit programs: Expert 10 Weight-Loss Program, Expert 15 Weight-Loss Program, Advisor Weight-Loss Program, Jump Start Program, Retrofit Program, and Sustain Program. The Expert 10, Expert 15, Advisor, and Sustain programs were designed with a 6-month weight-loss phase and an additional 6-month weight maintenance phase. The Retrofit Program was designed with a 6-month weight-loss phase only with the option to continue into a maintenance program called Retrofit Next. The Jump Start Program was designed with a 3-month weight-loss phase only with the option to continue into Retrofit Next.

As part of the Retrofit weight-loss protocols, all participants are taught and encouraged to adhere to the same self-monitoring recommendations. All programs provided participants with the same technology, access to a weight-loss expert, education, accountability, feedback, and the opportunity to communicate with an expert coach via Web-based messages. Additional details of the Retrofit Program and expert coach qualifications can be found in a previous publication [38].

The participant was provided a Fitbit activity tracker, Wi-Fi-enabled scale, and access to a private dashboard (see Multimedia Appendix 2). The private dashboard allowed each participant to keep a personal food and exercise log, review his or her personal data, and enabled communication between the participant and his or her expert coach through a Web-based electronic messaging feature (see Multimedia Appendix 3). The private dashboard was accessed via the Retrofitme Web app, mobile website, or mobile phone app, which was available on Apple iOS and Android platforms.

Participants were encouraged to weigh in daily, wear their activity tracker daily, achieve their personalized daily step goal, and log all food and beverage choices consumed throughout the day. Expert coaches personalized participant's step goals by recommending the participants to increase step counts in increments of 500 to achieve their personal daily step goal at 6 months.

Measures

Weight

Participants were provided a Wi-Fi-enabled scale that securely transmitted weight data over the Internet to a Retrofit central data server. Participant weight data were collected through the use of the provided Wi-Fi scale (99.39%, 556,630/560,043 of recorded weights) or self-reported entry (0.61%, 3413/560,043). Instructions were provided for scale set up, as well as the option for help through Retrofit's customer support. Self-reported entry was permissible if a participant had difficulty setting up their Wi-Fi scale. Expert coaches reviewed weight data during 1:1 coaching sessions to confirm weight accuracy. Baseline weight was considered as the first weight measurement received from the participant during week 1. Percentage of baseline weight lost at 6 months was calculated and used as the primary outcome. Two weigh-in metrics were calculated to quantify participants' adherence to self-monitoring behavior and the potential impact of self-weigh-ins on weight loss: (1) the number of weigh-ins per week and (2) the percentage of weeks participants weighed in at least three times.

Activity

Participants were encouraged to wear Fitbit activity trackers every day. Activity data from any version of Fitbit activity trackers, such as steps, distance, calories burned, active minutes, etc, were wirelessly uploaded to Fitbit.com and later automatically synced to participants' personal Retrofit dashboards. Participants did not have the option to self-report activity data. A total of five different metrics were calculated to understand the impact of activity on weight loss. The number of tracker usage days per week was calculated to monitor participant engagement. Step count per day was considered one

XSL•FO
RenderX

of the metrics for measuring participant activity. To measure the intensity of the activities, three levels of active minutes were tracked. Fitbit trackers continuously estimate users' metabolic equivalents (METs) by calculating the intensity of the activity and classifying the active minutes as high, moderate, or low following the CDC's recommendation [28].

Nutrition Tracking / Food Logging

A private online/mobile dashboard allowed participants to track personal food logs. Participants were able to log meals, snacks, treats, and beverages along with the description, quantity, and photo of the food. Each individual meal, snack, treat, and/or beverage was considered a food log entry. Four food logging-specific measures were calculated to quantify participants' adherence to food logging behavior and the potential impact of food logging on weight loss. The number of days participants logged food entries per week and the number of food log entries per week were calculated to measure the level of adherence to the behavior of food logging. The following two measures were introduced to measure participants' engagement through food logging over the 6-month intervention: the percentage of weeks participants logged at least five food log entries and at least 15 food log entries.

Statistical Analysis

All measures associated with self-monitoring behaviors involving weight measurements, dietary intake, and physical activity were included in a multiple regression analysis to predict weight loss during the intervention. Measures with statistically significant contribution to predicting weight loss were identified. To determine self-monitoring behaviors/measures that could be considered as significant predictors of weight loss, three primary regression models were built. The first primary regression model assessed two weigh-in-related measures as predictors of weight loss. The second model included five activity-related measures as predictors of weight loss. The third primary regression model assessed four measures related to food logging as predictors of weight loss. All the significant predictors (ie, self-monitoring behaviors/measures) from the primary regression model were included in an overall regression model that considered all behaviors as predictors of weight loss. The significant predictors of the overall model were considered to be the most important measures/behaviors for weight loss. Finally, each significant self-monitoring measure was analyzed in depth to reveal the impact on outcomes during the intervention period to capture the significant association between high-level monitoring to higher outcome levels. For each behavior, one-way ANOVA tests were performed to determine the association between behavior frequency and weight loss and compare behavior frequency of participants with different weight-loss levels.

Data analyses were performed using R version 3.2.3, which included dplyr 0.4.3, ggplot2 2.1.0, data.table 1.9.6, and leaps 2.9 packages. In addition, *t* tests of equal variance were conducted on continuous variables at baseline and subsequent time points for two group comparisons. One-way ANOVA was utilized to determine mean differences for greater than two group comparisons. Subsequent Tukey tests were conducted to determine mean differences. Chi-square analyses were performed to determine differences among categorical variables when appropriate. To perform best subset selection in a multiple regression analysis, an "all possible regressions" method was used to derive the best-fitting overall model using the leaps package. Alpha was set at .05 for all statistical tests to determine statistical significance.

Results

The reported results are based on the retrospective analysis evaluating the effect of various self-monitoring behaviors during a weight-loss intervention using 2113 of 3166 participants (66.74%) who completed the Retrofit 6-month weight-loss program.

Baseline Characteristics

Table 1 shows the demographic details of participants at baseline. There were no differences in age and starting BMI at baseline between male and female participants. Male participants had a higher starting weight (P<.001). There were no differences between completers and noncompleters in starting weight (P=.07) or starting BMI (P=.55), but completers had a higher mean age (mean 44.54, SD 10.72 years vs mean 42.01, SD 10.69 years, P<.001; see Multimedia Appendix 4, Table S1).

Weight Change at 6 Months

The mean weight loss at 6 months was –5.58% (SE 0.12), the mean change in BMI was –1.91 (SE 0.04), and 51.87% (1096/2113) of participants lost 5% or more of their baseline weight (see Table 2). Male participants lost a higher percentage of weight (P=.02) and had a higher BMI change (P=.01) than female participants. However, there were no significant differences between males and females in terms of the percentage of group losing 5% or more weight at 6 months.

Table 1. Baseline demographics of participants.

Demographics	Total, mean (SD) (N=2113)	Male, mean (SD) (n=860)	Female, mean (SD) (n=1253)	*P*
Age (years)	44.54 (10.72)	44.61 (10.98)	44.49 (10.54)	.81
Starting weight, (kg)	99.76 (22.92)	110.56 (22.43)	92.35 (20.14)	<.001
Starting BMI (kg/m^2)	33.84 (6.80)	34.03 (6.35)	33.71 (7.09)	.27

XSL•FO
RenderX

Table 2. Weight-loss outcomes at 6 months.

Outcome measures	Total, mean (SE) (N=2113)	Male, mean (SE) (n=860)	Female, mean (SE) (n=1253)	*P*
Weight loss (%)	–5.58 (0.12)	–5.90 (0.12)	–5.36 (0.12)	.02
BMI change	–1.91 (0.04)	–2.04 (0.07)	–1.82 (0.05)	.01
Lost 5% of baseline weight (%)	51.87 (0.01)	54.30 (0.02)	50.20 (0.01)	.07

Table 3. Multiple regression models identifying predictors of weight loss at 6 months.

Models	Coefficients			Model summary		
	Beta (SE)	*t* (df)	*P*	R^2	Adjusted R^2	*P*
Self-weigh-in				.103	.102	<.001
Weigh-ins/week (n)	–1.25 (0.19)	–6.54 (2110)	<.001			
Weeks with ≥3 weigh-ins (%)	0.018 (0.01)	1.59 (2110)	.11			
Activity				.152	.150	<.001
Tracker days/week	–0.54 (0.10)	–5.419 (2107)	<.001			
Steps/day	–0.0002 (0.0001)	–1.863 (2107)	.06			
Highly active minutes/day	–0.06 (0.01)	–4.288 (2107)	<.001			
Fairly active minutes/day	0.003 (0.004)	0.693 (2107)	.49			
Lightly active minutes/day	–0.002 (0.003)	–0.818 (2107)	.41			
Nutrition/food logging				.123	.121	<.001
Food logs/week (n)	0.01 (0.04)	0.245 (2108)	.81			
Food log days/week (n)	–1.92 (0.20)	–9.362 (2108)	<.001			
Weeks with ≥5 logs (%)	0.08 (0.01)	5.935 (2108)	<.001			
Weeks with ≥15 logs (%)	–0.01 (0.01)	–0.654 (2108)	.51			
Overall				.197	.194	<.001
Weigh-ins/week (n)	–0.417 (0.07)	–5.619 (2106)	<.001			
Tracker days/week (n)	–0.112 (0.10)	–1.081 (2106)	.28			
Steps/day (n)	–0.0001 (0.00006)	–2.269 (2106)	.02			
Highly active mins/day (n)	–0.05 (0.01)	–4.420 (2106)	<.001			
Food log days/week (n)	–1.30 (0.19)	–6.777 (2106)	<.001			
Weeks with ≥5 logs (%)	0.06 (0.01)	5.097 (2106)	<.001			

Identifying Behaviors That Matter

Model for Self-Weigh-In Behavior as a Predictor of Weight Change

To identify important self-weigh-in measures for predicting weight change, a regression model was built (R^2=.103, *P*<.001) containing the number of weigh-ins per week and the percentage of weeks with three or more weigh-ins. Table 3 shows that only the number of weigh-ins per week was identified as a significant predictor of weight change (*P*<.001).

Model for Activity-Related Behaviors as Predictor of Weight Change

To identify significant activity-related measures, a multiple regression model was constructed (R^2=.152, *P*<.001) containing the number of activity tracker usage days per week, the number of steps per day, and the number of highly, fairly, and lightly active minutes per day. Table 3 displays that the number of activity tracker days per week (*P*<.001) and the number of highly active minutes per day (*P*<.001) were significant predictors of weight change. Though the number of steps per day was not significant (*P*=.06), it was selected to be included as a predictor for weight change in the overall model based on previous study indications [27,30,39].

Model for Food Logging-Related Behaviors as a Predictor of Weight Change

To identify significant nutrition/food logging-related measures, a multiple regression model was constructed (R^2=.123, *P*<.001) containing the number of food logs per week, the number of food log days per week, the percentage of weeks with five or more food logs, and the percentage of weeks with 15 or more food logs. Table 3 shows that the number of food log days and

XSL•FO
RenderX

the percentage of weeks with five or more food logs were significant predictors of weight change.

Overall Multiple Regression Model

An overall regression model was built to predict weight change at 6 months by including all significant predictors from the self-weigh-in, activity, and food logging model. This multiple regression model (R^2=.197, P<.001) included the number of weigh-ins per week, the number of tracker usage days per week, the number of steps per week, the number of highly active minutes per week, the number of food log days per week, and the percentage of weeks with five or more food logs. Except the tracker usage days per week, all other behaviors/measures were found to be significant predictors of weight change, as shown in Table 3.

To further verify the significance of the selected behaviors/measures, an "all possible regressions" method was used to derive the best-fitting overall model. This approach of model selection determined the final model by performing an exhaustive search for the best subsets of the 11 measures listed under the primary regression models for predicting weight loss. All possible regressions included only the main effects; interactions were beyond the scope of this analysis. The best regression model contained the same five significant predictors of the overall model reported in Table 3. The next section focuses on analyzing the five significant predictors.

The Five Significant Predictors of Weight Loss

Self-Weigh-In

Based on the self-weigh-in data from 0 to 6 months, a higher weigh-in frequency was significantly associated with a higher level of weight loss at 6 months. Clinically significant weight loss (5%) was associated with at least three weigh-ins per week (see Table 4). The results of one-way ANOVA showed a significant difference of mean weight loss between different weigh-in levels (P<.001). A subsequent Tukey test confirmed the significant differences between the "≥5" weigh-in level and the remaining three levels (P<.001 for all) and between weigh-in levels "3 to 4" and "1 to 2" (P=.002) and between weigh-in levels "3 to 4" and "<1" (P=.02). Similar ANOVA tests were performed on male and female participants separately and a significant difference in mean weight loss between different weigh-in levels was found (male: P<.001; female: P<.001; see Multimedia Appendix 4, Table S2).

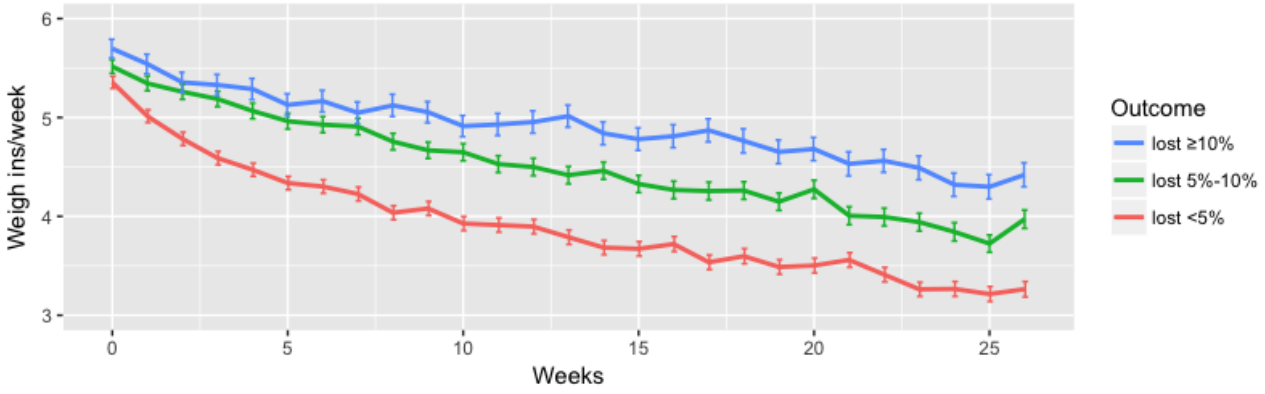
The analysis of self-weigh-in frequency of participants with different levels of weight loss showed that a higher weigh-in frequency was significantly associated with groups with higher levels of weight loss. Figure 2 presents the mean weekly weigh-in frequency among participants of three outcome levels: "lost ≥10%" (388/2113, 18.36%), "lost 5%-10%" (707/2113, 34.46%), and "lost <5%" (1018/2113, 48.18%). For all other analyses throughout the paper, behavior frequency based on outcome levels uses the same outcome-based participant groups. It showed a clear difference in weigh-in frequency throughout the 6-month program. The mean weigh-in frequencies over 6 months were mean 4.70 (SE 0.09), mean 4.21 (SE 0.07), and mean 3.40 (SE 0.05) weigh-ins per week for the lost ≥10%, lost 5%-10%, and lost <5% groups, respectively (P<.001). Additional ANOVA tests performed on male and female participants separately showed a similar significant difference of mean weigh-in frequency between different outcome levels (male: P<.001; female: P<.001, see Multimedia Appendix 4, Table S3).

XSL•FO
RenderX

Table 4. Weight-loss outcomes of participants for different behavior frequencies.

Self-monitoring behaviors	n (%)	Weight loss (%), mean (SE)	*P*
Weigh-in frequency per week			<.001
<1	89 (4.21)	–3.41 (0.58)	
1-2	636 (30.10)	–4.08 (0.20)	
3-4	690 (33.65)	–5.09 (0.19)	
≥5	698 (33.03)	–7.82 (0.20)	
Steps per day			<.001
<5000	797 (37.72)	–3.68 (0.17)	
5000-7499	604 (28.58)	–5.56 (0.20)	
7500-9999	429 (20.30)	–7.03 (0.26)	
≥10,000	283 (13.39)	–9.03 (0.34)	
Highly active minutes per week			<.001
<60	897 (42.41)	–4.14 (0.17)	
60-119	525 (24.82)	–5.71 (0.21)	
120-179	299 (14.14)	–5.85 (0.29)	
≥180	394 (18.63)	–8.64 (0.28)	
Food log days per week			<.001
<1	316 (14.96)	–3.67 (0.33)	
1-2	596 (28.21)	–4.32 (0.20)	
3-4	565 (26.74)	–5.15 (0.19)	
≥5	636 (30.10)	–8.20 (0.21)	
Food logs per week			<.001
<5	617 (29.20)	–4.37 (0.21)	
5-9	405 (19.17)	–4.66 (0.24)	
10-14	297 (14.06)	–5.11 (0.29)	
15-19	247 (11.69)	–5.46 (0.32)	
≥20	547 (25.89)	–8.10 (0.23)	

Figure 2. Weekly mean weigh-in frequency of participants between 0 and 6 months. Error bars indicate the standard error of the mean.



Daily Steps

Based on the steps data of participants between 0 and 6 months, a higher level of daily steps was significantly associated with a higher level of weight loss. Table 4 presents the results of the weight-loss analysis performed by dividing the participants to different levels of daily step counts. Daily steps of 5000 to 7499 or more were associated with clinically significant weight loss at 6 months. One-way ANOVA showed a significant difference in mean weight loss between different daily step levels (P<.001). A subsequent Tukey test confirmed significant mean differences between all levels of daily steps (P<.001). Similar ANOVA

XSL•FO
RenderX

tests performed on male and female participants separately showed a significant difference in mean weight loss between the different daily step levels (male: $P<.001$; female: $P<.001$; see Multimedia Appendix 4, Table S4).

The analysis of daily steps of participants with different levels of weight loss showed that a higher daily step count was significantly associated with groups with higher levels of weight loss. Figure 3 presents the weekly mean steps per day among participants of three outcome levels. The lost ≥10% group consistently maintained significantly higher daily steps throughout the 6-month program. The mean daily steps over 6 months were mean 8077.79 (SE 171.52), mean 6657.09 (SE 117.13), and mean 5276.91 (SE 95.08) steps per day for the lost ≥10%, lost 5%-10%, and lost <5% groups, respectively ($P<.001$). Male and female participants separately showed a similar significant difference in mean daily steps between the different outcome levels (male: $P<.001$; female: $P<.001$; see Multimedia Appendix 4, Table S5).

Highly Active Minutes

Higher levels of highly active minutes were significantly associated with higher levels of weight loss. Percentage of weight loss was calculated by dividing the participants into different levels of highly active minutes per week (Table 4). Higher-intensity activity for 60 minutes or more per week was associated with clinically significant weight loss. There was a significant difference in mean weight loss between different weekly active minutes levels ($P<.001$). A subsequent Tukey test showed significant differences between the "≥180" active minutes level and the remaining three levels ($P<.001$ for all) and between "120-179" active minutes level and "<60" ($P<.001$) and between "60-119" active minutes level and "<60" ($P<.001$). Similar ANOVA tests performed on male and female participants separately showed a similar significant difference in mean weight loss between different daily highly active minutes levels (male: $P<.001$; female: $P<.001$, see Multimedia Appendix 4, Table S6).

The analysis of highly active minutes of the participants with different levels of weight loss showed that higher amounts of highly active minutes were significantly associated with groups with higher levels of weight loss. Figure 4 shows the weekly mean highly active minutes among participants in three outcome levels. Similar to daily steps, the lost ≥10% group consistently had a significantly higher level of high-intensity activity throughout the 6-month program. The mean weekly highly active minutes over 6 months were mean 154.33 (SE 6.47), mean 115.63 (SE 3.91), and mean 79.03 (SE 2.53) minutes per week for the lost ≥10%, lost 5%-10%, and lost <5% groups, respectively ($P<.001$). Male and female participants separately showed a similar significant difference in mean highly active minutes between different outcome levels (male: $P<.001$; female: $P<.001$; see Multimedia Appendix 4, Table S7).

Figure 3. Weekly mean steps per day of participants between 0 and 6 months. Error bars indicate the standard error of the mean.

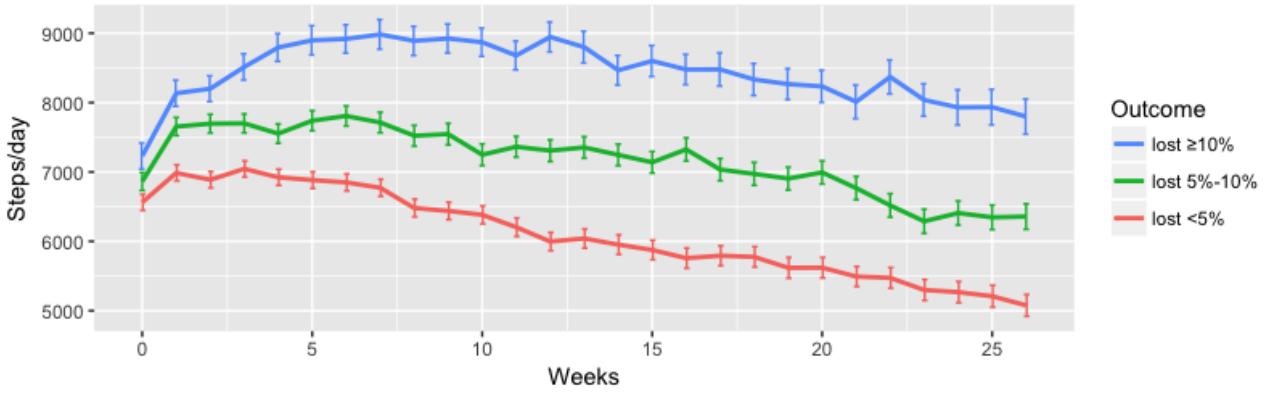
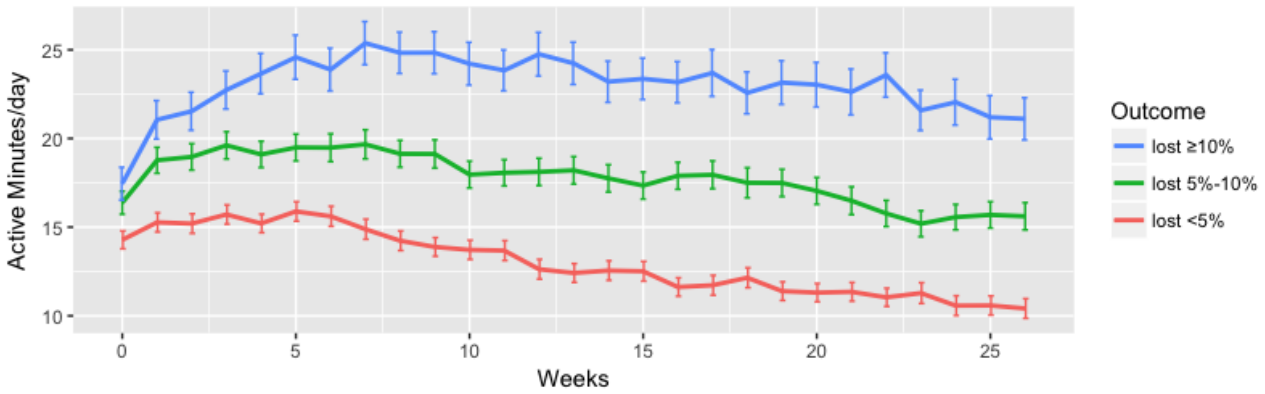


Figure 4. Weekly mean highly active minutes of participants between 0 and 6 months. Error bars indicate the standard error of the mean.



XSL•FO
RenderX

Food Log Days

Analysis of participants' food log data over 0 to 6 months showed that a higher number of food log days per week was significantly associated with a higher level of weight loss. Table 4 presents the weight-loss percentages of the groups of participants at different levels of food log days per week. Clinically significant weight loss was associated with at least three or more days of food logging per week. There was a significant difference in weight loss between different levels of weekly food log days (P<.001). A subsequent Tukey test found significant differences between "≥5" food log days and the remaining three levels (P<.001 for all) and between "3-4" and "1-2" food log days (P=.03) and between "3-4" and "<1" food log days (P<.001). Similar ANOVA tests performed on male and female participants separately showed a similar significant difference in mean weight loss between different numbers of food log days (male: P<.001; female: P<.001; see Multimedia Appendix 4, Table S8).

The analysis of food log days of the participants with different levels of weight loss showed that a higher number of food log days per week was significantly associated with groups with higher levels of weight loss. Figure 5 shows the weekly mean food log days among the participants in the three outcome levels. Participants in the higher outcome levels logged their food a significantly higher number of days throughout the 6-month program than the lowest outcome group. The mean weekly food log days over 6 months were mean 4.44 (SE 0.11), mean 3.92 (SE 0.08), and mean 2.90 (SE 0.60) days per week for the lost ≥10%, lost 5%-10%, and lost <5% groups, respectively (P<.001). Male and female participants separately showed a similar significant difference in mean food log days between different outcome levels (male: P<.001; female: P<.001; see Multimedia Appendix 4, Table S9).

Percentage of Weeks With Five or More Food Logs

Additional analysis of food logs showed that participants with a higher level of weight loss were significantly associated with a higher percentage of weeks with five or more food logs. There was a significant difference in percentage of weeks with five or more food logs between different weight loss outcome levels (lost ≥10%: mean 69.40%, SE 1.72; lost 5%-10%: mean 63.61%, SE 1.20; lost <5%: mean 49.14%, SE 0.97; P<.001). A subsequent Tukey test was performed, which found significant mean differences between all outcome levels (lost ≥10% and lost <5%: P<.001; lost ≥10% and lost 5%-10%: P=.01; lost 5%-10% and lost <5%: P<.001). Additional ANOVA tests performed on male and female participants separately showed a similar significant difference in percentage of weeks with five or more food logs between the different outcome levels (male: P<.001; female: P<.001; see Multimedia Appendix 4, Table S10).

Based on the analysis presented in previous sections, food logging is very critical for weight loss during the 6-month intervention. Hence, additional analysis is presented in Table 4 that shows the percentage weight loss for participants in different food log groups per week. A higher number of food logs per week was significantly associated with a higher level of weight loss (P<.001). Further analysis to understand differences in weight-loss outcomes for male and female participants between different food log groups showed a similar difference (male: P<.001; female: P<.001; see Multimedia Appendix 4, Table S11).

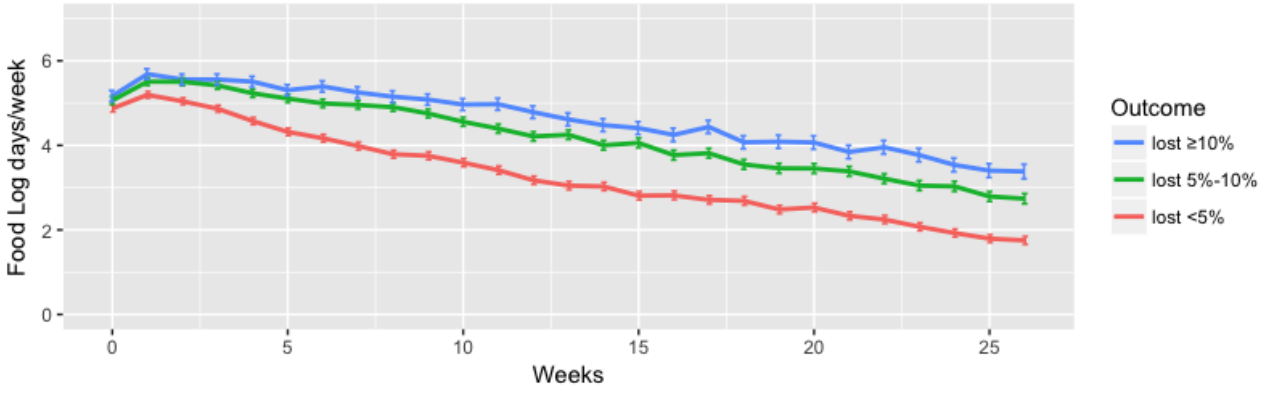
Discussion

Principal Findings

The results provide strong support for the use of self-monitoring in weight-management programs. Participants who complied more with body weight, physical activity, and food intake self-monitoring lost more weight than those who complied less. In a multiple regression equation, each category of self-monitoring contributed significantly to the prediction of weight loss. Furthermore, the independent analysis showed a significant association between each self-monitoring behavior and weight loss. Overall, the use of self-monitoring was found to have a high impact on weight management.

Advances in technology, such as wireless scales and physical activity trackers, make it easier to self-monitor weight and physical activity, and are recommended in weight-management programs. Food logging still requires that a participant take time to record food intake, but technology has made it a faster and simpler process. However, there is a great need for developing new technology to reduce the time, effort, and accuracy in self-monitoring food intake.

Figure 5. Weekly mean food log days of participants between 0 and 6 months. Error bars indicate the standard error of the mean.



XSL•FO
RenderX

Some report that self-monitoring of single behaviors, such as body weight and physical activity [2,27,40], may not be associated with greater weight loss. Our results found benefits of self-monitoring in each behavior category: weight, physical activity, and food intake. We found that self-monitoring all these behaviors together had the greatest predictive value for weight loss. Based on these results, it may be important to promote these self-monitoring behaviors together in an intervention or weight-loss program.

Significant Predictors of Weight Loss

Self-Weigh-In

The number of self-weigh-ins per week was identified as a significant predictor of weight loss (P<.001). Self-weighing at least three times per week is associated with higher weight loss. An even higher level of weight loss is associated when weighing more than five times per week. Evidence has shown that instructing participants to weigh in at least three times per week does increase weight loss during an active weight-loss period, and a higher frequency of weighing in is associated with greater weight-loss success [15-18]. Additionally, females tend to weigh in more frequently than males, which is a unique finding due to an historically small percentage of male participants in weight-management studies and a lack of evidence around gender comparison [15-18,20,25,40-48].

Daily Steps

The number of steps per week (P=.02) is a significant predictor of weight change. Higher step counts are associated with greater levels of weight loss, which has been shown in previous literature [27,30,39]. Also, our results confirm that men tend to have a higher daily step count than women, similar to that seen in the literature [26].

Highly Active Minutes

A minimum of 60 highly active minutes per week is significantly associated with higher levels of weight loss. Greater levels of highly active minutes are also significantly associated with higher weight loss outcomes. Males overall have a significantly higher level of active minutes than females at all weight-loss outcome levels (P<.001). Currently, there is a lack of evidence around measuring active minutes with activity trackers associated with weight-loss outcomes. However, there is some evidence that men log more exercise than females and have a greater exercise dependence [33,43,49].

Food Log Days

A higher number of food log days per week increases adherence to the self-monitoring behavior of food logging, which supports behavioral change as explained through self-regulation theory [31]. Food logging at least three days per week was significantly associated with higher levels of weight loss. Other studies have found that greater weight loss is achieved with a higher frequency of food logs, specifically three or more days per week [31-33,36,43].

Percentage of Weeks With Five or More Food Logs

A higher percentage of weeks with five or more food logs is significantly associated with higher levels of weight loss (P<.001). Additionally, the more times a participant logs their food per week increases their likelihood of successful weight loss [31-33]. Women tend to log their food more frequently than men do. However, this is a unique finding due to a historically small percentage of male participants in weight-management studies and a lack of evidence around gender comparison [31-35].

Strengths and Limitations

This study has several strengths, including the reporting of real-world weight-loss outcomes and providing a more focused analysis into weight-management behaviors to determine what behaviors are more significant in a behavioral weight-loss program. Participants were clients of Retrofit and not recruited or provided with incentives to participate in the study. All clients who met the starting BMI, age, and weight inclusion criteria and logged a weight at 6 months were included as participants and not removed from the population due to lack of success on the program, which is an uncommon research practice [44]. We conclude that this study adds value and brings a novel approach to the best practices around behaviors in weight management. Additionally, gender comparisons were able to be reported due to the unusually high population of males enrolled as participants, which is also a significant strength to understand which behaviors are more valuable to men and women in the weight-loss process.

The study also has some limitations, including the retrospective analysis study design, which does not allow any causal inferences based on the critical observations. Moreover, the adherence to different behaviors were evaluated using data from program completers, which limits the ability to generalize impact of the behavior on all participants. However, to determine effective levels of self-monitoring that can guarantee clinically significant weight loss, it is critical to study participants with known end weights. Lastly, due to use of the real-world population in this study, it is unknown if participants were integrating any other self-monitoring devices or practices outside of the Retrofit Program components.

Future Research

With a lack of real-world research in the commercial weight-loss industry, Retrofit encourages all commercial weight-loss programs to publish similar data to enhance understanding of which self-monitoring behaviors matter most in a weight-loss program. Reporting real-world data in relation to targeted behaviors allows commercial weight-loss programs to not only structure protocols and client strategies to increase weight-loss success, but also improve a participant's weight maintenance success. By narrowing in on the specific behaviors to build as life-long habits, commercial weight-loss programs will increase efficacy and establish our ability as an industry to overcome the obesity crisis.

Recommended future research includes studying self-monitoring behaviors beyond 6 months and each behavior's impact on weight maintenance. Also, further analysis around gender differences and self-monitoring behaviors is of interest to determine if specific behaviors should be encouraged more

XSL•FO
RenderX

frequently among female versus male participants, specifically around food logging and activity levels.

Conclusions

In conclusion, participants on the Retrofit Program lost a mean –5.58% (SE 0.12) and had a BMI change of mean –1.91 (SE 0.04) in 6 months with nearly 51.87% (1096/2113) of participants losing 5% or more of their baseline weight. Self-monitoring behaviors, such as self-weigh-in, daily step counts, high-intensity activity, and persistent food logging were shown to be significant predictors of weight change at 6 months. Specifically, weighing in three times or more per week, having a minimum of 60 highly active minutes per week, food logging for three days or more per week, and having a higher percentage of weeks with five or more food logs increased participant's weight-loss success.

Acknowledgments

Members of the Retrofit Advisory Board provided comments and professional insight around the data and results.

Conflicts of Interest

SP, RA, SB, and AM are employees of Retrofit, Inc, with equity in the company. JH, RK, and RL are active members of the Retrofit, Inc Advisory Board, with equity in the company.

Multimedia Appendix 1

Retrofit logo.

[JPG File, 64KB - jmir_v19i5e160_app1.jpg]

Multimedia Appendix 2

The technology provided included a Wi-Fi–enabled scale, activity tracker, and access to a private dashboard. The dashboard was accessible via Web and mobile apps.

[PNG File, 135KB - jmir_v19i5e160_app2.png]

Multimedia Appendix 3

Retrofit client dashboard for logging food and exercise accessible via Web and mobile apps.

[PNG File, 167KB - jmir_v19i5e160_app3.png]

Multimedia Appendix 4

Supplementary Tables.

[PDF File (Adobe PDF File), 129KB - jmir_v19i5e160_app4.pdf]

References

1. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
2. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. JAMA 2016 Sep 20;316(11):1161-1171. [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
3. Centers for Disease Control and Prevention. Adult obesity facts URL: https://www.cdc.gov/obesity/data/adult.html [accessed 2017-01-26] [WebCite Cache ID 6noA6pOKK]
4. Bray G, Bouchard C, James W, Wing R. Behavioral approaches to the treatment of obesity. In: Handbook of Obesity: Clinical Applications. New York: Marcel Dekker; 2004:147-167.
5. van der Mark M, Jonasson J, Svensson M, Linné Y, Rossner S, Lagerros YT. Older members perform better in an internet-based behavioral weight loss program compared to younger members. Obes Facts 2009;2(2):74-79. [doi: 10.1159/000209383] [Medline: 20054209]
6. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
7. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]

XSL•FO
RenderX

8. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: a pilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164. [doi: 10.1038/oby.2007.520] [Medline: 17228043]
9. Harvey-Berino J, Pintauro S, Buzzell P, Gold EC. Effect of internet support on the long-term maintenance of weight loss. Obes Res 2004 Feb;12(2):320-329. [doi: 10.1038/oby.2004.40] [Medline: 14981225]
10. Polzien KM, Jakicic JM, Tate DF, Otto AD. The efficacy of a technology-based system in a short-term behavioral weight loss intervention. Obesity (Silver Spring) 2007 Apr;15(4):825-830. [doi: 10.1038/oby.2007.584] [Medline: 17426316]
11. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
12. Rothert K, Strecher VJ, Doyle LA, Caplan WM, Joyce JS, Jimison HB, et al. Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. Obesity (Silver Spring) 2006 Feb;14(2):266-272 [FREE Full text] [doi: 10.1038/oby.2006.34] [Medline: 16571852]
13. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. JAMA 2003 Apr 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
14. Womble LG, Wadden TA, McGuckin BG, Sargent SL, Rothman RA, Krauthamer-Ewing ES. A randomized controlled trial of a commercial internet weight loss program. Obes Res 2004 Jun;12(6):1011-1018. [doi: 10.1038/oby.2004.124] [Medline: 15229342]
15. Helander EE, Vuorinen A, Wansink B, Korhonen IK. Are breaks in daily self-weighing associated with weight gain? PLoS One 2014 Nov;9(11):e113164 [FREE Full text] [doi: 10.1371/journal.pone.0113164] [Medline: 25397613]
16. Pacanowski CR, Bertz FC, Levitsky DA. Daily self-weighing to control body weight in adults: a critical review of the literature. Sage Open 2014;4(4):1-16 [FREE Full text] [doi: 10.1177/2158244014556992] [Medline: 27127719]
17. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. Daily self-weighing and adverse psychological outcomes: a randomized controlled trial. Am J Prev Med 2014 Jan;46(1):24-29. [doi: 10.1016/j.amepre.2013.08.006] [Medline: 24355668]
18. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265. [doi: 10.1002/oby.20946] [Medline: 25521523]
19. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. Psychiatr Clin North Am 2011 Dec;34(4):841-859 [FREE Full text] [doi: 10.1016/j.psc.2011.08.006] [Medline: 22098808]
20. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]
21. Vanwormer JJ, Martinez AM, Martinson BC. Self-weighing promotes weight loss for obese adults. Am J Prev Med 2009;36(1):70-73. [Medline: 20151204]
22. Akers JD, Cornett RA, Savla JS, Davy KP, Davy BM. Daily self-monitoring of body weight, step count, fruit/vegetable intake, and water consumption: a feasible and effective long-term weight loss maintenance approach. J Acad Nutr Diet 2012 May;112(5):685-692 [FREE Full text] [doi: 10.1016/j.jand.2012.01.022] [Medline: 22709772]
23. Linde JA, Jeffery RW, French SA, Pronk NP, Boyle RG. Self-weighing in weight gain prevention and weight loss trials. Ann Behav Med 2005 Dec;30(3):210-216. [doi: 10.1207/s15324796abm3003_5] [Medline: 16336072]
24. Butryn ML, Phelan S, Hill JO, Wing RR. Consistent self-monitoring of weight: a key component of successful weight loss maintenance. Obesity (Silver Spring) 2007 Dec;15(12):3091-3096. [doi: 10.1038/oby.2007.368] [Medline: 18198319]
25. 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]
26. Bassett DR, Wyatt HR, Thompson H, Peters JC, Hill JO. Pedometer-measured physical activity and health behaviors in US adults. Med Sci Sports Exerc 2010 Oct;42(10):1819-1825 [FREE Full text] [doi: 10.1249/MSS.0b013e3181dc2e54] [Medline: 20305579]
27. Cai X, Qiu SH, Yin H, Sun ZL, Ju CP, Zügel M, et al. Pedometer intervention and weight loss in overweight and obese adults with Type 2 diabetes: a meta-analysis. Diabet Med 2016 Aug;33(8):1035-1044 [FREE Full text] [doi: 10.1111/dme.13104] [Medline: 26926674]
28. Musto A, Jacobs K, Nash M, DelRossi G, Perry A. The effects of an incremental approach to 10,000 steps/day on metabolic syndrome components in sedentary overweight women. J Phys Act Health 2010 Nov;7(6):737-745. [Medline: 21088304]
29. Cadmus-Bertram L, Marcus BH, Patterson RE, Parker BA, Morey BL. Use of the Fitbit to measure adherence to a physical activity intervention among overweight or obese, postmenopausal women: self-monitoring trajectory during 16 weeks. JMIR Mhealth Uhealth 2015;3(4):e96 [FREE Full text] [doi: 10.2196/mhealth.4229] [Medline: 26586418]
30. Conroy MB, Yang K, Elci OU, Gabriel KP, Styn MA, Wang J, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574 [FREE Full text] [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]

XSL•FO
RenderX

31. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
32. Krukowski RA, Harvey-Berino J, Bursac Z, Ashikaga T, West DS. Patterns of success: online self-monitoring in a web-based behavioral weight control program. Health Psychol 2013 Feb;32(2):164-170. [doi: 10.1037/a0028135] [Medline: 22545978]
33. Johnson F, Wardle J. The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss programme: a retrospective analysis. Int J Behav Nutr Phys Act 2011 Aug 02;8(1):83 [FREE Full text] [doi: 10.1186/1479-5868-8-83] [Medline: 21810222]
34. Turk MW, Elci OU, Wang J, Sereika SM, Ewing LJ, Acharya SD, et al. Self-monitoring as a mediator of weight loss in the SMART randomized clinical trial. Int J Behav Med 2013 Dec;20(4):556-561 [FREE Full text] [doi: 10.1007/s12529-012-9259-9] [Medline: 22936524]
35. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of self-monitoring in a web-based weight loss program by extra individualized feedback and reminders: randomized trial. J Med Internet Res 2016 Apr 12;18(4):e82 [FREE Full text] [doi: 10.2196/jmir.4100] [Medline: 27072817]
36. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
37. Retrofit. URL: https://www.retrofitme.com/ [accessed 2017-05-06] [WebCite Cache ID 6qG26LCBV]
38. Painter S, Ditsch G, Ahmed R, Hanson NB, Kachin K, Berger J. Retrofit weight-loss outcomes at 6, 12, and 24 months and characteristics of 12-month high performers: a retrospective analysis. JMIR Mhealth Uhealth 2016 Aug 22;4(3):e101 [FREE Full text] [doi: 10.2196/mhealth.5873] [Medline: 27549134]
39. Dwyer T, Hosmer D, Hosmer T, Venn AJ, Blizzard CL, Granger RH, et al. The inverse relationship between number of steps per day and obesity in a population-based sample: the AusDiab study. Int J Obes (Lond) 2007 May;31(5):797-804. [doi: 10.1038/sj.ijo.0803472] [Medline: 17047641]
40. Madigan CD, Daley AJ, Lewis AL, Aveyard P, Jolly K. Is self-weighing an effective tool for weight loss: a systematic literature review and meta-analysis. Int J Behav Nutr Phys Act 2015 Aug 21;12:104 [FREE Full text] [doi: 10.1186/s12966-015-0267-4] [Medline: 26293454]
41. Saad L. Gallup. 2011. To lose weight, Americans rely more on dieting than exercise URL: http://www.webcitation.org/6qD3hIfW7 [accessed 2017-05-04] [WebCite Cache ID 6qD3hIfW7]
42. Wadden TA, Webb VL, Moran CH, Bailer BA. Lifestyle modification for obesity: new developments in diet, physical activity, and behavior therapy. Circulation 2012 Mar 6;125(9):1157-1170 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.039453] [Medline: 22392863]
43. Serrano KJ, Yu M, Coa KI, Collins LM, Atienza AA. Mining health app data to find more and less successful weight loss subgroups. J Med Internet Res 2016 Jun 14;18(6):e154 [FREE Full text] [doi: 10.2196/jmir.5473] [Medline: 27301853]
44. Gudzune KA, Bleich SN, Clark JM. Efficacy of commercial weight-loss programs. Ann Intern Med 2015 Sep 1;163(5):399. [doi: 10.7326/L15-5130-3] [Medline: 26322706]
45. Wing RR, Lang W, Wadden TA, Safford M, Knowler WC, Bertoni AG, Look AHEAD Research Group. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care 2011 Jul;34(7):1481-1486 [FREE Full text] [doi: 10.2337/dc10-2415] [Medline: 21593294]
46. Appel LJ, Clark JM, Yeh H, Wang N, Coughlin JW, Daumit G, et al. Comparative effectiveness of weight-loss interventions in clinical practice. N Engl J Med 2011 Nov 24;365(21):1959-1968 [FREE Full text] [doi: 10.1056/NEJMoa1108660] [Medline: 22085317]
47. Almeida FA, You W, Harden SM, Blackman KC, Davy BM, Glasgow RE, et al. Effectiveness of a worksite-based weight loss randomized controlled trial: the worksite study. Obesity (Silver Spring) 2015 Apr;23(4):737-745 [FREE Full text] [doi: 10.1002/oby.20899] [Medline: 25678325]
48. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):e236-e243 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]
49. Weik M, Hale BD. Contrasting gender differences on two measures of exercise dependence. Br J Sports Med 2009 Mar;43(3):204-207. [doi: 10.1136/bjsm.2007.045138] [Medline: 18308882]

Abbreviations

BMI: body mass index

XSL•FO
RenderX

Edited by G Eysenbach; submitted 09.02.17; peer-reviewed by L Chai, J Chaplin; comments to author 24.02.17; revised version received 27.03.17; accepted 14.04.17; published 12.05.17.

Please cite as:
Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, Margulies A
What Matters in Weight Loss? An In-Depth Analysis of Self-Monitoring
J Med Internet Res 2017;19(5):e160
URL: http://www.jmir.org/2017/5/e160/
doi:10.2196/jmir.7457
PMID:28500022

©Stefanie Lynn Painter, Rezwan Ahmed, James O Hill, Robert F Kushner, Richard Lindquist, Scott Brunning, Amy Margulies. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 12.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Web-Based Cognitive Behavior Therapy for Depression in People With Diabetes Mellitus: A Randomized Controlled Trial

Jill Newby[1,2], PhD; Lisa Robins[3,4,5], BSc, MPsych (Clinical), GradCert (DiabEdn&Mgt); Kay Wilhelm[3,4], MBBS, MD; Jessica Smith[6], BSc, MSc; Therese Fletcher[3], BAppSc, MPH, GradCert (DiabEdn&Mgt), GradDipEd; Inika Gillis[3], MPsych (Clinical); Trevor Ma[4], MBBS, BBus; Adam Finch[4], BSc, MPsych (Clinical), GradDipPsych; Lesley Campbell[5], MBBS, FRCP, FRACP; Gavin Andrews[6], AO, MD

[1]School of Psychology, The University of New South Wales, Randwick, Australia

[2]The University of New South Wales at St Vincent's Hospital, Clinical Research Unit for Anxiety and Depression, School of Psychiatry, Darlinghurst, Australia

[3]St Vincent's Health Australia, Faces in the Street, Darlinghurst, Australia

[4]St Vincent's Hospital, Sydney, Consultation Liaison Psychiatry, Darlinghurst, Australia

[5]Garvan Institute, St Vincent's Hospital, Sydney, Diabetes Centre, Darlinghurst, Australia

[6]School of Psychiatry, Clinical Research Unit for Anxiety and Depression, The University of New South Wales at St Vincent's Hospital, Darlinghurst, Australia

Corresponding Author:
Jill Newby, PhD
School of Psychology
The University of New South Wales
1302 Mathews Building
Randwick, 2052
Australia
Phone: 61 293853425
Email: j.newby@unsw.edu.au

Abstract

Background: Depression is twice as common in diabetes mellitus (DM) as the general population and is associated with adverse health outcomes, but access to evidence-based therapies such as cognitive behavioral therapy (CBT) is limited in routine diabetes care. Past research has shown that generic Internet-based cognitive behavioral therapy (iCBT) is an effective treatment for depression in the general population, but it has never been evaluated in people with comorbid depression and DM.

Objective: The aim of our study was to examine the efficacy of a generic 6-lesson iCBT delivered over 10 weeks in people with major depressive disorder (MDD) and DM.

Methods: Participants with comorbid MDD and DM (type 1 or 2) were recruited online and randomized to an iCBT program with therapist support provided by phone and email (n=42) or a treatment as usual (TAU, n=49) control group. Outcomes were assessed through Web-based self-report questionnaires and the trial was Web-based with no face-to-face components. Primary outcomes were self-reported depression (patient health questionnaire-9, PHQ-9), diabetes-related distress (problem areas in diabetes, PAID), and self-reported glycemic control (hemoglobin A1c, HbA1c). Secondary outcomes were general distress (Kessler 10-item psychological distress scale, K-10) and disability (short form 12-item, SF-12), generalized anxiety (generalized anxiety disorder 7-item, GAD-7), and somatization (PHQ-15). The iCBT group was assessed at 3 months.

Results: A total of 27 participants (66%; 27/41) completed the iCBT program. Analyses indicated between-group superiority of iCBT over TAU at posttreatment on PHQ-9 (g=0.78), PAID (g=0.80), K-10 (g=1.06), GAD-7 (g=0.72), and SF-12 mental well-being scores (g=0.66), but no significant differences in self-reported HbA1c levels (g=0.14), SF-12 physical well-being, or PHQ-15 scores (g=0.03-0.21). Gains were maintained at 3-month follow-up in the iCBT group, and the 87% (27/31) of iCBT participants who were interviewed no longer met criteria for MDD. Clinically significant change following iCBT on PHQ-9 scores was 51% (21/41) versus 18% (9/49) in TAU.

Conclusions: iCBT for depression is an efficacious, accessible treatment option for people with diabetes. Future studies should explore whether tailoring of iCBT programs improves acceptability and adherence, and evaluate the long-term outcomes following iCBT.

XSL•FO
RenderX

Trial Registration: Australian and New Zealand Clinical Trials Registry (ACTRN): 12613001198718; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=365208&isReview=true (Archived by WebCite at http://www.webcitation.org/6qCR8Fi9V)

(J Med Internet Res 2017;19(5):e157) doi:10.2196/jmir.7274

KEYWORDS

depression; diabetes mellitus; cognitive behavior therapy; diabetes-related distress; computer-assisted therapy

Introduction

Diabetes mellitus (DM) is one of the largest causes of global health burden, affecting an estimated 415 million adults, with prevalence rates expected to rise to 643 million by 2040. The burden of this disease is significant in personal as well as economic terms, accounting for 5 million deaths, and 12% of global health care expenditure [1]. Depression is twice as common for people with DM compared with the general population [2], and places a substantial burden on people living with the disease, both in terms of personal suffering, and contributing to a range of adverse mental and physical health outcomes. Experiencing depression in the context of DM is associated with a range of adverse health outcomes, including poorer quality of life [3], higher diabetes-related distress [4], and poorer diabetes self-management, including lower adherence to self-care regimes (eg, physical activity and diet) [5]. Depression in DM is also associated with increased health care utilization and expenditure (eg, [6]), poorer medical outcomes, including worse glycemic control [7], and increased risk for diabetes-related complications and early mortality [8].

Effective management of depression is essential to reduce individual suffering and prevent these adverse outcomes [9]. Studies show that screening for depression has little influence on depression outcomes [10], and that proactive detection and treatment of depression is essential to reduce the burden of depression in people living with DM. A range of proactive treatment approaches have been found to be effective for treating depression in DM, including collaborative care [11], pharmacological interventions such as selective serotonin reuptake inhibitors (SSRIs), and psychological treatments [12]. Of the available psychological treatments, cognitive behavioral therapy (CBT) has the most empirical support for treatment for depression in DM. Meta-analyses show that CBT achieves clinically significant reductions in depression symptoms and improved quality of life [12]. Leading clinical practice guidelines therefore recommend routine assessment, screening, and treatment of depression in patients with DM [13], but despite these recommendations, depression is both underrecognized and undertreated: in routine care for DM, depression remains untreated in 50% of patients [6]. We now need new ways to deliver evidence-based depression interventions to people with DM, and overcome barriers underlying the shortfall of depression treatment in DM care, such as the lack of integration of mental and physical health services, and limited access to clinicians with expertise in addressing mental health issues in DM.

Delivering evidence-based psychological treatments via the Internet has potential to overcome some of these barriers. The efficacy of therapist-supported iCBT is now well established for the treatment of depression and anxiety disorders in the general population [14], and these positive effects generalize to routine care [15]. iCBT programs are now successfully being used to encourage proactive self-management of a range of chronic health conditions including chronic pain and irritable bowel syndrome [16]. iCBT has been shown to be as efficacious as face-to-face CBT [17], but can be delivered at a fraction of the cost and clinical time [18]. In contrast to pharmacotherapies that often have significant side effects [19], iCBT does not lead to harm, and deterioration in symptoms is rare [20].

Despite the potential for delivering mental health treatments on the Web to depression in people with DM, there are only two published randomized controlled trials (RCTs) examining the impact of Web-based programs for the treatment of depression in people with DM. In a sample of 225 Dutch adults with elevated depression (Center for Epidemiologic Studies Depression Scale [CES-D] scores >16) and type 1 (T1) or type 2 (T2) DM, van Bastelaar and colleagues found that a guided Web-based CBT program for depression with DM-specific content was more effective than treatment as usual (TAU) in reducing depression symptoms, increasing depression remission rates, and reducing diabetes-related distress with small-to-moderate between-groups effective sizes at 1-month follow-up (d=0.29 for intention-to-treat [ITT] analyses) [21]. The second trial with 260 German adults with T1 and T2 DM and elevated depression (CES-D ≥23), found that a guided Web-based intervention to reduce depression with DM-specific content was effective in reducing both depressive symptoms (ITT d=0.89) and diabetes-specific emotional distress (d=0.56) compared with a brief unguided Web-based psychoeducation program for depression [22].

These trials show promising results for using Web-based CBT programs to treat depression symptoms in people with DM, although it is not known whether results generalize to clinically depressed individuals with major depressive disorder (MDD). To our knowledge, there are no studies examining whether general or unmodified iCBT depression programs—without any content tailored to DM—are effective for people with DM. With generic evidence-based iCBT programs for depression now available for general public use, they present a novel opportunity to reduce the problem of depression in routine DM care, but need to be tested first to evaluate whether they are acceptable and effective in people with DM. Although RCTs have shown that face-to-face protocol-driven CBT interventions are effective for treating depression in DM, in-person CBT differs to iCBT in that it is able to be tailored and individualized to the individuals presenting problems. In contrast iCBT

XSL•FO
RenderX

programs are typically highly standardized and fixed in format, and therefore need to be evaluated further in the context of DM.

This study is the first RCT to evaluate the efficacy of a generic clinician-guided iCBT program for MDD in people with T1 or T2 DM. The iCBT program has been previously demonstrated to be effective in RCTs [23] and effectiveness trials [15]. Whereas our primary aim was to examine the acceptability of the program for patients with DM and the efficacy of the iCBT program in reducing depression, we also sought to examine its impact on diabetes-related distress, generalized anxiety, mental and physical well-being, and glycemic control (hemoglobin A1c [HbA1c] levels). We hypothesized that adults with T1 or T2 DM who received iCBT would demonstrate significantly lower levels of depression, diabetes-related distress, and better self-reported glycemic control compared with the TAU control group at posttreatment.

Methods

This study was approved by the St Vincent's Hospital Human Research Ethics Committee (HREC/13/SVH/291). The trial was registered with the Australian and New Zealand Clinical Trials registry (ACTRN12613001198718).

Design

Study details are reported in the published protocol [24]. Eligible participants were randomly allocated to the iCBT program or the TAU group who received the intervention after a 10-week waiting period. Simple randomization was used (1:1 ratio), and randomization numbers were generated by an independent research assistant using random.org. Group allocation was concealed in sequentially numbered opaque sealed envelopes. All participants provided electronic informed consent before participating.

Eligibility

Inclusion criteria included: Australian resident, age 18 years or older, fluent in English, access to a computer and Internet, self-reported diagnosis of T1 or T2 DM, meet criteria for MDD according to telephone-administered diagnostic interview, and provide personal and general practitioner (GP) contact details. Participants were excluded if they had a self-reported diagnosis of bipolar affective disorder, psychotic disorder or substance use disorder, or were taking antipsychotics or benzodiazepines. Participants were excluded if they had commenced CBT in the past month, or changed antidepressant medication in the past 2 months. Participants scoring either <5 (normal range) or >23 (very severe) on the patient health questionnaire-9 (PHQ-9) were excluded, and those identified as being at significant risk of suicide or deliberate self-harm in the telephone risk assessment were also excluded and referred to appropriate treatments.

Setting and Procedure

Participants were recruited from September 2013 to June 2015 by advertisements on DM websites, social media, and fliers. Interested applicants applied on the Web via the St Vincent's Hospital, Sydney's Virtual Clinic website, with an email and username, and completed Web-based screening questionnaires. Those who passed the Web-based screening criteria were assessed via telephone to confirm whether they met DSM-IV-TR diagnostic criteria [25] for MDD according to the Mini International Neuropsychiatric Interview (MINI) Version 5.0.0 [26]. There were no face-to-face components of the study.

Interventions

Internet Cognitive Behavioral Therapy (CBT)

The iCBT Program is described in detail elsewhere [24,27] and a demo can be accessed by contacting the corresponding author. In brief, participants completed 6 automated cartoon-style Web-based lessons teaching CBT skills (eg, behavioral activation) over 10 weeks, with a minimum wait-time of 5 days between lessons. Participants downloaded a "homework" document which included practical assignments (eg, thought monitoring) after each lesson, and had access to extra resources, frequently asked questions, and recovery stories of former participants. Automated reminder emails were also sent to participants when lessons became available. Participants in the iCBT group were able to continue to receive usual care from their health services during the intervention period.

Treatment as Usual (TAU) Control Group

Participants in the TAU control condition waited 10 weeks to gain access to the iCBT program. During the waiting period, they were able to continue to receive usual care from their health services.

Clinician-Guidance

Minimal clinician-assistance was provided to encourage adherence and engagement with the program by trained clinical psychologists with either Masters (LR) or PhD-level (JN) qualifications or psychiatry registrars (TM). Clinicians contacted the patient after lesson 1 and lesson 2 by email or phone to encourage progress. During the remainder of the program, clinician contact was made primarily by email, but if clinically indicated, or if patients' Kessler 10-item psychological distress scale (K-10) or PHQ-9 scores deteriorated significantly, telephone contact was made by the clinician.

Power Calculations

With a sample size of 40 participants per group, the study was powered (0.8 power) to detect a medium between-group difference of 0.65 on the primary depression measure at posttreatment (Cronbach alpha set at .05).

Statistical Analyses

All analyses were undertaken in Statistical Package for the Social Sciences (SPSS) version 23 (IBM Corp, 2014). ITT linear mixed models analyses were used to account for missing data due to participant dropouts. This approach is appropriate for RCTs with multiple time points [28] and does not assume that the last measurement was stable (an assumption of the the last observation carried forward approach [29]). Linear mixed models were conducted separately for each of the dependent variable (DV) measures, with time, treatment group, and the time by group interaction entered as fixed factors in the model, with a random intercept for subject. For each outcome, an identity covariance structure was specified to model the

XSL•FO
RenderX

covariance structure of the random intercept. Initial model building focused on the selection of the most appropriate covariance structure for the residual correlation matrix. Model fit indices and inspection of the variance-covariance matrix supported the selection of the identity covariance structure for each of the outcome measures. The fixed effect of age was added to each of the models. For each outcome measure except for problem areas in diabetes (PAID) scores, the fixed effect of age was not statistically significant and was removed from the model. Chi-square difference testing of the −2 log-likelihoods indicated that the removal of these fixed effects did not decrease model fit for any of the outcome variables, and they were excluded from further analyses.

For each group, planned contrasts were used to compare changes within and between groups from baseline to posttreatment (and 3-month follow-up for the iCBT group only). Between-group effect sizes using the pooled standard deviation and adjusted for sample size (Hedges *g*) were calculated to compare between groups at posttreatment. Within-group effect sizes (Cohen *d*) were calculated between pre- and posttreatment for both groups, and between pre- and 3-month follow-up for the iCBT group only. Effect sizes of 0.2, 0.5, and 0.8 were considered to be small, moderate, and large respectively in line with Cohen recommendations [30]. To investigate whether there were changes between posttreatment and follow-up for the iCBT group (n=21) for each dependent variable (eg, PHQ-9 scores), linear mixed models were conducted with time entered as a fixed factor and subject as a random intercept.

Reliable Change

Reliable change index (RCI) values [31] were calculated for the PHQ-9 scores to determine the proportion of each group who evidenced reliable improvements (or deterioration) between baseline and posttreatment RCI values were calculated using test-retest reliability values of .84 from Kroenke et al (2001). In order to calculate standard error of measurement values, standard deviations were derived from current sample (PHQ-9 pretreatment pooled, SD 5.34). We compared the demographic and clinical characteristics of individuals in the iCBT group who were considered responders (reliable improvements in PHQ-9 scores) and nonresponders (no reliable change in PHQ-9 scores) using independent-samples *t* tests and chi-square for categorical data.

Completers Versus Noncompleters

We also sought to compare the baseline demographic and clinical characteristics of individuals in the iCBT group who completed the entire program versus those who did not complete the program, using independent-samples *t* test and chi-square for categorical data.

Measurements

All measures were Web-based self-report questionnaires, with the exception of the MINI diagnostic interview which was administered by telephone.

Baseline Measures

At baseline, sociodemographic and sample characteristics were assessed including DM- and depression-related illness and treatment history.

Primary Outcome Measures

The primary outcome measure was the PHQ-9 [32], which is a validated 9-item self-report measure of depression symptom severity over the past 2 weeks. The PHQ-9 contains items answered on a 4-point Likert scale; the total score ranges between 0 and 27 [33]. The PHQ-9 has been validated in diabetes samples [34].

Glycemic control was measured via self-reported HbA1c values. Diabetes-related distress was measured using the PAID [35] questionnaire, a well-validated 20-item measure with a 5-point Likert scale; total scores are multiplied by 1.25 and range from 0 to 100 (with higher scores indicating greater emotional distress). The PAID has demonstrated sensitivity to change [36] and good internal and test-retest reliability [35].

Secondary Outcome Measures

The secondary outcome measures included the K-10 [37] for psychological distress; short form 12-item (SF-12) scale to measure of mental well-being (SF-12 MCS) and physical well-being (SF-12 PCS) [38], the generalized anxiety disorder 7-item (GAD-7) [39] for anxiety severity; and the PHQ-15—physical symptoms module for somatic symptom severity [40]. Other measures were also administered as part of the trial, which will be reported elsewhere. These include the relationships questionnaire (to assess attachment style), the alcohol and eating modules of the PHQ, and the fantastic checklist to assess specific lifestyle behaviours such as smoking status and alcohol use.

Treatment Expectancy, Acceptability, and Satisfaction

Participants rated their expectancy of benefit from the intervention at baseline, and the acceptability and satisfaction with the program at posttreatment using the treatment credibility or expectancy questionnaire (CEQ) [41].

Diagnostic Status

Current DSM-IV MDD diagnosis was assessed with the MINI version 5.0.0 [26] at baseline for both groups, and 3-month follow-up for the iCBT group only. The MINI possesses excellent interrater reliability (*k*=0.88-1.00) and good concurrent validity with the Composite International Diagnostic Interview (CIDI, World Health Organization, 1990) [42].

Measurement Time-Points

Outcomes were assessed at 3 time points: baseline, posttreatment (11 weeks), and 3-month follow-up (for the iCBT group only, as the TAU group received the iCBT program after they completed the posttreatment assessment). The primary outcome measures (PHQ-9, PAID) were also administered at the mid-time point (5 weeks). Finally, the K-10 was administered prior to each lesson in the iCBT group to monitor distress.

XSL•FO
RenderX

Results

Participant Flow

Of the 334 individuals who started a Web-based application, 185 were eligible for phone interview. After phone interview, 106 individuals met the inclusion criteria and were randomized to either iCBT (n=49) or TAU (n=57). Of these participants, 42/49 allocated to the iCBT group and 49/57 allocated to the TAU group completed baseline assessment and were included in the ITT analysis. At posttreatment assessment, 31/49 provided data in the iCBT group and 46/49 provided data in the TAU group. At 3-month follow-up, 21 participants completed the questionnaires and 31 completed the diagnostic interview to assess MDD. See Figure 1 for study flow diagram.

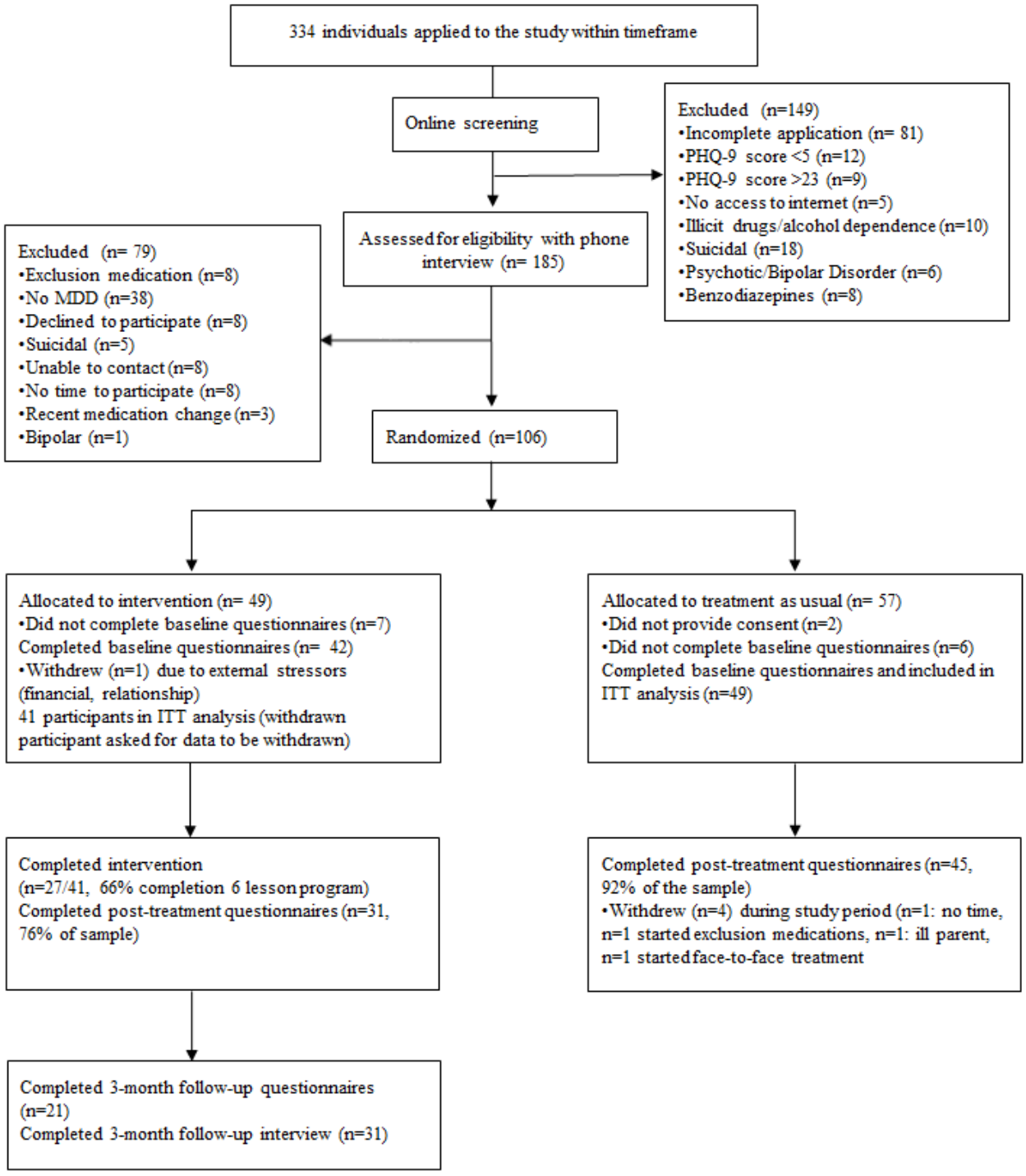
Baseline Characteristics

A total of 52 participants (57% of the sample) had T1 DM. Participants were 47 years on average (SD 12.61, range 20-71), and the majority were female (71%, 64/90) and married or living in a defacto relationship (55/90, 61%). Education status was mixed: one-fifth (16/90, 18%) had not completed high school, whereas 29 completed tertiary education (32%, 29/90). The majority were in full-time or part-time paid work (50/90, 56%), with 12 on the disability support pension (12/90, 6%). Only 16% (14/90) of the total sample were receiving psychological therapy, and 40% were taking medications for depression at baseline (37/90, 41%; see Table 1 for sample characteristics). Participants' baseline depression levels were moderate to severe on the PHQ-9 (mean 15.0, SD 5.3). The majority reported 3 or more episodes of depression (80/90, 89%); two-thirds of the sample had not been depression free for at least 2 years (57/90, 63%), and half (46/90, 51%) reported being depressed for more than 4 years during their lifetime.

Almost three quarters of the sample (65/90, 72%) reported chronic disease comorbidities, with the average being 1.25 comorbid conditions (SD 1.07, range 0-4). Circulatory conditions were the most common (43/90, 48%), followed by arthritis conditions (30/90, 37%). It was found that 40% of the sample (n=37) reported having at least one DM-related complication, with eyesight problems the most common (n=14 individuals (16%, 14/90) followed by nerve damage (10/90, 11%).

XSL•FO
RenderX

Figure 1. Consolidated standards of reporting trials (CONSORT) participant flow diagram. ITT: intention-to-treat; PHQ-9: patient health questionnaire-9.



Baseline Between-Group Comparisons

There were no significant differences between the groups on baseline scores, the self-report measures (PHQ-9, PAID, K-10), and demographic variables (eg, DM type, gender, education, employment, depression onset, comorbid chronic conditions). The TAU group were older on average than the iCBT group (t_{88}=2.22, P=.029).

Adherence

It was found that 27 out of 41 individuals in the iCBT group completed all 6 lessons of the program, resulting in a 66% adherence rate. Of the noncompleters, 4 participants completed 1 lesson only, 3 completed 2 lessons, 2 completed 4 lessons, and 3 completed 5 lessons.

Expectancy of Benefit

Prior to lesson 1, participants in the iCBT group were asked to provide a rating ranging from 1 to 9 about how logical the

XSL•FO
RenderX

therapy offered to them seemed (where 1=not at all, 9=very logical), and how useful they thought the treatment would be in reducing their symptoms of depression (where 1=not at all, 9=very useful). The scores on these items were summed to derive an "expectancy of benefit" rating. On average, scores in the treatment group were positive (mean 11.27, SD 3.89, range 2-18).

Primary Outcome Measures at Posttreatment and Effect Sizes

Table 2 includes estimated marginal means and linear mixed model results, and effect sizes for each of the outcome measures at baseline, midtreatment, and posttreatment. See Figure 2, PHQ-9 and PAID results. There were significant group by time interactions for the PHQ-9 ($F_{3,122.84}$=10.41, P<.001), PAID ($F_{3,123.00}$=10.32, P=.01), K-10 ($F_{3,153.20}$=21.86, P<.001), SF-12 MCS scores ($F_{1,87.40}$=9.07, P=.01), and GAD-7 scores ($F_{1,81.40}$=13.18, P<.001). In contrast, the group by time interactions were not significant for SF-12 PCS ($F_{1,78.92}$=0.27, P=.60), HbA1c levels ($F_{1,74.23}$=0.11, P=.75), or PHQ-15 scores ($F_{1,86.59}$=1.84, P=.18).

Within-Group Effect Sizes (Baseline to Posttreatment)

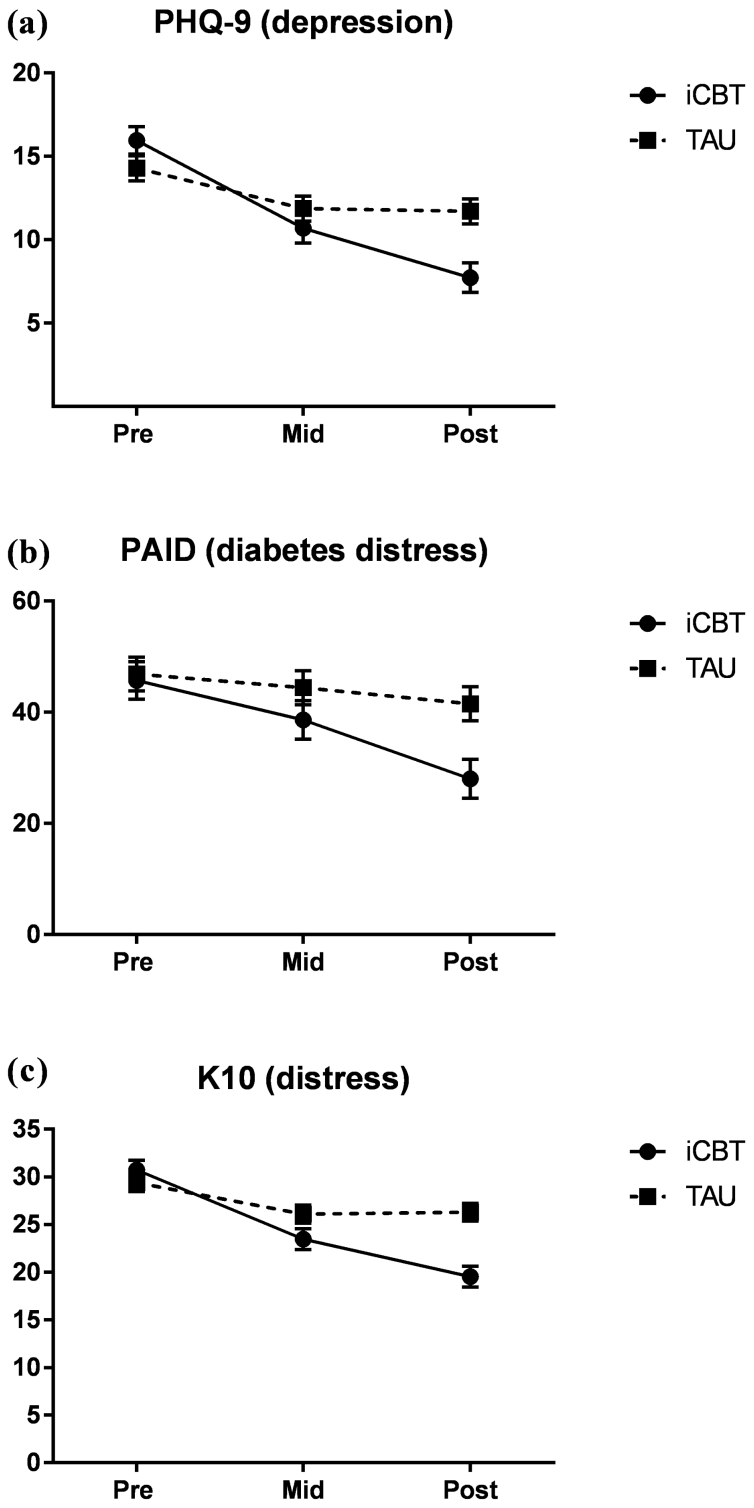
Within-group comparisons for the iCBT group revealed large effect sizes between pretreatment and posttreatment on the PHQ-9 (d=1.90), PAID (d=1.18), K-10 (d=2.59), SF-12 mental well-being subscale (d=−0.85), and GAD-7 (d=1.38, 95% CI 0.86-1.90), and moderate for somatic symptom severity on the PHQ-15 (d=0.63). The changes on the SF12 physical well-being subscale and HbA1c levels were not significant (P>.05, see Table 2). The TAU group demonstrated medium statistically significant reductions in PHQ-9 (d=0.53), GAD-7 (d=0.42), and the K-10 (d=0.41), although the changes on the remaining outcome measures were not significant.

Between-Group Effect Sizes (Posttreatment)

Posttreatment scores were significantly lower in the iCBT group relative to TAU on the PHQ-9, PAID, K-10, GAD-7, and SF-12 mental well-being subscales with moderate between-groups effect sizes on the PHQ-9 (g=0.78, 95% CI 0.30-1.25), GAD-7 scores (g=0.72, 95% CI 0.25-1.19), and SF-12 MCS scores (g=−0.66, 95% CI −1.12 to −0.19). In addition, we found large between-group differences for PAID scores (g=0.80, 95% CI 0.32-1.27) and K-10 (g=1.06, 95% CI 0.57-1.54). The between-group effect sizes for the other measures were small and not-significant (PHQ-15: g=0.21, 95% CI −0.24 to 0.67; HbA1c levels: g= − 0.14, 95% CI −0.62 to 0.34; SF-12 PCS: g=0.15, 95% CI −0.30 to 0.60).

XSL•FO
RenderX

Figure 2. Estimated marginal means and standard errors for Internet-based cognitive behavioral therapy (iCBT) and treatment-as-usual (TAU) control group for (a) patient health questionnaire-9 (PHQ-9) (depression), (b) problem areas in diabetes (PAID) (diabetes-related distress), and (c) Kessler 10-item psychological distress scale (K-10) (general distress) at baseline, midtreatment, and posttreatment.



XSL•FO
RenderX

Table 1. Baseline demographics and sample characteristics for the Internet-based cognitive behavioral therapy (iCBT) and treatment as usual (TAU) groups.

Variable	iCBT[a]group (n=41)	TAU[b]group (n=49)	Total (N=90)
Diabetes type, n (%)			
Type 1	24 (59)	28 (57)	52 (58)
Type 2	17 (42)	21 (43)	38 (42)
Diabetes treatment, n (%)			
Insulin	29 (71)	34 (69)	63 (70)
Diet alone	1 (2)	2 (4)	3 (3)
Tablets	8 (20)	11 (22)	19 (21)
Other	3 (7)	2 (4)	5 (6)
Age of onset (diabetes), mean (SD)	27.6 (18.1)	32.9 (16.2)	30.5 (17.2)
Diabetes complications (total), mean (SD)	0.83 (1.24)	1.15 (1.9)	1.0 (1.63)
Age (years), mean (SD)[c]	43.5 (13.3)	49.3 (11.5)	46.7 (12.6)
Baseline K-10[d], mean (SD)	30.7 (5.9)	29.4 (6.7)	30.0 (6.4)
Baseline PHQ-9[e], mean (SD)	15.95 (5.1)	14.3 (5.5)	15.0 (5.3)
Baseline PAID[f], mean (SD)	38.1 (16.1)	36.5 (18.6)	37.2 (17.5)
Gender, n (%)			
Male	8 (20)	18 (37)	26 (29)
Female	33 (81)	31 (63)	64 (71)
Marital status, n (%)			
Single or never married	9 (22)	10 (20)	19 (21)
Married or defacto	29 (71)	26 (53)	55 (61)
Separated or divorced or widowed	3 (7)	13 (27)	16 (18)
Educational status, n (%)[i]			
Less than high school	6 (15)	10 (20)	16 (18)
High school	3 (7)	5 (10)	8 (9)
Tertiary (diploma)	6 (15)	8 (16)	14 (16)
Tertiary (university degree)	11 (27)	10 (20)	21 (23)
Tertiary (postgraduate degree)	5 (12)	3 (6)	8 (9)
Other certificate	6 (15)	6 (12)	12 (13)
Trade certificate	4 (10)	7 (14)	11 (12)
Employment status, n (%)			
Full-time paid work	17 (42)	16 (33)	33 (37)
Part-time paid work	5 (12)	12 (25)	17 (19)
Unemployed	3 (7)	4 (8)	7 (13)
Student	3 (7)	1 (2)	4 (4)
Retired	5 (12)	0 (0)	12 (13)
Disability support	5 (12)	7 (14)	12 (6)
At home parent	3 (7)	2 (4)	5 (8)
Age of onset (depression), n (%)			
Under 12	3 (7)	5 (10)	8 (9)

XSL•FO
RenderX

Variable	iCBT[a]group (n=41)	TAU[b]group (n=49)	Total (N=90)
13-21 years	14 (34)	15 (31)	29 (32)
22 years or older	23 (56)	29 (59)	52 (58)
Number of episodes (depression), n (%)			
1-2 episodes	5 (12)	5 (10)	10 (11)
3-4 episodes	10 (24)	18 (37)	28 (31)
5-8 episodes	10 (24)	6 (12)	16 (18)
More than 8 episodes	16 (39)	20 (42)	36 (40)
Total duration (depression) during lifetime, n (%)			
Less than 1 month	2 (5)	1 (2)	3 (3)
1-12 months	10 (24)	9 (18)	19 (21)
1-4 years	8 (20)	14 (29)	22 (24)
More than 4 years	21 (51)	25 (51)	46 (51)
Free of depression past 2 years, n (%)			
No	28 (68)	29 (59)	57 (63)
Yes	13 (32)	20 (41)	33 (37)
Comorbid chronic conditions, n (%)			
Asthma	13 (32)	15 (31)	28 (31)
Cancer	2 (5)	4 (8)	6 (7)
Stroke	2 (5)	1 (2)	3 (3)
Circulatory condition	17 (42)	26 (53)	43 (48)
Gout or rheumatism or arthritis	18 (44)	15 (31)	33 (37)
Current psychotherapy (psychology, social work, or counseling), n (%)	7 (17)	7 (15)	14 (16)
Current medication, n (%)	13 (32)	24 (50)	37 (41)
Current medication (class), n (%)			
SSRI[g]	4 (10)	16 (33)	20 (22)
SNRI[h]	8 (20)	8 (16)	16 (18)
Other	1 (2)	0 (0)	1 (1)

[a]iCBT: Internet-based cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]Mean (SD): significant between-groups difference at $P<.05$ level.

[d]K-10: Kessler 10-item psychological distress scale.

[e]PHQ-9: patient health questionnaire-9.

[f]PAID: Problem areas in diabetes.

[g]SSRI: selective serotonin reuptake inhibitor.

[h]SNRI: selective noradrenaline reuptake inhibitor.

[i]Educational status: refers to the highest level of education received.

XSL•FO
RenderX

Table 2. Estimated marginal means (standard deviations) for primary and secondary outcome measures, within-group effect sizes, and between-group effect sizes.

Measure	Baseline Mean[a] (SD)	Mid Mean[b] (SD)	Post Mean[c] (SD)	Within *t* (df) Baseline, Post	Within ES[d] (95% CI) Baseline, Post	Between ES[e] (95% CI) Post	*F* (time by group)
PHQ-9[f] iCBT[g]	15.95 (5.25)	10.69 (4.98)	7.72 (4.96)	10.55 (159.32)	1.90 (1.34-2.45)	0.78 (0.30-1.25)	$F_{3,122.84}$=10.41, *P*<.001
PHQ-9 TAU[h]	14.29 (5.25)	11.88 (5.15)	11.70 (5.15)	3.91 (152.40)	0.53 (0.12-0.94)	-	-
HbA1c[f][i] iCBT	7.87 (1.79) or 63 mmol/mol	-	7.98 (1.73) or 64 mmol/mol	−0.48 (75.48)	−0.03 (−0.52 to 0.45)	−0.14 (−0.62 to 0.34), ns[j]	$F_{1,74.23}$=0.11 *P*=.75, ns
HbA1c TAU	7.72 (1.82) or 61 mmol/mol	-	7.73 (1.76) or 61 mmol/mol	−0.05 (72.47)	-0.01 (−0.43 to 0.41)	-	-
PAID[k] iCBT	45.73 (21.45)	38.65 (19.74)	28.00 (19.65)	7.74 (155.89)	1.18 (0.68-1.69)	0.80 (0.32-1.27)	$F_{3,87.03}$=10.32, *P*=.01
PAID TAU	46.87 (21.28)	44.43 (20.82)	41.55 (20.82)	2.83 (152.09)	0.24 (−0.16 to 0.65), ns	-	-
K-10[l] iCBT	30.73 (6.47)	23.46 (6.17)	19.54 (6.12)	12.03 (157.66)	2.59 (1.96-3.22)	1.06 (0.57-1.54)	$F_{2,155.20}$=21.86, *P*<.001
K-10 TAU	29.39 (6.51)	26.10 (6.38)	26.28 (6.44)	3.99 (151.50)	0.41 (0.00-0.81)	-	-
SF-12 MCS[m] iCBT	30.22 (10.12)	-	39.26 (9.80)	−5.38 (84.04)	−0.85 (−1.34 to −0.34)	−0.66 (−1.12 to −0.19)	$F_{1,87.40}$=9.07, *P*=.01
SF-12 MCS TAU	29.79 (10.01)	-	32.70 (9.97)	−2.09 (77.55)	−0.27 (−0.68 to 0.13), ns	-	-
SF-12 PCS[n] iCBT	40.15 (12.49)	-	41.1 (11.64)	−0.59 (80.95)	−0.08 (−0.55 to 0.38), ns	0.15 (−0.30 to 0.60), ns	$F_{1,78.92}$=0.27, *P*=.60, ns
SF-12 PCS TAU	42.94 (11.27)	-	42.8 (11.12)	0.11 (75.97)	−0.01 (−0.39 to 0.41), ns	-	
GAD-7[o] iCBT	10.73 (4.80)	-	4.60 (4.62)	7.21 (84.11)	1.38 (0.86-1.90)	0.72 (0.25-1.19)	$F_{1,81.40}$=13.18, *P*<.001
GAD-7 TAU	10.11 (4.76)	-	8.02 (4.75)	2.97 (77.74)	0.42 (0.02-0.83)	-	
PHQ-15[p] iCBT	11.90 (4.61)	-	8.65 (4.57)	3.23 (90.24)	0.63 (0.15-1.11)	0.21 (−0.24 to 0.67), ns	$F_{1,86.59}$=1.84, *P*=.18, ns
PHQ-15 TAU	11.10 (4.55)	-	9.65 (4.68)	1.68 (81.88)	0.33 (−0.08 to 0.73), ns		

[a]Baseline: iCBT: n=41, TAU: n=49.

[b]Mid-treatment: iCBT: n=32, TAU: n=46.

[c]Posttreatment: iCBT: n=31, TAU: n=46.

[d]Within-group ES=Cohen *d*.

[e]Between-group ES=Hedges *g* with Hedges pooled SD.

[f]PHQ-9: patient health questionnaire-9.

[g]iCBT: Internet-based cognitive behavioral therapy group.

[h]TAU: treatment as usual group.

[i]HbA1c: haemoglobin A1c.

XSL•FO
RenderX

[j]ns: not significant, *P*>.05.

[k]PAID: problem areas in diabetes scale.

[l]K-10: Kessler 10-item psychological distress scale.

[m]SF-12 MCS: short form 12-item mental health subscale.

[n]SF-12 PCS: short form 12-item physical health subscale.

[o]GAD-7: generalized anxiety disorder 7-item scale.

[p]PHQ-15: patient health questionnaire 15-item somatization scale.

Reliable Change

Of the iCBT group, 21/31 (51%) reliably improved compared with 9 (18%) in the TAU group. Of the iCBT group, no participants evidenced reliable deterioration, compared with only 2 participants in the TAU group (4%). The difference in proportions of participants who evidenced reliable change was significant (χ^2_2=18.4, *P*<.001). There were no significant differences between iCBT participants who showed reliable change versus those who did not show reliable change on any of the clinical or demographic variables at baseline (*t*=0.46-1.79, *P*>.08).

Completers Versus Noncompleters in the Internet-Based Cognitive Behavioral Therapy (iCBT) Group

We compared baseline demographic and clinical characteristics of participants who completed all 6 lessons versus those who completed fewer than 6 lessons to explore whether there were any key differences at baseline in completers versus noncompleters. There were no significant differences on any of the clinical or demographic variables, including expectancy ratings at baseline. However, there was a trend toward higher K-10 distress scores in completers compared with noncompleters (K-10: completer sample (n=21): mean 31.96, SD 5.61, noncompleter sample (n=14): mean 28.36, SD 5.98, t_{39}=1.91, *P*=.06, *g*=0.61, 95% CI 0.08-1.30).

Outcomes for iCBT Group Between Posttreatment and 3-Months Follow-Up

There were no statistically significant effects of time for any of the outcome measures between posttreatment and follow-up (n=19). The within-group effect sizes were small and not significant (see Table 3).

Diagnostic Status at Follow-Up

Of the total 31 participants in the iCBT group who completed a diagnostic interview to assess for MDD, 27 (87.1%) no longer met criteria for MDD.

Clinician Time

The clinician spent on average 27.3 min per participant on email and telephone contact in the iCBT group (SD 14.9, range 8-71 min) over the course of the program. The clinician spent on average 13.51 min (SD 13.74, range 1-69 min) on the control group. This difference was significant (t_{88}=4.55, *P*<.001).

Patient Satisfaction

The iCBT participants were asked to provide a rating about: (1) how satisfied they were that the program taught them the skills to manage depression and (2) their confidence in recommending the program to a friend with similar problems (where 1=not at all, 5=somewhat, and 9=very). The overall mean scores were acceptable (satisfaction: mean 6.06, SD 2.14; recommend to friend (mean 6.84, SD 2.20, range 1-9). The majority of participants reported feeling somewhat to very satisfied with the program (n=27; 85%), although only 3 of these (9% of the sample collected at posttreatment) were “very” satisfied. The majority of participants reported feeling somewhat to very confident in recommending the program to a friend (n=28; 88%); of these 9 (28%) reported feeling very confident in recommending this program to a friend.

Discussion

Principal Findings

Our RCT in adults with T1 or T2 DM and MDD aimed to test whether a generic iCBT program for depression was more effective than usual care in improving depression. This is the first RCT to show that a generic iCBT program for the treatment of depression was superior to TAU in reducing depression, diabetes-related distress, anxiety, general distress, and improving mental well-being for people with comorbid T1DM or T2DM and MDD. On the primary outcome measure at posttreatment (PHQ-9), between-group effect sizes were moderate (*d*=0.78), and we also found large between-group differences at posttreatment on measures of diabetes-related distress (*g*=0.80) and general distress on the K-10 (*g*=1.06), and moderate between-groups effect sizes for generalized anxiety (*g*=0.72) and mental well-being (*g*=−0.66). These findings support the use of iCBT in the treatment of depression in DM.

Patients in the iCBT group experienced, on average, large reductions in depressive symptoms between baseline and posttreatment (ES=1.90). Benchmarked against previous studies of iCBT for depressive and anxiety disorders in the general population, these effect sizes are larger than those seen in patients who do not have chronic physical diseases [43]. At posttreatment, 51% of the iCBT group showed evidence of statistically reliable change compared with only 18% in the TAU group. There was minimal evidence of relapse between posttreatment and follow-up in the iCBT group, with 87% of the sample interviewed no longer meeting diagnostic criteria for MDD at 3-months follow-up. These findings are consistent with previous trials showing that iCBT leads to sustained improvements in depression symptoms beyond the completion of treatment, and even at long-term follow-up [44].

XSL•FO
RenderX

Table 3. Estimated marginal means (standard deviations) on primary and secondary outcomes between posttreatment and 3-month follow-up for the Internet-delivered cognitive behavioral therapy group.

Measure	Post mean (SD) T3[j] (n=30)	3-month follow-up mean (SD) T4[k] (n=19)	Within-group *t* (df) T3, T4	*r*	Within effect size[l] (95% CI) T3, T4
PHQ-9[a]	10.09 (3.56)	10.98 (4.49)	−0.91 (27.53)	.68	−0.21 (−0.78 to 0.37), ns[b]
HbA1c[c]	7.84 (1.06)	7.71 (1.18)	0.48 (22.06)	.72	0.04 (−0.54 to 0.62), ns
PAID[d]	35.88 (14.19)	33.10 (16.17)	0.87 (22.18)	.76	0.17 (−0.41 to 0.74), ns
K-10[e]	23.69 (4.55)	23.20 (5.23)	0.47 (24.55)	.79	0.13 (−0.45 to 0.70), ns
GAD-7[f]	6.75 (2.85)	7.88 (3.57)	−1.45 (27.89)	.67	−0.34 (−0.93 to 0.22), ns
SF-12 MCS[g]	35.37 (7.07)	37.23 (10.11)	−0.77 (43.39)	.46	−0.16 (−0.73 to 0.42), ns
SF-12 PCS[h]	41.12 (7.01)	41.92 (8.24)	−0.47 (26.26)	.81	−0.09 (−0.66 to 0.49), ns
PHQ-15[i]	9.24 (2.79)	10.10 (3.66)	(22.49)	.86	−0.26 (−0.84-0.32), ns

[a]PHQ-9: patient health questionnaire-9.

[b]ns: not significant (*P*>.05).

[c]HbA1c: haemoglobin A1c.

[d]PAID: problem areas in diabetes.

[e]K-10: Kessler 10-item psychological distress scale.

[f]GAD-7: generalized anxiety disorder 7-item scale.

[g]SF-12 MCS: short form 12-item mental health subscale.

[h]SF-12 PCS: short form 12-item physical health subscale.

[i]PHQ-15: patient health questionnaire 15-item somatisation scale.

[j]T3: posttreatment.

[k]T4: 3-month follow-up.

[l]within-group ES=Hedges *g*.

Despite finding significant improvements in mental health and well-being, the positive effects of this program did not appear to translate to improved physical health outcomes or well-being. We failed to find a difference at posttreatment on the physical well-being subscale of the SF-12, somatic symptom severity on the PHQ-15, and self-reported HbA1c levels, and there was no evidence of improvement in self-reported physical health outcomes between posttreatment and 3-months follow-up in the iCBT group. Although these results may be due to lack of power, these preliminary findings suggest that although iCBT for depression improved mental health outcomes, at least in the short-term, it did not improve physical health outcomes. The relatively short follow-up period of this study is likely to have precluded us from finding a positive effect on physical health outcomes and self-management behaviors, which may only be observable at long-term follow-up [45].

This sample had poor health status, with 40% reporting one or more DM-related complications and 75% of the sample reporting one or more comorbid chronic diseases alongside DM. Despite living with severe and complicated physical health problems, participants in the iCBT group still reported better mental health at the end of treatment. Going forward, these improvements in mental health may help them to better cope with the challenges of managing diabetes and adjusting to changes in health by improving their resilience, or facilitating the use of positive coping strategies and self-care behaviors [46]. Further research is now needed to investigate whether improved physical health outcomes are observed at longer-term follow-up without any further intervention. We also recommend that future research focus on the mechanisms that underlie the positive changes we have observed in people with T1 or T2 DM and comorbid depression to better inform the development and tailoring of future treatments.

Overall, these findings are consistent with other studies showing that CBT for depression leads to improved depression outcomes, but has a limited effect on HbA1c levels [47]. The future challenge in this area of research is to identify treatment components and interventions that continue to positively impact on depression symptoms, but in addition derive beneficial changes in physical health outcomes, self-management, and biomarkers such as HbA1c levels. Diabetes distress is a broad-ranging construct covering many domains associated with living with diabetes, from adjustment to a new diagnosis through to the ongoing burden and burnout of managing a chronic condition. Although we noted large and significant reductions in diabetes distress as a consequence of our

XSL•FO
RenderX

depression intervention, perhaps if diabetes distress were to be specifically targeted as part of the treatment, we would see broader improvements and even more significant change on this variable, with greater potential to impact on self-management behavior and physical outcomes. A multidisciplinary, multicomponent, Web-based treatment which addresses depression symptoms as well as the way that the emotional effects of living with diabetes can impact on self-care and self-management might be the key to seeing improvements in both mental and physical health.

We found only 66% of the participants completed the program and there was significant loss of data at follow-up (25% did not complete posttreatment assessments). Whereas these adherence rates are consistent with previous studies of Web-based depression management programs that incorporated diabetes-specific content in DM populations (eg, 62% adherence in 22), they are lower than rates of up to 80-90% found in previous trials of iCBT for depression and anxiety using the same protocols for providing clinician guidance to participants [48]. We failed to find any consistent differences between completers and noncompleters of the program in demographic characteristics, or baseline depression, anxiety, and diabetes distress severity, which may have been due to lack of power to detect key differences between these groups. Preliminary results suggested a trend toward higher distress scores in those who completed the program, although these findings need to be replicated in a larger sample.

Future research needs to determine the most effective ways to engage people with comorbid DM and depression in iCBT programs and examine the possible reasons for these lower adherence rates. First, it is possible that the program was not as acceptable for the participants who dropped out of the program, or may not have adequately addressed their unique concerns or difficulties that impacted on their depression symptoms. Although the general CBT skills may be useful and effective for improving depression for some individuals with DM, tailoring content to DM-related may improve the acceptability of the program for those who dropped out, and therefore improve engagement and completion rates. Second, in this study, participants had minimal guidance and monitoring from a therapist, but it is possible that more frequent or intensive guidance was needed to better engage the individuals who dropped out of the iCBT program. Therapist guidance during iCBT has been shown to promote adherence and influence program completion rates [49], although it is unclear how much, and what type of support is needed to achieve optimal adherence to iCBT interventions. Future research could benefit from examining differing levels of support on adherence. Finally, it is possible that the relatively low adherence rates in this study were reflective of broader issues with adherence in depressed DM populations [50]. Treatment adherence is an important consideration in diabetes self-management, with adherence to lifestyle, medication, and self-monitoring interventions critical to prevent the development of long-term complications, yet depressed individuals with DM have been shown to have poorer adherence to treatments in general [5].

Limitations

The findings should be interpreted in the context of its limitations. Due to the fact that we do not have follow-up data in the TAU group, more research is needed to confirm the efficacy of iCBT for depression in people with DM in the long-term (eg, 6-12 months following treatment). Follow-up diagnostic interviews were not blinded to treatment allocation, introducing bias. The use of self-reported HbA1c levels to assess glycemic control was also a limitation of the study and we did not assess when the participants had their HbA1c levels tested. We chose this measure to mirror what occurs in usual clinical care because we wanted to increase the coverage of recruitment across Australia, and it was impractical to collect blood samples in other Australian states and territories. Finally, the high proportion of females and people with T1 DM may influence the generalizability of the findings.

Conclusions

An unmodified Internet-delivered CBT program for depression is an effective intervention to improve depression and comorbid anxiety, general distress, and diabetes-specific distress in individuals with DM. Further research comparing unmodified versus tailored approaches for addressing depression in DM will clarify whether tailored approaches are more acceptable, engaging and therefore lead to better adherence rates. Given the increasing availability of evidence-based generic iCBT programs for depression in the general community and their ability to be implemented in routine care, they should be considered as a treatment option for those suffering from comorbid depression and DM. iCBT programs represent a scalable, accessible, evidence-based treatment option for people looking for effective treatment for their depression, who either are unable or do not wish to seek face-to-face psychological treatment.

Acknowledgments

This research was conducted by the authors as part of their employment with St Vincent's Hospital Australia. We would like to acknowledge the support of St Vincent's Hospital Australia. We also thank all the participants who took part in this research. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Dr Newby is supported by an Australian NHMRC Early Career Australian Clinical Fellowship (#1037797). Lisa Robins is supported by a grant from the St Vincent's Clinic Foundation and an APA scholarship. Prof Kay Wilhelm and Dr Jill Newby take responsibility for the contents of this article.

XSL•FO
RenderX

Authors' Contributions

JN and GA conceived of the study. JN, KW, LR, and GA initiated the study design and JN, TF, JS, AF, TM, KW, and LR helped with implementation. JN conducted the primary statistical analysis. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT E-Health Checklist.

[HTML File, 386KB - jmir_v19i5e157_app1.html]

References

1. International Diabetes Federation. IDF Diabetes Atlas: Seventh Edition. Brussels, Belgium: International Diabetes Federation; 2015:1-144.
2. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. Diabetes Care 2001 Jun;24(6):1069-1078. [Medline: 11375373]
3. Baumeister H, Hutter N, Bengel J, Härter M. M. Quality of life in medically ill persons with comorbid mental disorders: a systematic review and meta-analysis. Psychother Psychosom 2011;80(5):275-286. [doi: 10.1159/000323404]
4. Hermanns N, Kulzer B, Krichbaum M, Kubiak T, Haak T. How to screen for depression and emotional problems in patients with diabetes: comparison of screening characteristics of depression questionnaires, measurement of diabetes-specific emotional problems and standard clinical assessment. Diabetologia 2006 Mar;49(3):469-477. [doi: 10.1007/s00125-005-0094-2] [Medline: 16432706]
5. Gonzalez JS, Peyrot M, McCarl LA, Collins EM, Serpa L, Mimiaga MJ, et al. Depression and diabetes treatment nonadherence: a meta-analysis. Diabetes Care 2008;31(12):2398-2403. [doi: 10.2337/dc08-1341]
6. Egede LE, Ellis C. Diabetes and depression: global perspectives. Diabetes Res Clin Pract 2010 Mar;87(3):302-312. [doi: 10.1016/j.diabres.2010.01.024] [Medline: 20181405]
7. Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM. Depression and poor glycemic control: a meta-analytic review of the literature. Diabetes Care 2000;23(7):934-942.
8. Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary heart disease mortality among adults with and without diabetes. Diabetes Care 2005 Jun;28(6):1339-1345. [Medline: 15920049]
9. MacDonald G, Campbell L. Mental illness: the forgotten burden on diabetes populations? The Lancet 2016;388(10044):561-561. [doi: 10.1016/S0140-6736(16)31213-2]
10. Pouwer F, Tack CJ, Geelhoed-Duijvestijn PH, Bazelmans E, Beekman AT, Heine RJ, et al. Limited effect of screening for depression with written feedback in outpatients with diabetes mellitus: a randomised controlled trial. Diabetologia 2011;54(4):741-748. [doi: 10.1007/s00125-010-2033-0]
11. Atlantis E, Fahey P, Foster J. Collaborative care for comorbid depression and diabetes: a systematic review and meta-analysis. BMJ Open 2014;4:e004706. [doi: 10.1136/bmjopen-2013- 004706]
12. Baumeister H, Hutter N, Bengel J. Psychological and pharmacological interventions for depression in patients with diabetes mellitus: an abridged Cochrane review. Diabetic Med 2014;31(7):773-786. [doi: 10.1111/dme.12452]
13. National Collaborating Centre for Chronic Conditions (Great Britain), Royal College of Physicians of London. Type 2 Diabetes: National clinical guideline for management in primary secondary care (update). London, United Kingdom: Royal College of Physicians; 2008.
14. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016;21(5):290. [doi: 10.1002/14651858.CD011565.pub2]
15. Newby JM, Mewton L, Andrews G. Cognitive behaviour therapy for anxiety and depression in primary care. J Anxiety Disord 2017;46:25-35. [doi: 10.1016/j.janxdis.2016.06.002]
16. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012;12(6):745-764. [doi: 10.1586/erp.12.67]
17. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014;13(3):288-295. [doi: 10.1002/wps.20151]
18. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
19. Gøtzsche PC. Why I think antidepressants cause more harm than good. Lancet Psychiatry 2014 Jul;1(2):104-106. [doi: 10.1016/S2215-0366(14)70280-9] [Medline: 26360561]

XSL•FO
RenderX

20. Ebert D, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does Internet-based guided-self-help for depression cause harm? an individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol Med 2016 Oct;46(13):2679-2693. [doi: 10.1017/S0033291716001562] [Medline: 27649340]
21. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011;34(2):320-325. [doi: 10.2337/dc10-1248]
22. 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;38(5):776-783. [doi: 10.2337/dc14-1728]
23. Watts S. CBT for depression: A pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49. [doi: 10.1186/1471-244X-13-49]
24. Robins L, Newby J, Wilhelm K, Smith J, Ma T, Finch A, et al. Internet-delivered cognitive behaviour therapy for depression in people with diabetes: study protocol for a randomised controlled trial. BMJ Open Diabetes Res Care 2015;3(1):e000144. [doi: 10.1136/bmjdrc-2015-000144]
25. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association; 2000.
26. 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 [FREE Full text] [Medline: 9881538]
27. Perini S, Titov N, Andrews G. The climate sadness program of internet-based treatment for depression: a pilot study. E-J Appl Psychol 2008;4(2):18-24.
28. Salim A, Mackinnon A, Christensen H, Griffiths K. Comparison of data analysis strategies for intent-to-treat analysis in pre-test-post-test designs with substantial dropout rates. Psychiatry Res 2008 Sep 30;160(3):335-345. [doi: 10.1016/j.psychres.2007.08.005] [Medline: 18718673]
29. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
30. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition. New York, NY: Academic Press, Inc; 1988.
31. Jacobson N, 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]
32. 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]
33. Wittkampf K, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the patient health questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29(5):388-395. [doi: 10.1016/j.genhosppsych.2007.06.004] [Medline: 17888804]
34. van Steenbergen-Weijenburg KM, de VL, Ploeger RR, Brals JW, Vloedbeld MG, Veneman TF, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. BMC Health Serv Res 2010 Aug 12;10:235 [FREE Full text] [doi: 10.1186/1472-6963-10-235] [Medline: 20704720]
35. Welch GW, Jacobson AM, Polonsky WH. The problem areas in diabetes scale. An evaluation of its clinical utility. Diabetes Care 1997 May;20(5):760-766. [Medline: 9135939]
36. Welch G, Weinger K, Anderson B, Polonsky WH. Responsiveness of the problem areas in diabetes (PAID) questionnaire. Diabet Med 2003 Jan;20(1):69-72. [Medline: 12519323]
37. Kessler R, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [Medline: 12214795]
38. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]
39. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [Medline: 11914441]
41. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [Medline: 11132119]
42. Kessler RC, Ustün TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res 2004;13(2):93-121. [Medline: 15297906]
43. Perini S, Titov N, Andrews G. Clinician-assisted internet-based treatment is effective for depression: randomized controlled trial. Aust N Z J Psychiatry 2009 Jun;43(6):571-578. [doi: 10.1080/00048670902873722] [Medline: 19440890]
44. Andersson G, Hesser H, Hummerdal D, Bergman-Nordgren L, Carlbring P. A 3.5-year follow-up of Internet-delivered cognitive behavior therapy for major depression. J Ment Health 2013 Apr;22(2):155-164. [doi: 10.3109/09638237.2011.608747] [Medline: 21957933]

XSL•FO
RenderX

45. Lustman PJ, Griffith LS, Freedland KE, Kissel SS, Clouse RE. Cognitive behavior therapy for depression in type 2 diabetes mellitus. A randomized, controlled trial. Ann Intern Med 1998 Oct 15;129(8):613-621. [Medline: 9786808]
46. Yi JP, Vitaliano PP, Smith RE, Yi JC, Weinger K. The role of resilience on psychological adjustment and physical health in patients with diabetes. Br J Health Psychol 2008 May;13(Pt 2):311-325 [FREE Full text] [doi: 10.1348/135910707X186994] [Medline: 17535497]
47. Petrak F, Herpertz S, Albus C, Hermanns N, Hiemke C, Hiller W, et al. Cognitive behavioral therapy versus sertraline in patients with depression and poorly controlled diabetes: the diabetes and depression (DAD) study: a randomized controlled multicenter trial. Diabetes Care 2015 May;38(5):767-775. [doi: 10.2337/dc14-1599] [Medline: 25690005]
48. Newby J, Mackenzie A, Williams AD, McIntyre K, Watts S, Wong N, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. Psychol Med 2013;43(12):2635-2648. [doi: 10.1017/S0033291713000111] [Medline: 23419552]
49. Baumeister H, Reichlera L, Munzingera M, Lin J. The impact of guidance on Internet-based mental health interventions? A systematic review. Internet Interventions 2014;1(4):205-215.
50. Lin Elizabeth HB, Katon W, Von KM, Rutter C, Simon GE, Oliver M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. Diabetes Care 2004 Sep;27(9):2154-2160. [Medline: 15333477]

Abbreviations

CBT: cognitive behavioral therapy
CES-D: Center for Epidemiologic Studies Depression Scale
CEQ: credibility or expectancy questionnaire
CIDI: Composite International Diagnostic Interview
DM: diabetes mellitus
DV: dependent variable
GAD-7: generalized anxiety disorder 7-item
GP: general practitioner
iCBT: Internet-based cognitive behavioral therapy
K-10: Kessler 10-item psychological distress scale
MDD: major depressive disorder
MINI: Mini International Neuropsychiatric Interview
PAID: problem areas in diabetes
PHQ-9: patient health questionnaire-9
RCI: reliable change index
RCT: randomized controlled trial
SF-12: short form 12-item
SPSS: Statistical Package for the Social Sciences
SSRI: selective serotonin reuptake inhibitor
TAU: treatment as usual

Edited by G Eysenbach; submitted 09.01.17; peer-reviewed by M Rogers, A Cross; comments to author 08.02.17; revised version received 13.02.17; accepted 14.02.17; published 15.05.17.

Please cite as:
Newby J, Robins L, Wilhelm K, Smith J, Fletcher T, Gillis I, Ma T, Finch A, Campbell L, Andrews G
Web-Based Cognitive Behavior Therapy for Depression in People With Diabetes Mellitus: A Randomized Controlled Trial
J Med Internet Res 2017;19(5):e157
URL: http://www.jmir.org/2017/5/e157/
doi:10.2196/jmir.7274
PMID:28506956

©Jill Newby, Lisa Robins, Kay Wilhelm, Jessica Smith, Therese Fletcher, Inika Gillis, Trevor Ma, Adam Finch, Lesley Campbell, Gavin Andrews. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Support Interventions for the Self-Management of Low Back Pain: A Systematic Review

Barbara I Nicholl[1*], BSc, PGDip, PhD; Louise F Sandal[2*], BSc, MSc, PhD; Mette J Stochkendahl[2,3], MSc, PhD; Marianne McCallum[1], BMedSci, MBChB, MRCGP; Nithya Suresh[1], MBChB; Ottar Vasseljen[4], BSc, PhD; Jan Hartvigsen[2,3], DC, PhD; Paul J Mork[4], MSc, PhD; Per Kjaer[2], PT, MSc, PhD; Karen Søgaard[2], MSc, PhD; Frances S Mair[1], MBChB, MD, DRCOG, FRCGP

[1]General Practice & Primary Care, Institute of Health & Wellbeing, University of Glasgow, Glasgow, United Kingdom

[2]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

[3]Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark

[4]Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[*]these authors contributed equally

Corresponding Author:
Barbara I Nicholl, BSc, PGDip, PhD
General Practice & Primary Care
Institute of Health & Wellbeing
University of Glasgow
1 Horselethill Road
Glasgow, G12 9LX
United Kingdom
Phone: 44 141 3308327
Fax: 44 141 3308327
Email: barbara.nicholl@glasgow.ac.uk

Abstract

Background: Low back pain (LBP) is a common cause of disability and is ranked as the most burdensome health condition globally. Self-management, including components on increased knowledge, monitoring of symptoms, and physical activity, are consistently recommended in clinical guidelines as cost-effective strategies for LBP management and there is increasing interest in the potential role of digital health.

Objective: The study aimed to synthesize and critically appraise published evidence concerning the use of interactive digital interventions to support self-management of LBP. The following specific questions were examined: (1) What are the key components of digital self-management interventions for LBP, including theoretical underpinnings? (2) What outcome measures have been used in randomized trials of digital self-management interventions in LBP and what effect, if any, did the intervention have on these? and (3) What specific characteristics or components, if any, of interventions appear to be associated with beneficial outcomes?

Methods: Bibliographic databases searched from 2000 to March 2016 included Medline, Embase, CINAHL, PsycINFO, Cochrane Library, DoPHER and TRoPHI, Social Science Citation Index, and Science Citation Index. Reference and citation searching was also undertaken. Search strategy combined the following concepts: (1) back pain, (2) digital intervention, and (3) self-management. Only randomized controlled trial (RCT) protocols or completed RCTs involving adults with LBP published in peer-reviewed journals were included. Two reviewers independently screened titles and abstracts, full-text articles, extracted data, and assessed risk of bias using Cochrane risk of bias tool. An independent third reviewer adjudicated on disagreements. Data were synthesized narratively.

Results: Of the total 7014 references identified, 11 were included, describing 9 studies: 6 completed RCTs and 3 protocols for future RCTs. The completed RCTs included a total of 2706 participants (range of 114-1343 participants per study) and varied considerably in the nature and delivery of the interventions, the duration/definition of LBP, the outcomes measured, and the effectiveness of the interventions. Participants were generally white, middle aged, and in 5 of 6 RCT reports, the majority were female and most reported educational level as time at college or higher. Only one study reported between-group differences in

XSL•FO
RenderX

favor of the digital intervention. There was considerable variation in the extent of reporting the characteristics, components, and theories underpinning each intervention. None of the studies showed evidence of harm.

Conclusions: The literature is extremely heterogeneous, making it difficult to understand what might work best, for whom, and in what circumstances. Participants were predominantly female, white, well educated, and middle aged, and thus the wider applicability of digital self-management interventions remains uncertain. No information on cost-effectiveness was reported. The evidence base for interactive digital interventions to support patient self-management of LBP remains weak.

(J Med Internet Res 2017;19(5):e179) doi:10.2196/jmir.7290

KEYWORDS

low back pain; self-management; mHealth; eHealth

Introduction

The point prevalence of low back pain (LBP) is estimated to be 12% and one-month prevalence 23% across the globe [1]. The Global Burden of Disease study reported that LBP is the greatest contributor to disability in 12 of 21 world regions studied [2]. When considering years lived with disability, LBP is one of the leading causes of burden worldwide out of 291 conditions considered [2,3]. It is among the most common causes of long-term work absence and has a major impact on productivity at work [4,5]. Annual costs of LBP have been estimated to be approximately £10.7 billion for indirect factors in the United Kingdom [6,7] and up to US $200 billion in the United States [8], including workplace productivity costs; personal costs include a reduction in everyday functioning and quality of life [9].

Optimizing treatment strategies that are cost-effective, safe, and easy to administer for individuals with LBP is essential. Self-management is consistently recommended in international guidelines on the management of LBP [10,11]. Self-management focuses on the patient's ability to manage their own condition rather than treatment being based within the health care system or centered on a health care professional. The aim is to restore autonomy to the patient and include educational, or learning, components to position the patient at the center of their own management process and to help them acquire and maintain competencies to enable them to efficiently manage their condition [12].

A systematic review of the effectiveness of the self-management of LBP published in 2012 reports moderate quality evidence that self-management interventions have small, but clinically relevant, effects on reducing pain and disability for people with LBP when compared with minimal interventions [13]. The content and mode of delivery varied across the studies included, from receiving written information, attending face-to-face educational programs, functional movement training programs to information from websites [13].

Digital interventions (ie, interventions accessed via computer, mobile phone, or other handheld devices, including Web-based, desktop computer programs, or apps), providing self-management information have been proposed as a promising mode of delivery for self-management interventions. In a Cochrane Review from 2005, the use of such digital interventions was evaluated in people with chronic diseases and found to have a significant positive effect on knowledge, social support, and clinical outcomes in conditions such as diabetes and obesity [14]. Digital interventions have also been shown to effectively improve chronic pain, including chronic LBP, when compared with control groups (no care, waiting list, placebo, or care as usual) [15]. Providing supported self-management through digital platforms may enable individuals with LBP to better manage their symptoms. Garg et al [16] identified 9 randomized controlled trials (RCTs) for a systematic review of Web-based interventions to support individuals with LBP; included studies were grouped into cognitive behavioral therapy (CBT), a dialogue-based therapy that has been shown to have some efficacy for individuals with LBP [17], or knowledge improvement approaches with an interactive component [16]. Web-based methods were found to be useful, particularly CBT and those that offered an interactive support component; however, there was caution placed on the external validity of all studies included. Consequently, it appears that digital interventions hold potential in supporting the self-management of LBP but not enough is known about their content, delivery, and benefits, if any, or whether these interventions can be expected to be an improvement on traditional self-management approaches. Although it appears that the majority of digital interventions in this area have targeted individuals with chronic LBP (LBP for 3 months or longer), there is little known about the sociodemographic characteristics of individuals with LBP who are either targeted or who subsequently engage with such interventions.

The purpose of this systematic review was to synthesize published evidence concerning the characteristics, components, and effects of interactive digital interventions to support patient self-management of LBP. More specifically, the review aimed to address the following questions:

What are the key characteristics and components of digital self-management interventions for LBP, including theoretical underpinnings?

What outcome measures have been used in randomized trials of digital self-management interventions in LBP and what effect, if any, did the intervention have on these?

What specific characteristics or components, if any, of interventions appear to be associated with beneficial outcomes?

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

Participants: adults (18 years or above) with nonspecific LBP

Digital intervention:

- Any intervention accessed through a computer (work or home), mobile phone, or hand-held device, and included Web-based or desktop computer programs or apps that provided self-management information or material, which is in keeping with previous reviews in this sphere [19].
- Element of interaction between the user and digital interface: interaction was defined as patients entering data into the program or app, either by entering personal data or making choices that alter the pathways in the program and produce feedback in response to the patients' inputted data or choices.
- Interactive component as an add-on to face to-face health professional contact (eg, regularly seeing doctor but reporting pain levels electronically and receiving automated messages advising on physical activity level between visits).

Control group: usual care or digital noninteractive or nondigital self-management interventions for LBP

Study design: published randomized controlled trials (RCTs) or protocols for RCTs from peer-reviewed journals

Language: studies published in English, Danish, or Norwegian

Exclusion criteria

Digital intervention:

- Studies that only involved sending information to a remotely located health professional and receiving advice directly from the health professional.

Study design: all non-RCT reports and protocols

Methods

Study Design

The systematic literature review followed an *a priori* defined protocol as registered in PROSPERO (reference number 42016037954) and reporting is consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [18]. Inclusion and exclusion criteria are outlined in Textbox 1.

Information Sources and Search Strategy

A systematic search of the following databases was undertaken: Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Library (including Database of Abstracts of Reviews of Effects [DARE] and Health Technology Assessment [HTA] databases), Database of Promoting Health Effectiveness Reviews (DoPHER), Embase, MEDLINE, PsycINFO, Trials Register of Promoting Health Interventions (TRoPHI) and Web of Science (Social Science Citation and Science Citation Index). All databases were searched from 2000 until March 2016. Reference and citation searching were also undertaken. The searches were performed by an experienced Librarian at the Norwegian University of Science and Technology (NTNU). The search strategy included subject indexing terms and free-text terms for title, abstract, and keyword searching. The search terms were grouped into 3 concepts: (1) back pain, (2) digital interventions, and (3) self-management. The search terms were selected with reference to previous systematic reviews of interactive digital interventions for hypertension [19,20] and asthma [21,22] and after discussion with the review team. The full version of the search terms used, including specifications on use of title, keywords, or abstract screening, is documented for the example of MEDLINE in Multimedia Appendix 1.

Study Selection

All identified citations from the searched databases were uploaded to Distiller software (Evidence Partners). An integrated duplication detection tool was used to identify duplicates. All suggested duplicate pairs were screened for correctness by one reviewer (LS). Title and abstract screening was performed for each article by two independent reviewers from four (LS, BN, MM, NS). Disagreement between the two reviewers resulted in inclusion of the citation to full-text screening. Full-text screening was similarly performed by two independent reviewers from four (LS, BN, MM, NS), assessing the eligibility of the citation. Any disagreement was resolved through discussion mediated by a third reviewer (PJM).

Data Collection

Similar to the study selection process, data extraction was performed independently by two of four reviewers (LS, BN, MM, NS) using the Distiller software. Discrepancies in data extracted were considered by LS by revisiting the original paper to adjudicate on appropriateness and discussed and finalized with BN where required. Data were systematically extracted on study settings (country, inclusion and exclusion criteria, recruitment and participation numbers); study population (baseline characteristics such as age, gender, ethnicity, duration of symptoms, comorbidities); description of the intervention (details on the key components, characteristics, and underlying theoretical concepts); and outcome measures (time-points for outcome assessment, choice of primary outcomes, included secondary outcomes and effects, if any, noted as well as attrition rates, where available).

XSL•FO
RenderX

Outcome Measures

Our primary and secondary outcomes of particular interest are outlined in Textbox 2. These outcomes were a priori defined as of interest, however all outcomes reported were included in the data synthesis. For this review, pain-related disability was of special interest, as it measures a construct of the physical functioning domain, which has been recommended as a core domain in LBP research by several authors and guidelines [23-25].

Quality Appraisal

The methodological quality of all included studies was assessed using the Cochrane Collaboration tool for assessing risk of bias in randomized trials [26]. Two reviewers independently assessed selection bias (allocation concealment and randomization procedure); blinding of participants, personnel, and outcomes assessors; completeness of data; selective outcome reporting; and other potential biases. Any disagreements were resolved through discussion by the two independent reviewers (BN, PK). Papers were not excluded from the study on the basis of quality.

Data Synthesis

The study population, intervention components, outcomes, and characteristics of the included studies were narratively described. In our protocol we stated that we would conduct a meta-analysis if included studies were sufficiently homogeneous; however, due to the heterogeneity of identified studies, meta-analysis was not possible. Quantitative results from all outcomes reported in the completed RCT studies were described as either favoring the intervention group, no difference between groups, or favoring the control group. The outcomes reported in Textbox 2 were used as a basis to structure the results for research question 2. Included protocols for future RCTs were used to consider intervention components, characteristics, and outcome measures, but were not included in synthesis of intervention effects.

Results

Study Selection

We identified a total of 7014 citations, including 8 from searching reference lists of included studies. From these, 2316 were excluded as duplicates, and thus a total of 4698 titles were screened, resulting in the screening of 729 abstracts and 89 full-text papers. A total of 11 references concerning 9 different studies that described 5 RCT study protocols and 6 RCT reports met the inclusion criteria [27-37]. The PRISMA flow diagram demonstrating the screening process is illustrated in Figure 1 (adapted from Moher et al [18]).

Textbox 2. Outcome measures of interest.

Primary outcome

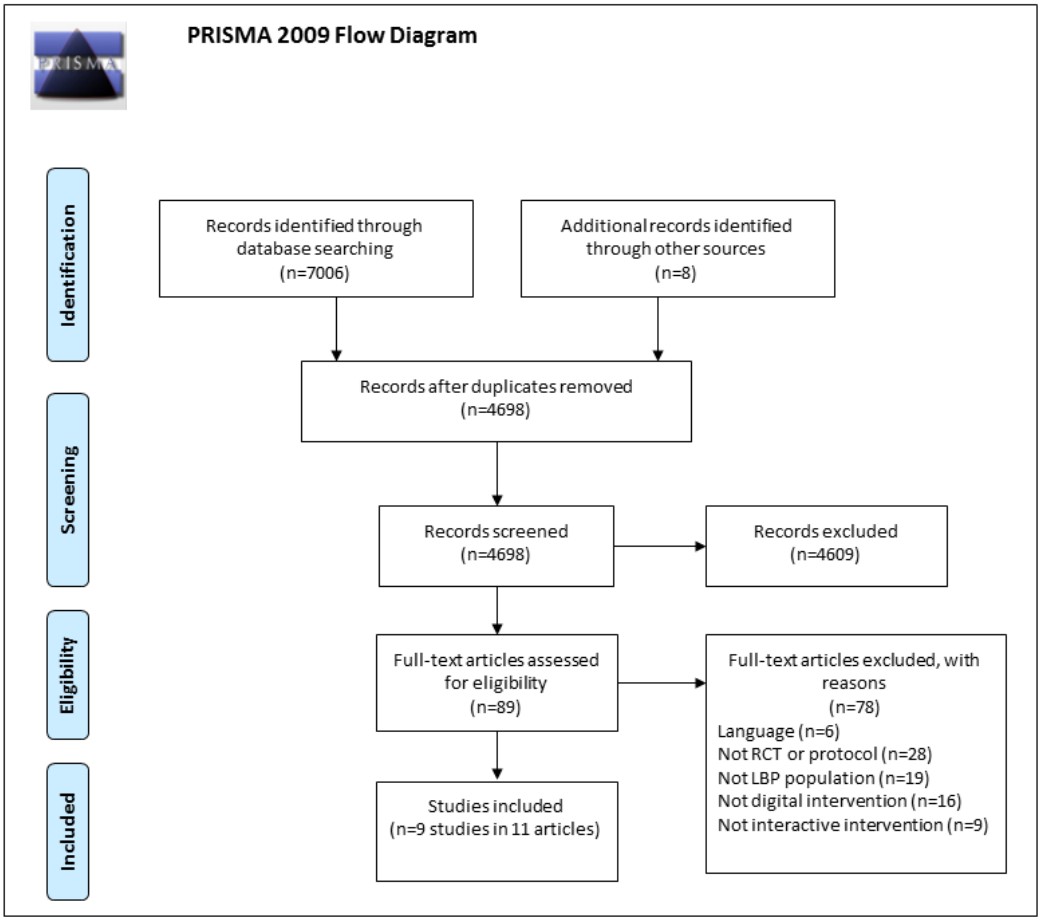
Details of outcome measures used to determine the effects of interventions for self-management of LBP pain-related disability

Secondary outcomes

- Pain intensity
- Quality of life
- Depression
- Fear avoidance
- Pain catastrophizing
- Physical activity
- Medication use
- Health care utilization (eg, primary and secondary care visits, emergency department visits)
- Health care costs
- Knowledge of LBP
- Markers of self-care
- Self-efficacy

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the screening process of papers.



Description of Included Studies

Four of the 9 separate studies were undertaken in the United States [28,29,32,33], 2 in Germany [35,37], one in the United Kingdom [31], one in Australia [27], and one in Spain [36]. The studies were published between 2010 and 2016. The 6 completed RCT reports included a total of 2706 participants, with a range of 114-1343 participants per study (Table 1).

Study Population

The characteristics of the study population in each of the studies are described in Table 1. There was considerable variation between studies in the duration of LBP symptoms, content and delivery of the interventions, and the measured outcomes. In 6 studies, LBP was defined by participant self-report [27-29,32,35,37] and in 3 studies by general practitioner evaluations [31,36] or diagnosis codes from medical records [33]. Seven studies included participants with pain for more than 3 months [27-29,32,33,36,37]. Only one study included participants with current LBP at the time of screening (or within the past 2 weeks) [31], while Simon et al [35] included only acute LBP participants, defined as participants who had experienced pain for less than 3 months. The included populations had a mean age ranging from 42.5 to 52.7 years; one study did not report the age of the population, except to say they were 18-65 years [32], a further study also had an upper age limit of 65 years [36]. All the remaining 7 studies did not report any upper limit in their inclusion criteria, yet only one study reported the age range of participants, 18-79 years [29]. In 5 [28,29,32,35,37] of the 6 RCT reports, the majority of the participants were female (58%-83%). The 6th study, which was conducted within the American Department of Veterans Affairs, included only 11% females in the intervention group and 14% in the control group [33]. Included participants were generally Caucasian (74%-87%) and the majority (42%-75%) reported educational levels as time spent at college or higher.

XSL•FO
RenderX

Table 1. Included studies and population characteristics.

Study	Definition of LBP[a]	Number	Age, mean (SD)[b]	Sex (%)	Ethnicity (%)	Comorbid (%)	SES (%)[c]
Chiauzzi et al [29] United States	LBP ≥10 days/month for at least 3 consecutive months	N=209 I[d]=104 C[e]=105	I=47.3 (12.2) C=45.0 (11.7)	I=F[f] (67) C=F (68)	White I=85 C=87	N/R[g]	Education I=72 C=77 report partial college or associates degree or higher Income, I=79 C=84 report ≥US $25,000/year
Simon et al [35] Germany	Acute LBP (<3 months)	N=1343 I=691 C=652	I=45.8 (12.7)[h] C=45.3 (13.0)[h]	F (82)[h]	N/R	N/R	Education 60% "high" education level[h]
Carpenter et al [28] United States	Noncancer LBP ≥6 months	N=141 I=70 C=71	42.5 (10.3)	F (83)	White (77)	N/R	Education 54% ≤2 years college
Krein et al [33] Krein et al [34][i] United States	Patients with ≥2 outpatient encounters within the past 12 months with a diagnosis of back pain with no neurologic findings (ICD-9-CM codes 724.2, 724.5, 846.0-846.9)	N=229[h] I=111 C=118 130/group[j]	I=51.2 (12.5) C=51.9 (12.8)	I=F (11%) C=F (14)	White I=74 C=86	N/R	Education I=72 C=75 reported some college or higher Income, I=82 C=87 ≥US $10,000/year
Irvine et al [32] United States	Nonspecific LBP within the past 3 months	N=398 I=199 C=199	N/R	I=F (58%) C=F (63)	White I=76 C=82	N/R	Education I=87 C=94 report some college or higher Income, I=89 C=94 ≥US $ 20,000/year
Weymann et al [37] Dirmaier et al [30][i] Germany	Chronic LBP: pain almost every day for >12 weeks	N=368[h] Randomized I=190 C=188 207/group[j]	I=52.2 (13.1)[h] C=52.7 (13.0)[h]	I=F, 162 (59)[h] C=F, 162 (59)[h]	N/R	N/R	Education[h] I=53 C=51 reported >10 years of education
Geraghty et al [31][i] United Kingdom	LBP in the past 3 months recorded in General Practitioner records and current LBP (or within the past 2 weeks) at the time of screening	20-30/group	-	-	-	-	-
Valenzuela-Pascual et al [36][i] Spain	Chronic LBP >6 months, confirmed by clinician	29/group	-	-	-	-	-

XSL•FO
RenderX

Study	Definition of LBP[a]	Number	Age, mean (SD)[b]	Sex (%)	Ethnicity (%)	Comorbid (%)	SES (%)[c]
Amorim et al [27][i] Australia	Chronic LBP persisting for >12 weeks but without radicular symptoms	34/group	-	-	-	-	-

[a]LBP: low back pain.

[b]SD: standard deviation.

[c]SES: socioeconomic status.

[d]I: intervention group.

[e]C: control group.

[f]F: female.

[g]N/R: not reported.

[h]Population comprising more conditions than LBP, numbers refer to the general population and were not available for LBP group only.

[i]Protocol paper, no data available unless reported alongside the RCT results paper.

[j]Planned number to recruit based on protocol paper.

Quality Assessment

Results for the 7 items of the Cochrane Risk of Bias tool for the 6 completed RCTs are reported in Table 2. Four studies had one or more items rated as unclear risk of bias [28,29,32,35], 4 studies had one item with high risk of bias [28,32,35,37], and only one study had a low risk of bias for all 7 items [33]. One study was assessed to have a high risk of incomplete data [35] because the attrition rate was extremely high. Two studies were rated as high risk of selective outcome reporting as one had no published protocol and reported the outcome with highest effect size as a primary outcome [28], and the other did not report on the primary outcome stated in the published protocol [32]. One study was assessed to have high risk of bias for other potential biases due to differences in educational level between the groups [37].

Key Intervention Components and Theoretical Underpinning of Digital Self-Management Interventions for LBP

Content

The extent of descriptions of the intervention content varied across studies (Table 3), but the level of details provided was generally sparse. The content of the digital interventions can be grouped into the following categories: (1) Pain education material: all studies report educational material as part of the intervention, which included information on pain origin, mechanisms and management, epidemiology of LBP, psychological aspects (eg, role of depression and mood), diagnostics and treatment-options; (2) General well-being activities: information concerning well-being, such as meditation, relaxation, general physical activity, and sleep hygiene, was reported in 4 studies [27-29,31]; (3) Exercise advice and goals: 5 studies described exercise advice, such as recommendations and goal-setting [27,29,31-33]. Two studies included short videos of exercises [32,33]; (4) E-community: one study reported a discussion forum with peers and health professionals in addition to the educational material [33]; (5) Narratives: one study included patient stories as part of the content [28].

Table 2. Quality assessment: risk of bias assessment with Cochrane Risk of Bias tool (low risk of bias, unclear risk of bias, high risk of bias).

Study	Random sequence generation	Allocation concealment	Blinding		Incomplete data	Selective reporting	Other biases
			Patients/personnel	Assessor			
Chiauzzi et al [29]	Low	Unclear	Unclear	Low	Low	Unclear	Unclear
Simon et al [35]	Low	Low	Unclear	Low	High	Low	Low
Carpenter et al [28]	Low	Unclear	Unclear	Low	Low	High	Unclear
Krein et al [33]	Low	Low	Low	Low	Low	Low	Low
Irvine et al [32]	Unclear	Unclear	Unclear	Low	Low	High	Unclear
Weymann et al [37]	Low	Low	Low	Low	Low	Low	High

XSL•FO
RenderX

Theoretical Underpinnings

Four of the 9 studies reported a theoretical underpinning to their intervention development (Table 3) [28,29,32,33]; cognitive behavior theory [28,29], collaborative decision making [29], social cognitive theory [32,33], theory of planned behavior [32], and acceptance and commitment therapy [28] were reported. The following approaches were mentioned as underpinning or rationales for the intervention: mindfulness [28], person-based approach [31], and self-management principles (not specified further) [29]; and tools such as goal setting [33] and information on pain and pain etiology [32,36]. Finally, 2 studies also reported that the advice given to participants was based on treatment guidelines, either evidence-based or recommendations from Governmental Institutes [27,37]. With regard to the tailoring element of interventions, 2 of the 7 studies that reported a tailoring element to their intervention described a systematic theoretical underpinning for the tailoring: Simon et al used the Ottawa Decision Support framework [35], whereas Weymann et al used the Avoidance Endurance Model and Health-Literacy as tailoring frameworks, as described in the study protocol [30,37].

Table 3. Intervention components and theoretical underpinnings.

Study	Content	Theoretical underpinning of content
Chiauzzi et al [29]	Educational material: content not more specifically described	Cognitive behavior theory
	Wellness activities: enhance good sleep, nutrition, stress management, exercise practices	Collaborative decision making
Simon et al [35]	Condition-specific information: epidemiology, etiology, diagnostics, treatment options	N/R[a]
Carpenter et al [28]	Educational chapters: all about pain, thoughts and pain, stress and relaxation, getting active	Cognitive therapy, behavioral activation
	Didactic material and interactive exercises	Acceptance and commitment therapy
	Patient stories	
	Guided relaxation and meditation exercises	
Krein et al [33] Krein et al[c][34]	Educational material: Handouts about topics (body mechanics, use of cold packs, lumbar rolls, and good posture); videos demonstrating specific strengthening and stretching exercises	Social cognitive theory[b]
	Pedometer data: weekly goals for steps	
	E-community: participants to post suggestions, ask questions, and share stories	
Irvine et al [32]	Education and behavioral strategies to manage and prevent pain: thirty 1-4 min videos on pain management, cognitive, and behavioral strategies; videos gain-framed messages with animated whiteboard-style coach; videos of ergonomics and exercises	Social cognitive theory Theory of planned behavior
Weymann et al [37] Dirmaier et al[c][30]	Educational information: physiology of pain, acute versus chronic pain; "chronification"; epidemiology; psychological aspects; coping and pain management	N/R
	Diagnostic procedures	
	Treatment options	
Geraghty et al[c][31]	Educational information: goal review; feedback on achievements; sessions on sleep, pain relief, flare-up, work, mood daily living.	N/R
	Supporting advice: managing pain; modeling expectation through patient stories; reinforcing positive behavior through automated feedback; simple instructions on back exercises/behavior	
Valenzuela-Pascual et al[c] [36]	Content not yet developed, but will be based on qualitative study including interviews with patients	N/R
Amorim et al[c] [27]	Educational material: "make your move—sit less, be active for life!"	N/R
	Information on how to increase physical activity and decrease sedentary behavior	
	Health-coaching by health care professional	
	FitBit activity monitor/feedback device	

[a]N/R: not reported.

[b]Information given in the protocol but not stated in the randomized controlled trial report.

[c]Protocol paper.

XSL•FO
RenderX

Outcome Measures Used in Digital Self-Management Interventions for LBP

Primary Outcomes

A wide range of outcomes were included in the RCTs (Table 4), with a total of 16 different outcomes being reported as a "primary outcome" measure. The number of primary outcomes per study ranged from 1 to 4. The primary outcome measures covered the domains of pain-related disability, pain intensity, attitude, depression, physical activity, knowledge of LBP, markers of self-care, and participant's assessment of change over time. Of the 6 completed trials, 4 studies [32,33,35,37] did not find a statistically significant effect on the primary outcome measures in favor of the intervention group; one study [28] reported a statistically significant effect in favor of the intervention compared with the control group on 6 of 7 subscales of their primary outcome—Survey of Pain Attitudes (SOPA)—following 3 weeks of intervention use (*F* statistic ranged from 5.1 to 44.7); while Chiauzzi et al [29] reported a favorable effect in the intervention arm but only in one of 4 primary outcomes that they measured (the Patient Global Impression Change Scale).

Pain-Related Disability

Pain-related disability was considered as the primary outcome in 4 of the 9 studies. The Roland-Morris Disability Questionnaire (RMDQ) was used in 2 of the 6 completed RCTs [28,33]. Carpenter et al [28] reported a significant difference in favor of the intervention group in RMDQ after 3 weeks of Web-based intervention compared with a waiting list control group (a reduction in RMDQ score of 2.8 for the intervention group compared with 0.8 for the control group; *P*=.01). Krein et al [33] similarly used the RMDQ, but observed reduced disability in chronic LBP with a 12-month, pedometer-based, Internet-supported, intervention of the same magnitude as the control group. The 3 protocols for RCT trials [27,31,36] all expected to use RMDQ as a measure of pain-related disability. The Oswestry Disability Index (ODI) was stated as the primary outcome measure in two RCT reports [29,32]. Chiauzzi et al [29] did not find a difference in ODI score between the intervention and control group after 4 weeks of access to a pain information website compared with static participant information. Irvine et al [32] did not report the trial results for ODI even though it was stated as a primary outcome in their Web-based trial registration.

Secondary Outcomes

A large variety of secondary outcome measures were described (Table 4 and Multimedia Appendix 2 provide a more detailed view). The outcome measures covered the following domains: pain-related disability; pain; health-related disability; depression/mood; fear of movement; pain catastrophizing; physical activity; knowledge of LBP, markers of self-care, and a range of other outcomes not held within our *a priori* defined domains. For the 3 protocols of future RCTs [27,31,36], a more consistent choice of outcomes was seen, as 2 outcomes—RMDQ and pain intensity—were planned to be measured in all 3 RCT protocols and 3 outcomes—the Tampa Scale of Kinesiophobia (TSK) [31,36], Pain Catastrophizing Scale (PCS) [31,36], and the International Physical Activity Questionnaire (IPAQ) [27,31]—were planned in 2 of the 3 RCT protocols. Below we provide an overview from the 6 included RCT reports of the treatment effects observed for the secondary outcomes that we had identified as being of interest in our systematic review protocol.

Pain Intensity

Pain intensity measured with either an 11-point Numerical Rating Scale (NRS) or a 100-mm Visual Analogue Scale (VAS) was reported in 3 of 6 RCT reports [28,32,33]. Only one study [32] reported that the digital intervention had a beneficial effect on pain intensity, 16 weeks post-intervention (eta-square = 0.43, *P*=.002); however, this was reported as a composite pain measure combining pain intensity, duration, and frequency.

Quality of Life

Health-related quality of life was reported in 2 studies using the Dartmouth Primary Care Cooperative Information Project (CO-OP) [32] and the Short-Form 12-Item questionnaire (SF-12) [34]. Only one of these studies actually reported the effect in the RCT report. Again Irvine et al used a composite outcome measure, incorporating functionality, well-being, and quality of life; however, they reported a beneficial effect of the intervention compared with the control arm (eta-square = 0.033, *P*=.001) [32].

Depression

Depression was reported in 3 of the 6 RCT reports [28,29,33] but only one study reported beneficial effects of the digital intervention using the Negative Mood Regulation Scale (an increase in score of 0.4 in the intervention group compared with 0.1 in the control group after 3 weeks of the intervention, *P*<.001) [28].

Fear Avoidance

Three studies reported fear of movement with the Fear Avoidance Belief Questionnaire (FABQ) [28,29,33], but only one reported an effect in favor of the digital intervention group [28], just for the physical activity subscale (decrease in score of 1.0 compared with an increase of 0.1 in the control group, after 3 weeks of the intervention, *P*<.001). One study used the TSK as a measure of fear avoidance; no between-group difference was reported [32].

Pain Catastrophizing

The PCS questionnaire was used in 2 RCT reports [28,29] but again only one study reported an effect in favor of the digital intervention compared with the waiting list control for the 3 subscales [28].

Physical Activity

Only one of the completed RCTs assessed physical activity outcomes and observed no difference in daily steps achieved between the control and intervention group [33].

Medication Use

No studies reported medication use.

XSL•FO
RenderX

Health Care Utilization

No studies reported details of health care utilization (eg, primary and secondary care visits, emergency department visits).

Health Care Costs

No studies reported on health care costs or cost-effectiveness.

Knowledge of LBP

Three of the RCT reports used participants' knowledge of LBP as an outcome measure [32,35,37]. Simon et al [35] and Weymann et al [37] used the same self-developed questionnaire, but neither study found a difference between the digital intervention and control group. Irvine et al assessed knowledge using a self-developed questionnaire and reported an effect in favor of the intervention group, however, as a composite score of 3 different outcomes (self-efficacy, behavior intentions, and knowledge) [32].

Markers of Self-Care

In total, 14 different outcomes were identified as markers of self-care, such as the Decision Conflict Scale [35,37], Patient Activation Measure, [32] and Preparation for Decision Making Scale [35,37]. Overall, 5 of the 14 outcomes showed an effect in favor of the digital invention when compared with a control group. Of the 14 outcomes, 10 were reported in only 3 of 6 RCT reports. Of these, the studies by Simon et al [35] and Weymann et al [37] originate from the same research group, and consequently there is considerable overlap between the interventions described and outcomes assessed in both trials. Irvine et al reported an effect on 3 outcomes of self-care in favor of the digital mobile app FitBack, when compared with the control group (behavior intentions, Patient Activation Measure, and prevention helping behaviors) [32].

Self-Efficacy

Four different measures of self-efficacy were reported in 4 RCT reports. The Self-Efficacy for Exercise Scale was used by Carpenter et al, who found an effect on self-efficacy in favor of the digital intervention group as compared with the waiting list control [28]. Irvine et al used a self-developed self-efficacy scale in a composite outcome score and reported a difference in favor of the digital intervention, however, again reported in a composite score [32]. The two other studies reported no benefits [29,33].

Other Outcomes

Ten outcomes could not be classified within our *a priori* defined outcome domains. These 10 included work-related outcomes, such as the Stanford Presenteeism Scale (SPS) [32], time off work, [32] and the Work Limitations Questionnaire (WLQ) [32], and procedural and implementation outcomes, which included issues such as feasibility [31], treatment adherence [35], as well as credibility and expectations of the intervention [31]. Four outcomes were additionally placed in an "other" category: the Chronic Pain Coping Inventory (CPCI) [29]; Participants' Global Impression of Change (PGIC) [29]; StartBack Screen Tool [31]; and the Problematic experience of Therapy Scale [31]. For these outcomes in the completed RCTs, Irvine et al reported a between-group difference favoring the digital intervention for the SPS and WLQ in a composite score [32] and Chiauzzi et al reported between-group differences favoring the digital intervention for the CPCI and PGIC outcomes [29].

Specific Characteristics or Components of Digital Self-Management Interventions for LBP Associated With Beneficial Outcomes

Key characteristics of the digital interventions are summarized in Table 5.

Aim of Interventions

Eight of the 9 studies aimed to investigate the effectiveness of the digital intervention in relation to pain intensity, attitudes toward pain, or pain-related disability by comparison with a control group (usual care or a nondigital intervention; as summarized in Table 4) [27-29,32,33,35-37]. One study had its main objective to explore the feasibility of the digital intervention [31].

Intervention Characteristics

Format and Delivery

Seven of 9 studies assessed digital interventions that were accessed over the Internet and by use of a computer [28,29,31,33,35-37], and 2 studies assessed digital interventions, which were app based, but accessible from both computer and handheld devices (tablets or smartphones) [27,32].

Frequency, Duration of Use, and Intervention Duration

Large variation was seen in the reported frequency and duration of use of the digital interventions. Six studies reported unlimited access to the programs with no report of recommendations given regarding frequency of use [27,32,33,35-37]. Geraghty et al [31] recommended a frequency of 1 session per week; Carpenter et al [28] recommended participants complete 2 chapters of the program per week over the 3-week study period; and Chiauzzi et al [29] instructed participants to log in for sessions twice per week. In 3 studies, weekly reminders to visit the website or app were sent to participants in the intervention groups [28,32,33]. Although all studies provided participants with a recommended frequency of use, only 2 of the 9 studies reported their recommended duration of use per visit with a range of 20 min per session to 1-1.5 hours per session [28,29]. Several studies reported that they registered user data but did not give results. Intervention duration also varied greatly, with 3 RCTs lasting between 2 and 4 weeks [28,29,36], one lasting 8 weeks [32], 3 were 3-month long [31,35,37], one study was 6-month long [27], and the longest duration was reported to be of 12 months [33].

Interactive Elements

The interactive elements reported in the studies included (1) keeping a log or journal of use of the intervention [29,32]; (2) simulated dialogue between the user and the system, where the user's answer(s) was (were) used to create individualized information [28,35,37]; (3) small exercises, such as quizzes, drag-and-drop questions [28]; (4) patient's report of outcome data and receiving feedback in the form of revised goals, for example, goals for steps per day based on pedometer data [27,33] or graphs illustrating changes in pain intensity [27,32];

XSL•FO
RenderX

(5) targeted messages with information and motivational feedback from the system [27,31-33]; and (6) Web-based discussion forums with peers and health care professionals [33].

XSL•FO
RenderX

Table 4. Study aim, available outcomes, and main results.

Study	Aim	Primary analysis		Secondary outcomes	Main result	Control condition
		Outcomes	Measurement Times			
Chiauzzi et al [29]	Compare interactive self-management website for chronic LBP to standard text-based materials; hypothesized improved emotional management, coping, self-efficacy to manage pain, pain levels, and physical functioning	BPI (Brief Pain Inventory) ODQ (Oswestry Disability Questionnaire) DASS (Depression/Anxiety and Stress Scale) PGIC (Patient Global Impression of Change scale)	Baseline, post-intervention (4 weeks), 3 months, 6 months	PCS (Pain Catastrophizing Scale) FABQ (Fear Avoidance Belief Questionnaire)	Hypothesis not supported	Educational material: "A back pain guide" No reminder emails
Simon et al [35]	Whether insurees with depression or LBP experienced more favorable decision-related outcomes after using a Web-based tailored decision aid compared with non-tailored, static patient information	DCS (Decisional Conflict Scale)	Baseline, post-intervention, 3 months	Preparation for decision-making scale Preference for participation, knowledge Doctor facilitation Information exchange Decision regret Treatment adherence	Intervention effective in short term Follow-up data of >3 months did not suggest further effects of intervention	Same information as intervention, website, but no tailoring to the individual user
Carpenter et al [28]	Efficacy of a pilot version of a Web-based CBT (cognitive behavioral therapy) intervention for chronic LBP	SOPA (Survey of Pain Attitudes)	Baseline, 3 weeks, 6 weeks	FABQ NMR (Negative Mood Regulation scale) PCS RMDQ PSES (Pain Self-Efficacy Scale) Demographics and pain assessment questionnaire	Difference in favor of the intervention group on all SOPA subscales in the SOPA questionnaire except "medical cure"	Wait list, received no care for 3 weeks, then access to website
Krein et al [33] Krein et al [34][a]	Whether a pedometer-based, Internet-mediated intervention would reduce pain-related disability and functional interference in chronic LBP	RMDQ (Roland-Morris Disability Questionnaire) SF-36 function scale	Baseline, 6 months, 12 months	Pain intensity (NRS, numerical rating scale) Walking (steps/day) FABQ PA (physical activity) subscale Self-efficacy 6-min walking test[b] CES-D 100[b](Centre for Epidemiologic Studies Depression Scale)	No between-group difference reported at any time-points	Usual care (attending Back Class) and uploading pedometer data after receiving monthly email reminders to upload; no goal-setting or feedback received; no access to website

XSL•FO
RenderX

Study	Aim	Primary analysis		Secondary outcomes	Main result	Control condition
		Outcomes	Measurement Times			
Irvine et al [32]	Test FitBack for adults at increased risk for chronic LBP due to a recent episode of NLBP	*No primary outcome stated* ODQ stated as primary outcome in trial registration	Baseline, 8 weeks, 16 weeks	Pain: level, frequency, intensity and duration MPI (Multidimensional Pain Inventory Interference Scale) Dartmouth CO-OP Prevention-helping behaviors (self-developed) WLQ (Work Limitations Questionnaire) SPS (Stanford Presenteeism Scale) PAM (Patient Activation Measures) Knowledge Behavioral intensions Self-efficacy SOPA (modified) TSK (Tampa Scale of Kinesiophobia; modified)	*No data available for primary outcome analysis*	Usual care, emails to request completion of questionnaire
Weymann et al [37] Dirmaier et al [30][a]	Investigate effectiveness of a Web-based, tailored, fully automated intervention for patients with type-2 diabetes or chronic LBP against a standard website with identical content without tailoring	Knowledge (post-intervention) Patient empowerment (heiQ, Health Education Impact Questionnaire; 3 months)	Baseline, post-intervention, 3 months	DCS PDMS (Preparation for Decision Making Scale)	The tailored intervention had no effect on the total study population	Same website material as intervention but not tailored; not presented in a dialogue format; no guidance through the content
Geraghty et al [31][a]	Explore feasibility of providing an Internet intervention for patients with LBP in primary care, with and without physiotherapist telephone support (in addition to usual care), compared with usual care alone	Feasibility outcome *Number need to screen* *Recruitment rates* *Login and usage information*	Baseline, 3 months	Pain: days, duration, intensity RMDQ StartBack Screen Tool TSK PCS IPAQ (International Physical Activity Questionnaire) PEI (Patient Enablement Instrument) EQ-5D (Euro-Qol 5D) LBP related health care use Time off work CEQ (Credibility and Expectancy Questionnaire) SESE (Self-Efficacy for Exercise Scale) PETS (Problematic Experiences of Therapy Scale)	-	Usual care from their general practitioner; this may consist of education and self-management advice, including advice to stay active

XSL•FO
RenderX

Study	Aim	Primary analysis		Secondary outcomes	Main result	Control condition
		Outcomes	Measurement Times			
Valenzuela-Pascual et al [36][a]	Evaluate effect of a biopsychosocial Web-based, educational intervention for chronic LBP based on pain intensity compared with normal care	Pain intensity (100-mm VAS [visual analogue scale] scale)	Baseline, 2 weeks	FABQ TSK PCS RMDQ SF-36	-	No intervention; asked to return to webpage to complete questionnaire at 2 weeks
Amorim et al [27][a]	Investigate effect of a patient-centered PA intervention supported by health coaching and technology in chronic LBP	Care-seeking Pain levels (NRS) RMDQ	Baseline, weekly during intervention, 6 months, 12 months	IPAQ *Actigraph* accelerometer GAS (Goal Attainment Scale)	-	Educational material same as intervention: "Make your move—Sit less, be active for life!"; advice to work toward increasing PA and achieving long-term goals

[a]Protocol paper, no data available.

[b]Difference between the protocol paper and RCT report.

Tailoring

Two of the 9 studies did not report any tailoring element to the content of their digital intervention [27,28]. Valenzuela-Pascual et al [36] did not specify the information they used for tailoring. Of the other 6 studies, all used some form of patient characteristics to inform tailoring, for example, Krein et al [33] used gender as a tailoring variable; Chiauzzi et al [29] used participant responses and characteristics (not further specified); Irvine et al [32] used job-type assessed by questionnaires; and Geraghty et al [31] used the extent to which LBP obstructed daily activities as a tailoring variable.

XSL•FO
RenderX

Table 5. Intervention characteristics.

Study	Mode of delivery	Recommended frequency	Recommended duration of visit	Interactive element	Tailoring	Intervention Duration Attrition rate[g]
Chiauzzi et al [29]	Website	2 times/week for 4 weeks, then unlimited	<20 min/session	Log of activities and content viewed during sessions	Yes Matched patient characteristics to educational content, articles, and interactive tools	4 week intervention period, access for 6 months Attrition rate: 6 months I[a]=67/104 C[b]=88/105
Simon et al [35]	Website *Small information units presented in combination of text and graphics*	Unlimited access but no required frequency	N/R[c]	Simulated dialogue between user and system Text or graphics varied based on needs of users	Yes Ottawa Decision Support Framework Tailoring based on ≥4 tailoring concepts, including patient characteristics and preferences	One-time use required, access for 3 months Attrition rate: Post-use I[a]=147/691 C[b]=195/652 3 months I[a]=40/691 C[b]=25/652
Carpenter et al [28]	Website *Text and graphic with audio narration* *Animation used in educational material*	Two times/week, email reminders	1-1.5 hour/log-in	Reflective and interactive exercises	No	3-week intervention period Attrition rate: 3 weeks I[a]=63/70 C[b]=68/71
Krein et al [33] Krein et al[34][d]	Website *Graphical and written feedback* *Motivational messages* *Weekly news updates*	Unlimited access with weekly reminders to upload data	N/R[c]	Pedometer data, used to create weekly PA[e]goals and track progress Targeted messages Discussion on Web-based forum with peers and health personnel	Yes Gender[f] Written and graphical information as targeted messages[f]	12-month intervention period Attrition rate: 12 months I[a]=102/111 C[b]=105/118
Irvine et al [32]	Web app, accessible from Internet and mobile *Gain-framed text and video messages*	Unlimited access, weekly reminders to visit app	N/R[c]	Pain and PA[e]self-monitoring tool Journal-keeping function 7- and 30-day graphs of pain	Yes Job-type assessed by questionnaires	8-week intervention period, access for 16 weeks Attrition rate: 8 weeks I[a]=192/199 C[b]=197/199
Weymann et al [37] Dirmaier et al [30][d]	Website	Unlimited access, designed to be used in 1 sitting	N/R[c]	Simulated dialogue between user and system User-control to navigate site by replying to at least 3 options after each text passage	Yes Avoidance Endurance Model Health literacy[f] Motivational Interviewing Tunnelled design developed	3-month intervention period Attrition rate: 3 months I[a]=96/190 C[b]=106/188
Geraghty et al [31][d]	Website	One session/week	N/R[c]	User selects PA[e], system generates activity goals User may navigate the content as they find best	Yes Extent of pain obstructing daily activities	3-month intervention period

XSL•FO
RenderX

Study	Mode of delivery	Recommended frequency	Recommended duration of visit	Interactive element	Tailoring	Intervention Duration Attrition rate[g]
Valenzuela-Pascual et al [36][d]	Website *Changing delivery formats* *Video, 2-3D animation*	Unlimited access	N/R[c]	N/R[c] *(content not yet developed)*	Yes *(content not yet developed)*	2-week intervention period
Amorim et al [27][d]	App, accessed via computer or smartphone	Unlimited access, no recommendations on frequency or duration	N/R[c]	User reports PA[e]levels, pain intensity, and disability User receives encouragement based on PA[e]level	No	6-month intervention period

[a]C: control group.

[b]I: intervention group.

[c]N/R: not reported.

[d]Protocol paper, no data available.

[e]PA: physical activity.

[f]Information given in the protocol but not stated in the RCT report.

[g]Attrition rates reported as number of completed cases in relation to the total number of participants randomized to the group.

Discussion

Principal Findings

We have systematically searched and reviewed the literature pertaining to interactive, digital interventions for self-management of LBP. The effectiveness of interventions was mixed, with only 1 study reporting a positive effect on their primary outcome [28]. We found a large degree of heterogeneity regarding the description of intervention content and delivery, theoretical underpinnings, and outcomes reported, making comparison between interventions difficult. A comprehensive description of intervention development and use of theory has been recommended when reporting on RCTs of digital interventions [38]; however, such descriptions were either brief or completely lacking in the included studies. Participants were predominantly female, white, younger, and well educated, which renders the external validity of the identified studies as low.

Despite international recommendations for reporting core outcome domains (physical functioning, pain intensity, and health-related quality of life) in LBP studies [23], we identified 16 different primary outcome measures and a total of 52 outcomes covering a wide range of domains. Better consistency in choice of outcome measures was seen in the 3 RCT protocols of planned RCTs [27,31,36]. We expect that these trials will provide more useful information and data for future meta-analyses. Generally, the included studies were not able to demonstrate significant beneficial effects on either the primary or secondary outcomes and we were unable to identify specific characteristics of interventions to explain these findings. However, it may be that the most important factors related to whether an individual engages with a digital support tool were not taken into account, for example, low mood and additional physical comorbidities. Surprisingly, physical activity, which is considered one of the mainstays of the treatment of LBP [10], was only included as a key component in one study. Consequently, no evidence was presented to support effects on physical activity behavioral changes from digital self-management for LBP. This should be a matter for focus for future RCTs in this area. None of the studies showed any evidence of harm from interactive digital interventions. There was no evidence regarding cost-effectiveness of interactive digital interventions.

Strengths and Limitations

This systematic review was undertaken by a team with extensive experience in conducting such reviews. We used multiple databases, and a thorough search strategy that was designed iteratively by the research team and an information specialist to account for the 3 different dimensions of the search (back pain, digital interventions, and self-management). The methodological assessment tool used in our systematic review has been specifically developed to assess the risk of bias in RCTs [26], and its constructs are in line with the recommendations of the PRISMA statement [18]. All aspects of data extraction, quality appraisal, and data analysis were carried out independently by two researchers, with a third party available for adjudication in case of disagreements.

The primary limitation of this systematic review is the sparse literature related to our objectives. Due to the sparsity and heterogeneity of the data, a formal meta-analysis was not possible. Additionally, our search was limited to studies published in English, Danish, or Norwegian, which could be construed as a limitation, although there is increasing evidence that this is not a particular problem [39]; six papers were excluded at the abstract screening stage of this review based on language. Finally, gray literature was not included; however, given the nature of this review and that there is no suggestion of publication bias, it is unlikely that this will have any impact on the results.

XSL•FO
RenderX

Comparison With Previous Literature

To the best of our knowledge this is the first systematic review of RCTs of interactive digital interventions for self-management of LBP. However, systematic reviews of Web-based interventions for LBP (not specifically self-management) [16], nondigital self-management for LBP [13], and chronic musculoskeletal pain [40] have been published. The first review suggests that CBT-based approaches and interventions that offer Web-based support may have some effect on reducing pain-related catastrophizing and improving patient attitudes; however, study quality was relatively low and further studies were recommended [16]. Reviews of interventions targeted specifically at self-management have suggested that there is only moderate-quality evidence that self-management has small effects on pain and disability in people with LBP [13,40]. These reviews have not dismissed self-management as a treatment option for LBP, but rather suggested that further research is needed to understand the limitations of self-management and whether or how effectiveness can be increased. In addition, these reviews have suggested that future studies should extend the outcomes of interest to include aspects of self-efficacy, and also consider the impact of the duration of the intervention [13,40], increase the length of follow-up [16], and also consider the impact of such interventions on health care utilization [16]. Similar conclusions have been made in systematic reviews of digital self-management interventions in conditions like asthma [22], hypertension [19], and problematic cannabis use [41]. Tailoring digital interventions to individual patient needs has been advocated to enhance engagement [42]. Our review highlights that although 5 out of 6 of the RCT reports included some form of tailoring, there was a lack of detail on exactly what this involved and the role it played in the outcome of the RCT or in user engagement. Finally, small and very similar effects across types of interventions such as different types of exercises, manual treatment, or acupuncture for people with LBP are well recognized [43]; however, because of the enormous societal impact of LBP and LBP-related disability, these interventions may still have worthwhile effects both at the patient and population level [43,44]. In this context, digital interventions aiming to promote self-management are particularly attractive because they are easy to deliver, inexpensive, and safe.

Study Implications

The populations within the identified studies were predominantly female, white, well-educated, and middle-aged, and thus the wider applicability of digital self-management interventions remains uncertain and therefore further investigation including a broader range of participants is merited. Seven of the 9 included studies specifically aimed to address the self-management of chronic LBP, and thus the usefulness of supporting self-management for acute LBP using digital tools remains underinvestigated; any such interventions for acute LBP would possibly require different advice and support to that offered for chronic LBP, as directed in clinical guidelines [10,11]. In addition, the absence of any health economics data was surprising and certainly needs to be addressed in future studies. There were a number of areas of reporting that were identified as deficient in the majority of studies in this systematic review. This suggests that going forward greater adherence to published guidelines that have recommended increasing focus on reporting of the technical aspects of the digital intervention as well as reporting the content of the intervention and its theoretical underpinnings [38,45] would be valuable. Finally, there is growing evidence that tailoring of digital interventions may be an important ingredient for success [42], and this will be an important issue to address in future RCTs of digital interventions aimed at promoting self-management of LBP. We are aware of at least one such study currently underway [46,47].

Conclusions

Our review has highlighted that the published literature is extremely heterogeneous and that digital intervention studies for LBP are generally poorly described. The literature provides insufficient detail regarding target and participating populations, and intervention components, theoretical underpinnings, and the rationale for the wide variety of outcome measures used. This makes it difficult to gain a clear impression of what might work best, for whom and in what circumstances. It is clear that the existing evidence has not yet proven the wider utility of digital interventions for self-management of LBP for the population at large, a knowledge gap that future research should address by better characterizing participants and interventions in a way that would allow replication and by providing clear rationales for intervention components and outcome measure selection.

Acknowledgments

We would like to thank our librarian adviser Ingrid Ingeborg Riphagen, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU). This project has received funding from the European Union Horizon 2020 research and innovation program under grant agreement no. 689043.

Authors' Contributions

BN, LS, MS, OV, JH, PM, PK, KS, and FM contributed to the design of the study. LS and BN managed the review process. LS, MM, NS, BN, PJM, and PK were involved in one or more of the following stages of the review: screening, data extraction, quality appraisal, or resolution of conflicts. LS led on data synthesis with input from BN, FM, JH, PK, and KS. BN and LS drafted the manuscript. FM oversaw manuscript preparation.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE search strategy.

[PDF File (Adobe PDF File), 285KB - jmir_v19i5e179_app1.pdf]

Multimedia Appendix 2

Overview of all outcome measures included.

[PDF File (Adobe PDF File), 374KB - jmir_v19i5e179_app2.pdf]

References

1. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012 Jun;64(6):2028-2037 [FREE Full text] [doi: 10.1002/art.34347] [Medline: 22231424]
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, de Vaccaro, Fowkes Francis Gerry R, De Leòn Felipe Rodriguez, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196. [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
3. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jun;73(6):968-974. [doi: 10.1136/annrheumdis-2013-204428] [Medline: 24665116]
4. Walker BF, Muller R, Grant WD. Low back pain in Australian adults. health provider utilization and care seeking. J Manipulative Physiol Ther 2004 Jun;27(5):327-335. [doi: 10.1016/j.jmpt.2004.04.006] [Medline: 15195040]
5. Driscoll T, Jacklyn G, Orchard J, Passmore E, Vos T, Freedman G, et al. The global burden of occupationally related low back pain: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jun;73(6):975-981. [doi: 10.1136/annrheumdis-2013-204631] [Medline: 24665117]
6. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8(1):8-20. [doi: 10.1016/j.spinee.2007.10.005] [Medline: 18164449]
7. Maniadakis N, Gray A. The economic burden of back pain in the UK. Pain 2000 Jan;84(1):95-103. [Medline: 10601677]
8. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. J Bone Joint Surg Am 2006 Apr;88(Suppl 2):21-24. [doi: 10.2106/JBJS.E.01273] [Medline: 16595438]
9. Nolet PS, Kristman VL, Côté P, Carroll LJ, Cassidy JD. Is low back pain associated with worse health-related quality of life 6 months later? Eur Spine J 2015 Mar;24(3):458-466. [doi: 10.1007/s00586-014-3649-4] [Medline: 25391622]
10. National Institute for HealthCare Excellence (NICE). NICE.: National Institute for Health and Care Excellence (NICE); 2016 Nov. Low back pain and sciatica in over 16s: assessment and management. NICE guideline [NG59] URL: https://www.nice.org.uk/guidance/ng59 [accessed 2017-05-04] [WebCite Cache ID 6qDAgfu6k]
11. The Danish Health and Medicines Authorities. SST. 2016 Jun. National clinical guideline for treatment of recent onset low back pain URL: https://www.sst.dk/da/udgivelser/2016/nkr-laenderygsmerter [accessed 2017-05-04] [WebCite Cache ID 6qDAxDZtd]
12. Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, et al. The rising prevalence of chronic low back pain. Arch Intern Med 2009 Feb 9;169(3):251-258. [doi: 10.1001/archinternmed.2008.543] [Medline: 19204216]
13. Oliveira VC, Ferreira PH, Maher CG, Pinto RZ, Refshauge KM, Ferreira ML. Effectiveness of self-management of low back pain: systematic review with meta-analysis. Arthritis Care Res (Hoboken) 2012 Nov;64(11):1739-1748 [FREE Full text] [doi: 10.1002/acr.21737] [Medline: 22623349]
14. 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]
15. Beatty L, Lambert S. A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. Clin Psychol Rev 2013 Jun;33(4):609-622. [doi: 10.1016/j.cpr.2013.03.004] [Medline: 23603521]
16. Garg S, Garg D, Turin TC, Chowdhury MF. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
17. Richmond H, Hall AM, Copsey B, Hansen Z, Williamson E, Hoxey-Thomas N, et al. The effectiveness of cognitive behavioural treatment for non-specific low back pain: a systematic review and meta-analysis. PLoS One 2015;10(8):e0134192 [FREE Full text] [doi: 10.1371/journal.pone.0134192] [Medline: 26244668]
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535 [FREE Full text] [Medline: 19622551]

XSL•FO
RenderX

19. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, et al. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. J Hypertens 2016 Apr;34(4):600-612 [FREE Full text] [doi: 10.1097/HJH.0000000000000859] [Medline: 26845284]
20. McLean G, Murray E, Band R, Saunderson K, Hanlon P, Little P, et al. Digital interventions to promote self-management in adults with hypertension: protocol for systematic review and meta-analysis. JMIR Res Protoc 2015 Nov 20;4(4):e133 [FREE Full text] [doi: 10.2196/resprot.4648] [Medline: 26589728]
21. McLean G, Murray E, Band R, Moffat KR, Hanlon P, Bruton A, et al. Interactive digital interventions to promote self-management in adults with asthma: systematic review and meta-analysis. BMC Pulm Med 2016 May 23;16(1):83 [FREE Full text] [doi: 10.1186/s12890-016-0248-7] [Medline: 27215329]
22. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]
23. Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, et al. Core outcome domains for clinical trials in non-specific low back pain. Eur Spine J 2015 Jun;24(6):1127-1142. [doi: 10.1007/s00586-015-3892-3] [Medline: 25841358]
24. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. a proposal for standardized use. Spine (Phila Pa 1976) 1998 Sep 15;23(18):2003-2013. [Medline: 9779535]
25. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH Task Force on research standards for chronic low back pain. J Pain 2014 Jun;15(6):569-585 [FREE Full text] [doi: 10.1016/j.jpain.2014.03.005] [Medline: 24787228]
26. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
27. Amorim AB, Pappas E, Simic M, Ferreira ML, Tiedemann A, Jennings M, et al. Integrating mobile health and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot trial protocol. BMC Musculoskelet Disord 2016;17:36 [FREE Full text] [doi: 10.1186/s12891-015-0852-3] [Medline: 26787469]
28. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
29. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
30. Dirmaier J, Härter M, Weymann N. A tailored, dialogue-based health communication application for patients with chronic low back pain: study protocol of a randomised controlled trial. BMC Med Inform Decis Mak 2013;13:66 [FREE Full text] [doi: 10.1186/1472-6947-13-66] [Medline: 23768119]
31. Geraghty AW, Stanford R, Little P, Roberts L, Foster NE, Hill JC, et al. Using an internet intervention to support self-management of low back pain in primary care: protocol for a randomised controlled feasibility trial (SupportBack). BMJ Open 2015 Sep 23;5(9):e009524 [FREE Full text] [doi: 10.1136/bmjopen-2015-009524] [Medline: 26399575]
32. Irvine AB, Russell H, Manocchia M, Mino DE, Cox GT, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
33. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013 Aug 19;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
34. Krein SL, Metreger T, Kadri R, Hughes M, Kerr EA, Piette JD, et al. Veterans walk to beat back pain: study rationale, design and protocol of a randomized trial of a pedometer-based internet mediated intervention for patients with chronic low back pain. BMC Musculoskelet Disord 2010;11:205 [FREE Full text] [doi: 10.1186/1471-2474-11-205] [Medline: 20836856]
35. Simon D, Kriston L, von WA, Buchholz A, Vietor C, Hecke T, et al. Effectiveness of a web-based, individually tailored decision aid for depression or acute low back pain: a randomized controlled trial. Patient Educ Couns 2012 Jun;87(3):360-368. [doi: 10.1016/j.pec.2011.10.009] [Medline: 22154867]
36. Valenzuela-Pascual F, Molina F, Corbi F, Blanco-Blanco J, Gil RM, Soler-Gonzalez J. The influence of a biopsychosocial educational internet-based intervention on pain, dysfunction, quality of life, and pain cognition in chronic low back pain patients in primary care: a mixed methods approach. BMC Med Inform Decis Mak 2015 Nov 23;15:97 [FREE Full text] [doi: 10.1186/s12911-015-0220-0] [Medline: 26597937]
37. 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]
38. 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]

XSL•FO
RenderX

39. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care 2012 Apr;28(2):138-144. [doi: 10.1017/S0266462312000086] [Medline: 22559755]
40. Taylor S, Carnes D, Homer K, Pincus T, Kahan B, Hounsome N, et al. Improving the self-management of chronic pain: COping with persistent Pain, Effectiveness Research in Self-management (COPERS). Southampton, UK: NIHR Journals Library; Sep 2016.
41. Hoch E, Preuss UW, Ferri M, Simon R. Digital interventions for problematic Cannabis users in non-clinical settings: findings from a systematic review and meta-analysis. Eur Addict Res 2016;22(5):233-242 [FREE Full text] [doi: 10.1159/000445716] [Medline: 27160333]
42. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
43. Artus M, van der Windt DA, Jordan KP, Hay EM. Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. Rheumatology (Oxford) 2010 Dec;49(12):2346-2356. [doi: 10.1093/rheumatology/keq245] [Medline: 20713495]
44. Hancock MJ, Hill JC. Are small effects for back pain interventions really surprising? J Orthop Sports Phys Ther 2016 May;46(5):317-319. [doi: 10.2519/jospt.2016.0604] [Medline: 27133941]
45. 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]
46. Nicholl B, Bach K, Aamodt A, Mork P, Mair F. selfBACK: a decision support system for self-management of low back pain. 2016 Presented at: North American Primary Care Research Group Annual Meeting; November 11-15, 2016; Colorado Springs.
47. Mork PJ. Selfback. 2016. A decision support system for self-management of non-specific low back pain URL: http://cordis.europa.eu/project/rcn/199886_en.html

Abbreviations

BPI: Brief Pain Inventory
CBT: cognitive behavioral therapy
CDSR: Cochrane Database of Systematic Reviews
CENTRAL: Cochrane Central Register of Controlled Trials
CEQ: Credibility and Expectancy Questionnaire
CES-D-100: Centre for Epidemiologic Studies Depression Scale
CINAHL: Cumulative Index to Nursing and Allied Health Literature
CO-OP: Primary Care Cooperative Information Project
CPCI: Chronic Pain Coping Inventory
DARE: Database of Abstracts of Reviews of Effects
DASS: Depression/Anxiety and Stress Scale
DCS: Decisional Conflict Scale
DoPHER: Database of Promoting Health Effectiveness Reviews
FABQ: Fear Avoidance Belief Questionnaire
GAS: Goal Attainment Scale
heiQ: Health Education Impact Questionnaire
HTA: Health Technology Assessment
IPAQ: International Physical Activity Questionnaire
LBP: low back pain
MPI: Multidimensional Pain Inventory Interference Scale
NMR: Negative Mood Regulation scale
NRS: Numerical Rating Scale
NTNU: Norwegian University of Science and Technology
ODQ: Oswestry Disability Questionnaire
ODI: Oswestry Disability Index
PA: physical activity
PAM: Patient Activation Measures
PCS: Pain Catastrophizing Scale
PDMS: Preparation for Decision Making Scale
PEI: Patient Enablement Instrument
PETS: Problematic Experiences of Therapy Scale

XSL•FO
RenderX

PGIC: Patient Global Impression of Change
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSES: Pain Self-Efficacy Scale
RCT: randomized controlled trial
RMDQ: Roland-Morris Disability Questionnaire
SD: standard deviation
SES: socioeconomic status
SOPA: Survey of Pain Attitudes
SPS: Stanford Presenteeism Scale
TSK: Tampa Scale of Kinesiophobia
TROPHI: Trials Register of Promoting Health Interventions
TSK: Tampa Scale of Kinesiophobia
VAS: visual analogue scale
WLQ: Work Limitations Questionnaire

Edited by G Eysenbach; submitted 09.01.17; peer-reviewed by S Peterson, A Geraghty, K Morton; comments to author 29.01.17; revised version received 13.03.17; accepted 18.03.17; published 21.05.17.

Please cite as:
Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, Hartvigsen J, Mork PJ, Kjaer P, Søgaard K, Mair FS
Digital Support Interventions for the Self-Management of Low Back Pain: A Systematic Review
J Med Internet Res 2017;19(5):e179
URL: http://www.jmir.org/2017/5/e179/
doi:10.2196/jmir.7290
PMID:28550009

©Barbara I Nicholl, Louise F Sandal, Mette J Stochkendahl, Marianne McCallum, Nithya Suresh, Ottar Vasseljen, Jan Hartvigsen, Paul J Mork, Per Kjaer, Karen Søgaard, Frances S Mair. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 21.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study

Silje Stangeland Lie[1], PT, MSc; Bjørg Karlsen[1], RN, MSc, PhD; Ellen Renate Oord[2], RN; Marit Graue[3], RN, MSc, PhD; Bjørg Oftedal[1], RN, MSc, PhD

[1]Department of Health Studies, University of Stavanger, Stavanger, Norway

[2]Stavanger University Hospital, Stavanger, Norway

[3]Centre for Evidence-based Practice, Western Norway University of Applied Sciences, Bergen, Norway

Corresponding Author:
Silje Stangeland Lie, PT, MSc
Department of Health Studies
University of Stavanger
Kjell Arholms hus
Stavanger, N-4036
Norway
Phone: 47 97506752
Fax: 47 51 83 37 50
Email: silje.s.lie@uis.no

Abstract

Background: Adequate self-management is the cornerstone of type 2 diabetes treatment, as people make the majority of daily treatment measures and health decisions. The increasing prevalence of type 2 diabetes mellitus (T2DM) and the complexity of diabetes self-management demonstrate the need for innovative and effective ways to deliver self-management support. eHealth interventions are promoted worldwide and hold a great potential in future health care for people with chronic diseases such as T2DM. However, many eHealth interventions face high dropout rates. This led to our interest in the experiences of participants who dropped out of an eHealth intervention for adults with T2DM, based on the Guided Self-Determination (GSD) counseling method.

Objective: In this study, we aimed to explore experiences with an eHealth intervention based on GSD in general practice from the perspective of those who dropped out and to understand their reasons for dropping out. To the best of our knowledge, no previous qualitative study has focused on participants who withdrew from an eHealth self-management support intervention for adults with T2DM.

Methods: A qualitative design based on telephone interviews was used to collect data. The sample comprised 12 adults with type 2 diabetes who dropped out of an eHealth intervention. Data were collected in 2016 and subjected to qualitative content analysis.

Results: We identified one overall theme: “Losing motivation for intervention participation.” This theme was illustrated by four categories related to the participants’ experiences of the eHealth intervention: (1) frustrating technology, (2) perceiving the content as irrelevant and incomprehensible, (3) choosing other activities and perspectives, and (4) lacking face-to-face encounters.

Conclusions: Our findings indicate that the eHealth intervention based on GSD without face-to-face encounters with nurses reduced participants’ motivation for engagement in the intervention. To maintain motivation, our study points to the importance of combining eHealth with regular face-to-face consultations. Our study also shows that the perceived benefit of the GSD eHealth intervention intertwined with choosing to focus on other matters in complex daily lives are critical aspects in motivation for such interventions. This indicates the importance of giving potential participants tailored information about the aim, the content, and the effort needed to remain engaged in complex interventions so that eligible participants are recruited. Finally, motivation for engagement in the eHealth intervention was influenced by the technology used in this study. It seems important to facilitate more user-friendly but high-security eHealth technology. Our findings have implications for improving the eHealth intervention and to inform researchers and health care providers who are organizing eHealth interventions focusing on self-management support in order to reduce dropout rates.

XSL•FO
RenderX

(J Med Internet Res 2017;19(5):e187) doi:10.2196/jmir.7479

KEYWORDS

eHealth; Telehealth; type 2 diabetes; Internet; counseling; qualitative research; general practice, self-management; self-management support; patient dropouts

Introduction

eHealth interventions are promoted worldwide and hold a great potential in future health care for people with chronic diseases such as type 2 diabetes mellitus (T2DM). However, many eHealth interventions face adoption problems and high dropout rates [1-5]. This led to our interest in the experiences of participants who withdrew from an eHealth intervention for adults with T2DM at general practices in Norway.

Diabetes is a chronic disease affecting an estimated 415 million people worldwide. Most of them have T2DM and its prevalence is rapidly increasing [6]. People living with diabetes are recommended to engage in multiple self-care behaviors such as taking medications, following a diet, engaging in regular physical activity, and self-monitoring, in addition to problem-solving and coping [7]. These are all aspects of diabetes self-management and essential to blood glucose control for the prevention of long-term complications. Many people with T2DM find adequate self-management difficult to achieve and maintain [8]. Some of the recommended self-management behaviors do not coincide with peoples' priorities and desire for a "normal life." They may differ from people's habits and preferences and be perceived as burdensome [9,10]. Research indicates that only 1 in 8 patients with T2DM achieves the recommended treatment goals of glycemic control, cholesterol, and blood pressure [11]. Consequently, to achieve adequate self-management and optimal treatment outcomes, many patients need support from a health care professional. Given the increasing prevalence of T2DM, there is a need for innovative and effective ways to deliver self-management support interventions for people with T2DM. eHealth self-management support interventions can assist people with adopting and maintaining behaviors needed for adequate diabetes self-management [12-14].

Secure messaging is an eHealth technology that facilitates personal and interactive communication between health care providers and patients. A systematic review of participatory Web-based interventions found that asynchronous communication tools such as secure messaging was experienced as particularly useful for self-management support [2]. Such communication between patients and health care providers seems to improve effects and adherence in eHealth interventions [15-17]. Moreover, previous research has addressed the need for theory-based eHealth interventions for T2DM [14]. Theory-based interventions are valuable as the theory inform intervention strategies. These strategies translate into key components of the interventions that can be applied and assessed, thus facilitating explanation of observed effects or lack thereof [18,19].

As a response to the need for effective and theory-based interventions for people with T2DM, we adapted the self-management support intervention Guided Self-Determination (GSD) for T2DM [20], as an eHealth intervention via secure messaging in general practices (Table 1 and Textbox 1). GSD is a counseling approach founded on the self-determination theory (SDT). This theory proposes that in order to foster autonomous motivation for engagement in activities, it is important to support individuals' basic psychological needs for autonomy, relatedness, and competence [21]. The GSD intervention aims to support diabetes self-management by empowering self-determined goal-setting and competence-building [22,23]. The intervention is described in more detail in the Methods section.

Some eHealth interventions show dropout rates of up to 80% [3-5]. A systematic review, exploring Web-based interventions designed to support and promote diabetes education and health behavior change for management of T2DM, similarly shows that intervention-engagement and usage declined over time. About half of the interventions focused on support and coping skills, and the most targeted behaviors were physical exercise, diet, and blood glucose self-monitoring [15]. A meta-analysis of the effectiveness of Web-based tools for people with diabetes suggests that participants' difficulties in understanding the use of Web-based interventions led to higher dropout rates [24]. Moreover, a study investigating adherence to a Web-based intervention to support diabetes self-management through components derived from social cognitive theory (such as modeling-videos, information, and tools to monitor own target behavior), indicates that Web-based trials should plan for a 50% dropout rate in the first month of the intervention [25]. In a 2016 study, close to every second patient did not log on more than once to a personal health record with self-management support and personal feedback for patients with T2DM. Only five of 132 participants used the eHealth self-management support program with goal setting and action planning functionality. Three out of these five took advantage of the personal feedback offered by the health psychologist [26].

Dropout and nonuse are thus major challenges in eHealth interventions, including those offering self-management support and personalized feedback. This makes it imperative to explore experiences of such interventions among people who drop out. To the best of our knowledge, no previous study has conducted qualitative interviews with participants who dropped out of an eHealth counseling intervention designed to support self-management for people with T2DM. The aim of this study was therefore to explore experiences with the eHealth intervention based on GSD from the perspectives of those who dropped out and to provide insight into their reasons.

XSL•FO
RenderX

Table 1. Overview of the Guided Self-Determination counseling for adults with type two diabetes and the reflection sheets.

Consultations	Focus	Reflection sheets
The first session at the GP[a]'s office	Preparing for subsequent consultations	Invitation to work together The HbA_{1c}[b]measurement
eConsultation 1	Your life with diabetes	RS[c]1a. Important events and periods in your life RS 1b. At present, what do you find difficult about living with diabetes? RS 1c. Unfinished sentences – your needs, values, habits and opportunities RS 1d. A picture, metaphor or expression of your life with diabetes
eConsultation 2	Focus for change	RS 2a. Room for diabetes in your life RS 2b. Your plans for changing your way of life
eConsultation 3	Work with changes	RS 3a. Clarification of challenge in your life with diabetes RS 3b. Previous problem-solving: thoughts, feelings, goals, and actions RS 3c. Dynamic problem-solving
eConsultation 4	Changes in daily life	RS 4a. Blood glucose self-monitoring and your reasons for self-monitoring RS 4b. New strategies and long-term plan for change RS 4c. Dynamic judgment of current and future problem solving RS 4d. «Pros and cons»

[a]GP: general practitioner.

[b]HbA_{1c}: glycosylated hemoglobin.

[c]RS: reflection sheet.

Textbox 1. The Web portal.

The secure messaging service was provided by the portal MinJournal. The secure messaging system at the portal demands login with electronic identification (BankID), providing the highest level of security (security level 4). Norwegian law requires this for Web-based sensitive information transfer, such as asynchronous communication between patients and health care personnel. This platform is already in use in Norwegian health care.

Methods

Design

We used a qualitative design and collected data by means of individual telephone interviews with participants who withdrew from the GSD eHealth intervention.

Description of the Guided Self-Determination (GSD) eHealth Intervention

General practice was chosen as an applicable intervention site because general practitioners (GPs) and registered nurses working with GPs are primarily responsible for health care for T2DM in Norway. The GSD eHealth intervention was delivered in addition to regular care. Regular care consists of structured annual consultations with a GP and nurse, as well as recommended routine measurement of glycosylated hemoglobin (HbA_{1c}) and consultations with a GP every 3-4 months, or individually adapted [20,27].

The aim of the GSD intervention was to support diabetes self-management. The participants answer questions on reflection sheets, and the themes addressed are then discussed with the nurse [28]. Table 1 shows an overview of the 4 eConsultations and topics of the 13 reflections sheets used in the GSD eHealth intervention for T2DM.

In this study, 4 trained nurses experienced in diabetes care at general practices delivered the GSD eHealth intervention over 12 to 35 weeks from August 2015 to April 2016. To establish a relationship, the nurse and patients initially met face-to-face at the GPs office. The nurse explained the aim of the GSD counseling, how to work with the reflection sheets (Table 1), and how to log on to the Web portal to use the secure messaging system (Textbox 1). All patients received a manual describing how to use the portal, the process of downloading and uploading portable document formats (PDFs) to the secure messages, how to fill out the reflection sheets, and send secure messages. After this initial meeting, the patients and nurses were to conduct 4 eConsultations, each consisting of 2 to 4 message exchanges. The patients were to complete the reflection sheets belonging to each eConsultation at home on their own electronic device, using their own words to express and reflect on their experiences and difficulties with diabetes management in daily life. They also formulated goals and plans for self-management. The reflection sheets were sent to their nurses via secure messages. The purpose of the reflection sheets were to facilitate situational reflection and improve communication to enable autonomous problem-solving, goal setting, and action planning (Table 1) [23].The nurses responded with written feedback to the participants' reflections.

Recruitment

Overall, 18 people invited by nurses at 4 general practices in southwestern Norway agreed to participate in the GSD eHealth intervention. However, 13 of these 18 eventually left the intervention. The nurses who conducted the intervention invited

XSL•FO
RenderX

the participants who had dropped out to take part in telephone interviews with a researcher. One person declined and 12 agreed.

Data Collection

Data were collected through telephone interviews in the spring of 2016. Telephone interviews are useful for collecting qualitative data and are considered less time- and energy-consuming for participants than face-to-face interviews [29,30]. The first author performed all interviews according to a semistructured interview guide. The main question invited the participants to speak freely and was expressed this way: "What was your experience with the GSD eHealth counseling intervention?" Supplementary questions were asked during the conversation to invite clarification and elaboration. Examples were "When and why did you quit the intervention?" "What were your expectations?" and "How did you experience written communication with your nurse via secure messaging?" The interviews lasted an average of 20 min, were audiotaped, and subsequently transcribed verbatim. In addition, demographic and clinical data were collected by a questionnaire, which the participants completed at the start of the intervention.

Data Analysis

The transcribed interviews were subjected to qualitative content analysis as described by Graneheim and Lundman [31]. All interviews were the unit of analysis and were read by 4 members of the research team at the beginning of the analysis process to attain a comprehensive understanding of the data. Meaning units responding to the aim of the study were identified and shortened but with core content preserved. The condensed meaning units were then labeled with tentative codes, after which categories were created by comparing and grouping codes according to similarities and differences. The categories were interpreted and abstracted into a main theme. Next, to strengthen the credibility of the analysis, the research team discussed and revised the codes, categories, and main theme several times until consensus was reached.

Ethical Considerations

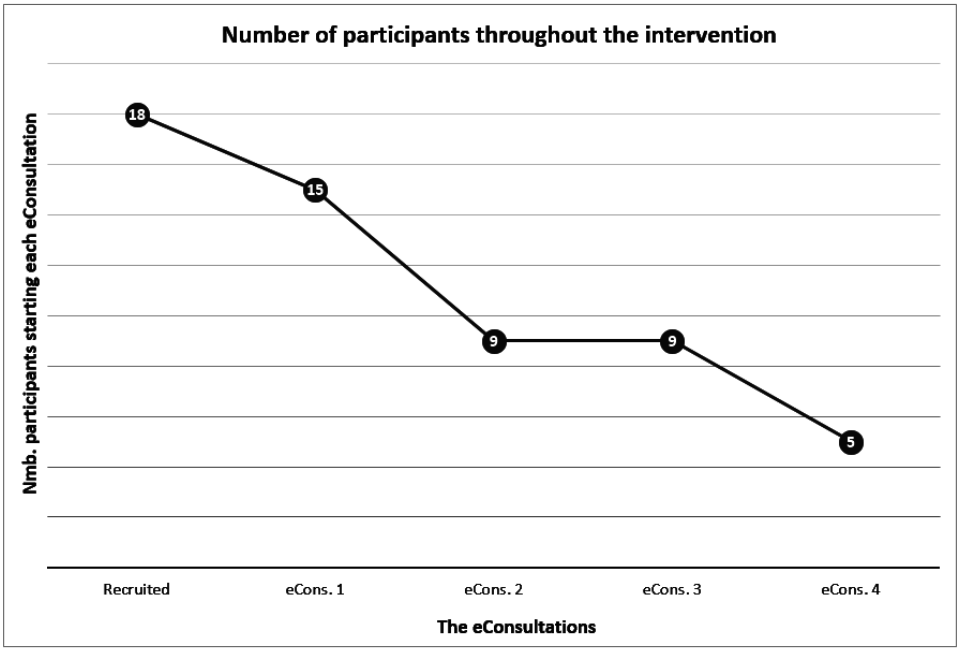
The Norwegian Regional Committee for Medical and Health Research Ethics (REK west No.2015/60) approved the study. All participants signed a written consent form and were guaranteed anonymity and the right to withdraw from the study at any time.

Results

Description of Participants

Participant characteristics are presented in Table 2. Of the 18 participants with T2DM recruited to the intervention, 14 were men and 4 were women. Of the 13 participants who dropped out, the majority (n=9) dropped out in the initial stage of the GSD eHealth intervention, before or during the first eConsultation. The last 4 participants withdrew during the third eConsultation (see Figure 1). Eleven of the 18 participants had an $HbA_{1c} \leq 7\%$, which is the expected treatment goal. The participants who dropped out from the intervention (n=13) did not differ considerably from those who completed the intervention (n=5). However, some small differences were detected; mean HbA_{1c} were 7.1% for the former and 7.7% for the latter. More men withdrew than women. All participants who regulated their diabetes with diet only withdrew from the intervention. Also, the median duration of diabetes was 9 years for those who dropped out and only 2 years for those who completed the intervention.

Figure 1. Dropout graph.



XSL•FO
RenderX

Table 2. Participant characteristics.

Demographics	All 18 participants recruited to the intervention	The 13[a]participants who dropped out of the intervention
Women (n)	4	2
Men (n)	14	11
Mean age (years, range)	55 (42-73)	57 (44-73)
Mean HbA_{1c}[b] (%, range)	7.3 (5.8-10.0)	7.1 (5.8-10.0)
Median diabetes duration (years, range)	9 (2-15)	9 (2-15)
Living situation (n)		
Alone	4	3
With family	14	10
Educational status (n)		
Higher education >4 years	1	0
Higher education <4 years	6	4
Upper secondary education	8	6
Primary school	3	3
Occupational status (n)		
Working full-time	15	10
Retirement pensioner	2	2
Receiver of disability benefit	1	1
Diabetes treatment (n)		
Diet	4	4
Oral or other medications	11	7
Insulin	3	2

[a]12 were interviewed in this study.

[b]HbA_{1c}: glycosylated hemoglobin.

Overview of Findings

The analysis resulted in identification of one theme related to experiences of the participants who dropped out of the GSD eHealth intervention: losing motivation for intervention participation. This theme described how motivation for participating in the intervention was influenced by some discouraging experiences. It was based on four categories: (1) frustrating technology, (2) perceiving the content as irrelevant and incomprehensible, (3) choosing other activities and perspectives, and (4) lacking face-to-face encounters. These categories are presented below and illustrated with quotations to facilitate transparency of interpretation. The quotations are attributed to the participants [P1-P12] to demonstrate their experiences and opinions.

Frustrating Technology

This category focuses on how participants felt frustrated by the technology used in this eHealth intervention. Initially, participants reported being receptive to participating in the GSD eHealth intervention. They valued the time and resource-saving potential of electronic communication with their nurse. However, they described difficulties in navigating the Web page due to errors with the portal and perceived the Web solution as time-consuming and tiring:

> *There was just too much trouble with it (the web page). In the end, I just gave up trying. Had it only been easier...* [P12]

Participants stated that it was cumbersome to download and save the PDFs before filling out the reflection sheets. They would have preferred completing the reflection sheets directly on the Web page. Participants also experienced Web page errors, for instance downtime, login problems, alerts from the firewall that it was an insecure Web page (which it was not), or that the nurse had not received the messages they sent. Some described being irritated and frustrated by technological problems. They pointed out that the Web solution bothered them when they were unable to send secure messages:

> *I answered the questions and tried to send, but it did not send. I tried several times, and I could not do it. This made the whole thing stressful for me...I bothered myself with it because I did not understand it and was not able to send anything. It was a bit silly, but it bothered me a lot, that I didn't get it...I feel like those kinds of things could be manageable, those forms,*

XSL•FO
RenderX

> *sending them. So I don't know what it was with this web page, why it didn't work.* [P2]

Although most participants experienced some challenges with the Web solution, some considered the problems minor. They said having to resend undelivered messages and change the browser to access the Web page were acceptable difficulties in an eHealth intervention.

Perceiving the Content as Irrelevant and Incomprehensible

Some participants did not see the content of the GSD as tailored to their needs and expectations for a diabetes self-management intervention. They expressed that they lost interest after reading some of the first issues raised in the reflection sheets because they could not familiarize themselves with these issues and did not consider the content relevant to their diabetes. As one participant noted:

> *I felt as if some constellations were made that I could not familiarize myself with. I live a completely normal life really; it's just the food, and the blood glucose level that makes me attend to it. But I have managed to adapt to the situation. And I keep adapting more gradually...I felt that it didn't suit me.* [P3]

The participants who reached the third eConsultation worked with reflection sheets intended to stimulate people to reflect on their goals and diabetes self-management behaviors. However, the purpose of these reflection sheets was described as difficult to understand:

> *When I came to "dynamic problem-solving" I started losing interest. I wondered: what do you want here? What method is this? I did not understand the purpose behind the form.* [P9]

Moreover, some of the participants stated that they did not fully understand what the intervention entailed when they signed up for it. Three of them said that they would prefer being able to send messages in free text to their nurse on their own schedule, instead of participating in a structured counseling intervention.

Choosing Other Activities and Perspectives

This category concerns the participants' narratives of more important priorities in their lives than the GSD eHealth intervention. Examples were other illnesses that needed more attention and other personal or work-related responsibilities. Daily life consisted of many complex tasks and commitments:

> *I am quite busy. I work full time and I really like to read. I have so much reading material, and I am active in politics as well. I have so much to read, so that just going online and having to spend much time there...It took too much of my time. Therefore, I felt it was a bit like...I didn't like that so much. I felt it took too much time.* [P11]

Going on the Web and engaging in the GSD eHealth intervention seemed to be considered less important than other matters requiring their attention, and the participants therefore chose to minimize their engagement with it:

> *It was the required time that did it. Some of the questions also, but that was not the main reason. It was more that it became a bit too much on top of everything else, having to sit down and spend time there, and remember to send and, yeah...There was too much else that had to be paramount somehow. Therefore, I simply had to downgrade it.* [P5]

Choosing not to focus on diabetes was also mentioned. Being uncomfortable with the issues raised in the reflection sheets or feeling pathologized by the demanding questions were articulated. Wanting to focus on living their life illustrates this perspective:

> *Because I feel healthy, and I do not want to be sick. But I am sick. Therefore I do have to look after it in the long run. But there is something in my head that I can't seem to get right...I have a diagnosis, but I do not run around being sick. I can explain some of this. My diet is what is wrong, or my life situation towards it (the diabetes). But I want to live as well. There is a limit there somewhere* [P9]

Lacking Face-to-Face Encounters

This category concerns the experience of lack of dialogue and a preference for face-to-face encounters with their nurse:

> *I would miss sitting down, see each other, and talk to each other. Because I'm not so into all the electronic communication. I really like to sit down and see the person I'm talking to.* [P4]

Meeting the nurse in person was emphasized as a motivating experience. One participant felt more obligated to try to reduce $HbA_{1c,}$ for example, when communicating with the nurse in person. Participants also stated that answering questions verbally was easier than writing down the answers, and that they would rather speak with the nurse in their regular consultations with the nurse. The following quotation illustrates this preference:

> *I think it is a lot better to sit and talk with her (the nurse) right in front of me. You know, and then we can discuss things and talk a little bit like that...And if there is any misunderstanding we can ask when we're sitting right next to each other.* [P8]

In addition, having eConsultations without a scheduled appointment with the nurse was considered less binding than regular health consultations:

> *It was allocating the time to it I had problems with...Although committing to answer, it does not have the same "disciplining" effect that one gets by meeting up at the doctor's office.* [P5]

At the same time, some participants emphasized that written messages could improve communication with the nurse by enabling carefully considered answers. They valued the ability to read and reflect upon the questions before answering:

> *The information you are able to provide about your health condition is much more thorough and better over the internet, when you sit and think through what you are going to answer and how to answer and that kind of thing. Than meeting up at the GPs office.* [P12]

XSL•FO
RenderX

Some of the participants insisted that they were accustomed to electronic and written communication. They appreciated the potential benefits of digital communication in health care, and some of them even preferred it, given they had the need for it. They mentioned that asynchronous digital communication could be time- and resource-saving. A combination of eHealth and regular encounters with the nurse was suggested as preferable when conducting the GSD, compared with merely written communication via secure messages.

Discussion

Principal Findings

This study provides insight into experiences with an eHealth intervention based on GSD from the perspective of those who dropped out and into their reasons for dropping out. Our findings indicate that the GSD eHealth intervention without face-to-face encounters influenced the participants' motivation for the intervention negatively and resulted in dropout. Other factors that diminished their motivation pertained to choosing other activities and perspectives in their lives, perceiving the content as irrelevant, and the technology as frustrating. We discuss these findings considering earlier research and in relation to the dimensions of autonomy, relatedness, and competence proposed by the SDT as important to develop and maintain autonomous motivation.

Comparison With Prior Work

Interventions With or Without Face-to-Face Encounters

Our findings indicate that participants missed face-to-face encounters with the nurse when communicating asynchronously via secure messages in the GSD eHealth intervention. They stated that they found it easier to discuss a variety of issues with the nurse and avoid misunderstandings when meeting face-to-face. Secure messages may have advantages for patient-nurse communication, such as efficient communication at convenient points of time in addition to the ability to think about the message before replying. However, our findings show the importance of acknowledging the drawbacks of written communication, such as the lack of nonverbal communication and the inability to ask immediate follow-up questions. Earlier research has demonstrated that support provided by clinicians via email enhanced adherence in eHealth interventions [32]. In contrast, our findings suggest that written communication alone is not experienced as motivating enough and that additional face-to-face encounters would have been preferred.

This could relate to the SDT, which proposes that a sense of relatedness is essential for motivation [21,33]. If people feel connected to their nurse in a warm, positive, and interpersonal manner, they may become more autonomously motivated to engage in health-related activities such as the GSD eHealth intervention [34]. Written communication via secure messages may not have been conducive to this sense of relatedness. Furthermore, we propose that our findings have some bearing on a previous study that suggests that the people with T2DM who presumably benefit the most from eHealth facilities actually use it the least [35]. This study furthermore suggests that patients' motivation to improve T2DM self-management is not sufficiently supported by eHealth facilities. This might have been the case for some of our participants. Combining eHealth with regular consultations has been suggested by earlier research as a promising way to improve engagement and reduce attrition [26]. Some of our participants also suggested that this would improve the GSD eHealth solution.

Moreover, our findings suggest that the current eHealth intervention was seen as less important when the participants had to engage in it on their own time and had no standing appointment with the nurse. This could reflect that asynchronous Web-based health consultations are regarded as less obligatory than regular health consultations with a scheduled appointment. This adds to findings from a recent study suggesting that planning for human support and interaction could be essential to upkeep motivation and use of digital interventions [36]. eHealth combined with regular consultations may be an important topic in future research, to facilitate the personal relationship between the participants and the health care personnel needed to motivate those who truly need and could benefit from self-management support interventions.

Lack of Perceived Value of the Intervention

Our findings indicate that participants had commitments that required more attention than diabetes and the GSD eHealth intervention. This was illustrated by narratives of other illnesses or daily responsibilities and competing life demands that required focus and reduced their motivation for participation. According to the SDT, the value people place on various activities affects their motivation [33]. Autonomous motivation is supported if people identify with behaviors or tasks, or place a value on projected results of behaviors [34]. If engaging in an eHealth intervention is not perceived valuable, people will not prioritize it. This intertwines our findings that when participants perceived the content irrelevant to their needs and expectations, the intervention was not perceived as valuable as other matters. Our findings relate to a previous investigation withdrawal from a telehealth intervention, revealing that the most frequent reason for withdrawal was that the participants did not perceive any benefit in using the telehealth service (eg, submitting their blood glucose readings to staff in local monitoring centers) [37]. One explanation for the lack of perceived value of the intervention is that some participants in our study said they already controlled their diabetes well, that they did not consider themselves as sick, or did not want to focus too much on diabetes in their daily lives. More than half of the participants had acceptable levels of HbA_{1c} prior to start, reaching the expected treatment goal of ≤ 7%. This could explain why they did not perceive a need for the intervention. Another explanation could be that even though their nurse deemed them suitable candidates for the intervention, they themselves did not want to put diabetes "up front." They were uncomfortable with, or regarded the issues raised in the reflection sheets as too demanding. Others preferred to focus on living their lives, not on the diabetes.

Patients' perspective of "wellness-in-the foreground" has been addressed in the shifting perspectives model, describing that people with chronic illness varies their attention of their disease [38]. Complex lives and competing priorities are important factors for developers to consider when designing "real-world"

XSL•FO
RenderX

eHealth interventions for diabetes self-management support, to create successful engagement strategies and approaches that are likely to reach and engage the target population.

Some participants did not see the relevance of the structured reflection sheets in the GSD eHealth intervention as relevant to them. This matter relates to the discussion of the consequences for motivation when an activity is not perceived as valuable enough and could indicate that the current intervention, with its complex aspects and delivery method, is not suitable for all participants. These findings can have two possible explanations. First, the reflection sheets address aspects of people's lives and emotions which may differ from what the participants are accustomed to and what they expect from communication with their nurse. The patients are asked to reflect on their challenges and make a plan for ideal problem solving (Table 1), which may differ from the traditional health care for people with diabetes, which are more concerned with education and information [7]. As the approach differs, it seems important to provide potential participants tailored information about the aim, the content, and the effort needed to remain engaged in the GSD intervention in order to recruit eligible participants who want to take part in and value such an intervention. Second, filling out reflection sheets electronically and communicating in writing could affect participants' perception of the purpose and value of the questions. The intervention aims to support each individual's autonomous goal setting and action planning [23], which are key features in self-management support interventions for people with diabetes. However, it was designed for face-to-face meetings. Perhaps the issues raised in the reflection sheets are so complicated that some participants would benefit from verbal explanation and discussion.

Technology

Previous research addresses technical problems as a continuous challenge in eHealth interventions resulting in high dropout rates [17,39]. Intelligible and user-friendly technology is imperative to maintain engagement and achieve benefits from digital health interventions [40]. Our findings concerning frustrating technology may therefore not be surprising. However, it is still important to address this issue, as most of our participants described difficulty with the technological solution. This finding may reflect that the demand for security level 4 (see Textbox 1) on patient-provider communication solutions is a barrier to engagement in such interventions. In addition, conducting the intervention depended on participants being able to download and upload PDFs to secure messages, which many participants found cumbersome. Our findings thus indicate that the eHealth technology offered in this study was not sufficiently user-friendly. Earlier research exploring patients' experiences with a diabetes self-management portal reveals technical challenges such as slow Internet access and time-consuming and difficult data entry as barriers to use. Improving the convenience of Web portals seems important to improve usability and reduce attrition [41]. Our findings add to this evidence, indicating that there is still a large potential for improvement in eHealth product design to ensure technology that patients will engage in and use. The frustrating technology may have thwarted the participants' sense of competence in managing the Web solution, and thus, reduced their engagement with the intervention. This points to the importance of facilitating more user-friendly but high security-level eHealth technology that would support users' sense of competence in managing the solution, and thus, increase their autonomous motivation for intervention engagement. However, experiencing a sense of competence supports autonomous motivation only when accompanied by self-determination [42]. This underlines the importance of creating successful engagement strategies and developing approaches that are likely to reach and engage the target population that can identify with or place a value on the projected results of engagement in the intervention.

Strengths and Limitations

The findings from this study may serve as a basis for future research aimed at broadening our understanding of the dynamics of withdrawing from eHealth interventions. However, generalizations from this small and situational study are not possible, nor are they intended. Out of 13 participants who dropped out of the intervention, 12 agreed to be interviewed. Although this could be considered a small sample, it is a strength of this study that most of the participants who dropped out were willing to be interviewed. The semistructured interview guide allowed the participants to express their genuine experiences, providing rich data. As the interviewer had no relationship with the participants, the participants might have felt more comfortable being candid. However, we cannot rule out the possibility that the nuances of face-to-face interaction are lost so that misleading information may not be detected [30]. Moreover, to reinforce the credibility of the data collection, the same researcher conducted all interviews. The findings and interpretations were discussed by a group of researchers, which also reinforced the credibility of the analysis.

A limitation that should be mentioned was the uneven gender distribution of the participants in this study. Initially, 14 men and 4 women were included, of which only 10 men and 2 women were interviewed. In relative terms, more men than women withdrew from the intervention. eHealth interventions may be used and experienced differently by men and women. A systematic literature review argues that there are gender differences in needs, preferences, and Web-based communication styles when engaging in Web-based health communication [43]. The dropout rate and the results of this study might have been different had we been able to include more women in the intervention. However, as this is a small sample, these are only speculations, and we cannot draw any definitive conclusions. Another limitation was interviewing only participants. Data from the study nurses about their experiences of conducting the intervention and their explanations concerning why patients left the intervention could have introduced other perspectives and improved our understanding of why some participants withdrew from the intervention.

Conclusions

Our findings indicate that the eHealth intervention based on GSD without face-to-face encounters with nurses reduced participants' motivation for engagement in the intervention. To maintain motivation, our study points to the importance of combining eHealth with regular face-to-face consultations. Our study also shows that the perceived benefit of the GSD eHealth

XSL•FO
RenderX

intervention intertwined with choosing to focus on other matters in complex daily lives are critical aspects in motivation for such interventions. This indicates the importance of giving potential participants tailored information about the aim, the content, and the effort needed to remain engaged in complex intervention so that eligible participants are recruited. Finally, motivation for engagement in the eHealth intervention was influenced by the technology used in this study. It seems important to facilitate more user-friendly but high-security eHealth technology. Our findings have implications for improving the eHealth intervention and to inform researchers and health care providers who are organizing eHealth interventions focusing on self-management support, in order to reduce dropout rates.

Acknowledgments

This study was conducted as a collaboration between Western Norway University of Applied Sciences and the University of Stavanger. It was funded by a grant from the Norwegian Research Council (Project No. 221065), University of Stavanger, and Western Norway University of Applied Sciences, Norway.

The authors would like to specially thank the patients participating in the study. In addition, we express our gratitude to the four study nurses and the involved general practices for recruiting the patients and conducting the intervention.

Authors' Contributions

SSL, BK, MG, and BO developed the study design. ERO contributed to the recruitment of participants and data collection. SSL performed the data collection, transcription, the tentative data analysis, and drafted the first version of the manuscript. BK, MG, and BO contributed to the data analysis. All authors contributed in editing the manuscript, and all authors contributed and agreed to the final draft of the article.

Conflicts of Interest

None declared.

References

1. van Gemert-Pijnen JE, Nijland N, van LM, 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;13(4):e111. [doi: 10.2196/jmir.1672] [Medline: 22155738]
2. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb;15(2):e35. [doi: 10.2196/jmir.2439] [Medline: 23410671]
3. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
4. Geraghty AW, Torres LD, Leykin Y, Perez-Stable EJ, Munoz RF. Understanding attrition from international internet health interventions: a step towards global eHealth. Health Promot Int 2012 Jul 10;28(3):442-452. [doi: 10.1093/heapro/das029]
5. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
6. IDF Diabetes Atlas. 2015. URL: http://www.diabetesatlas.org/ [accessed 2017-05-17] [WebCite Cache ID 6qXabQNCD]
7. Pillay J, Armstrong MJ, Butalia S, Donovan LE, Sigal RJ, Vandermeer B, et al. Behavioral programs for type 2 diabetes mellitus. Ann Intern Med 2015 Sep 29;163(11):848-860. [doi: 10.7326/M15-1400]
8. Heinrich E, Schaper N, de Vries N. Self-management interventions for type 2 diabetes: a systematic review. Eur Diabetes Nurs 2015 Feb 17;7(2):71-76. [doi: 10.1002/edn.160]
9. Oftedal B, Karlsen B, Bru E. Life values and self-regulation behaviours among adults with type 2 diabetes. J Clin Nurs 2010 Sep;19(17-18):2548-2556. [doi: 10.1111/j.1365-2702.2010.03243.x] [Medline: 20920080]
10. Snoek FJ. Breaking the barriers to optimal glycaemic control--what physicians need to know from patients' perspectives. Int J Clin Pract Suppl 2002 Jul(129):80-84. [Medline: 12166612]
11. Jenssen TG, Tonstad S, Claudi T, Midthjell K, Cooper J. The gap between guidelines and practice in the treatment of type 2 diabetes. Diabetes Res Clin Pract 2008 May;80(2):314-320. [doi: 10.1016/j.diabres.2007.12.025]
12. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Hess Fischl A, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Dia Care 2015 Jun 05;38(7):1372-1382. [doi: 10.2337/dc15-0730]
13. Haas L, Maryniuk M, Beck J, Cox CE, Duker P, Edwards L, et al. National standards for diabetes self-management education and support. Diabetes Care 2012 Dec 20;36(Supplement_1):S100-S108. [doi: 10.2337/dc13-S100]

XSL•FO
RenderX

14. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;3:CD008776. [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
15. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251. [doi: 10.1016/j.jdiacomp.2013.07.003]
16. Kelley H, Chiasson M, Downey A, Pacaud D. The clinical impact of ehealth on the self-management of diabetes: a double adoption perspective. J Assoc Inf Syst 2011;12(3):208-234 [FREE Full text]
17. Yu CH, Bahniwal R, Laupacis A, Leung E, Orr MS, Straus SE. Systematic review and evaluation of web-accessible tools for management of diabetes and related cardiovascular risk factors by patients and healthcare providers. J Am Med Inform Assoc 2012;19(4):514-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000307] [Medline: 22215057]
18. van Vugt M, de Wit M, Cleijne W, Snoek F. Use of behavioral change techniques in web-based self-management programs for type 2 diabetes patients: systematic review. J Med Internet Res 2013;15(12):e279. [doi: 10.2196/jmir.2800] [Medline: 24334230]
19. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of behaviour change theories. UK: Silverback Publishing; 2014.
20. Karlsen B, Oftedal B, Lie SS, Rokne B, Peyrot M, Zoffmann V, et al. Assessment of a web-based Guided Self-Determination intervention for adults with type 2 diabetes in general practice: a study protocol. BMJ Open 2016 Dec 13;6(12):e013026 [FREE Full text] [doi: 10.1136/bmjopen-2016-013026] [Medline: 27965253]
21. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. The European Health Psychologist 2008;10:2-5.
22. Oftedal B, Kolltveit BH, Zoffmann V, Hörnsten Å, Graue M. Learning to practise the guided self-determination approach in type 2 diabetes in primary care: a qualitative pilot study. Nurs Open 2017 Feb 10:- http://onlinelibrary.wiley.com/doi/10.1002/nop2.76/full. [doi: 10.1002/nop2.76]
23. Zoffmann V, Hörnsten Å, Storbækken S, Graue M, Rasmussen B, Wahl A, et al. Translating person-centered care into practice: a comparative analysis of motivational interviewing, illness-integration support, and guided self-determination. Patient Educ Couns 2016 Mar;99(3):400-407. [doi: 10.1016/j.pec.2015.10.015] [Medline: 26547303]
24. Angeles R, Howard M, Dolovich L. The effectiveness of web-based tools for improving blood glucose control in patients with diabetes mellitus: a meta-analysis. Can J Diabetes 2011 Jan;35(4):344-352. [doi: 10.1016/S1499-2671(11)54011-0]
25. Wangberg S, Bergmo T, Johnsen JAK. Adherence in internet-based interventions. Patient Prefer Adherence 2008;2:57-65.
26. van Vugt M, de Wit M, Sieverink F, Roelofsen Y, Hendriks SH, Bilo HJ, et al. Uptake and effects of the e-vita personal health record with self-management support and coaching, for type 2 diabetes patients treated in primary care. J Diabetes Res 2016;2016:5027356 [FREE Full text] [doi: 10.1155/2016/5027356] [Medline: 26955640]
27. Norwegian Directorate of Health. Helsedirektoratet. 2016. National treatment guideline for diabetes URL: https://helsedirektoratet.no/retningslinjer/diabetes [accessed 2017-05-18] [WebCite Cache ID 6qYXeiy5M]
28. Husted GR, Thorsteinsson B, Asbensen BA, Gluud C, Winkel P, Hommel E, et al. Effect of guided self-determination youth intervention integrated into outpatient visits versus treatment as usual on glycemic control and life skills: a randomized clinical trial in adolescents with type 1 diabetes. Trials 2014 Aug 12;15(321):1-12. [doi: 10.1186/1745-6215-15-321] [Medline: 25118146]
29. Mealer MJ, Jones J. Methodological and ethical issues related to qualitative telephone interviews on sensitive topics. Nurse Researcher 2014 Mar 27;21(4):32-37. [doi: 10.7748/nr2014.03.21.4.32.e1229]
30. Novick G. Is there a bias against telephone interviews in qualitative research? Res Nurs Health 2008 Aug;31(4):391-398 [FREE Full text] [doi: 10.1002/nur.20259] [Medline: 18203128]
31. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001]
32. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
33. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
34. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
35. Roelofsen Y, van Vugt VM, Hendriks SH, van Hateren KJ, Groenier KH, Snoek FJ, et al. Demographical, clinical, and psychological characteristics of users and nonusers of an online platform for T2DM patients (e-VitaDM-3/ZODIAC-44). J Diabetes Res 2016;2016:6343927 [FREE Full text] [doi: 10.1155/2016/6343927] [Medline: 26682232]
36. Aguilera A, Lyles CR. The case for jointly targeting diabetes and depression among vulnerable patients using digital technology. JMIR Diabetes 2017 Jan 17;2(1):e1. [doi: 10.2196/diabetes.6916]

XSL•FO
RenderX

37. Rixon L, Hirani S, Cartwright M, Beynon M, Selva A, Sanders C, et al. What influences withdrawal because of rejection of telehealth – the whole systems demonstrator evaluation. J Assist Technol 2013 Nov 29;7(4):219-227. [doi: 10.1108/JAT-06-2013-0017]
38. Paterson BL. The shifting perspectives model of chronic illness. J Nursing Scholarship 2001 Mar;33(1):21-26. [doi: 10.1111/j.1547-5069.2001.00021.x]
39. Verhoeven F, van GL, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007 Dec 14;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]
40. Lee C, Coughlin JF. PERSPECTIVE: Older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
41. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: a qualitative study of diabetes self-management portal. J Med Internet Res 2012;14(6):e158 [FREE Full text] [doi: 10.2196/jmir.2265] [Medline: 23195925]
42. Deci EL, Ryan RM. The support of autonomy and the control of behavior. J Pers Soc Psychol 1987;53(6):1024-1037 [FREE Full text]
43. Mo PK, Malik SH, Coulson NS. Gender differences in computer-mediated communication: a systematic literature review of online health-related support groups. Patient Educ Couns 2009 Apr;75(1):16-24. [doi: 10.1016/j.pec.2008.08.029] [Medline: 19233586]

Abbreviations

GPs: general practitioners
GSD: Guided Self-Determination
HbA1c: glycosylated hemoglobin
PDF: portable document format
SDT: self-determination theory
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 10.02.17; peer-reviewed by C Varsi, A Olinder; comments to author 02.03.17; revised version received 10.04.17; accepted 14.04.17; published 30.05.17.

Please cite as:
Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B
Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study
J Med Internet Res 2017;19(5):e187
URL: http://www.jmir.org/2017/5/e187/
doi:10.2196/jmir.7479
PMID:28559223

©Silje Stangeland Lie, Bjørg Karlsen, Ellen Renate Oord, Marit Graue, Bjørg Oftedal. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

High Level of Integration in Integrated Disease Management Leads to Higher Usage in the e-Vita Study: Self-Management of Chronic Obstructive Pulmonary Disease With Web-Based Platforms in a Parallel Cohort Design

Esther PWA Talboom-Kamp[1,2], MBA, MD; Noortje A Verdijk[1,2], MSc, PhD; Marise J Kasteleyn[1], MSc, PhD; Lara M Harmans[2], MSc; Irvin JSH Talboom[3]; Mattijs E Numans[1], MD, PhD; Niels H Chavannes[1], MD, PhD

[1]Public Health and Primary Care Department, Leiden University Medical Centre (LUMC), Leiden, Netherlands

[2]Saltro Diagnostic Centre, Utrecht, Netherlands

[3]Zorgdraad Foundation, Oosterbeek, Netherlands

Corresponding Author:
Esther PWA Talboom-Kamp, MBA, MD
Saltro Diagnostic Centre
Mississippidreef 83
Utrecht, 3565 CE
Netherlands
Phone: 31 302361170
Fax: 31 302361172
Email: e.talboom@saltro.nl

Abstract

Background: Worldwide, nearly 3 million people die of chronic obstructive pulmonary disease (COPD) every year. Integrated disease management (IDM) improves disease-specific quality of life and exercise capacity for people with COPD, but can also reduce hospital admissions and hospital days. Self-management of COPD through eHealth interventions has shown to be an effective method to improve the quality and efficiency of IDM in several settings, but it remains unknown which factors influence usage of eHealth and change in behavior of patients.

Objective: Our study, e-Vita COPD, compares different levels of integration of Web-based self-management platforms in IDM in three primary care settings. The main aim of this study is to analyze the factors that successfully promote the use of a self-management platform for COPD patients.

Methods: The e-Vita COPD study compares three different approaches to incorporating eHealth via Web-based self-management platforms into IDM of COPD using a parallel cohort design. Three groups integrated the platforms to different levels. In groups 1 (high integration) and 2 (medium integration), randomization was performed to two levels of personal assistance for patients (high and low assistance); in group 3 there was no integration into disease management (none integration). Every visit to the e-Vita and Zorgdraad COPD Web platforms was tracked objectively by collecting log data (sessions and services). At the first log-in, patients completed a baseline questionnaire. Baseline characteristics were automatically extracted from the log files including age, gender, education level, scores on the Clinical COPD Questionnaire (CCQ), dyspnea scale (MRC), and quality of life questionnaire (EQ5D). To predict the use of the platforms, multiple linear regression analyses for the different independent variables were performed: integration in IDM (high, medium, none), personal assistance for the participants (high vs low), educational level, and self-efficacy level (General Self-Efficacy Scale [GSES]). All analyses were adjusted for age and gender.

Results: Of the 702 invited COPD patients, 215 (30.6%) registered to a platform. Of the 82 patients in group 1 (high integration IDM), 36 were in group 1A (personal assistance) and 46 in group 1B (low assistance). Of the 96 patients in group 2 (medium integration IDM), 44 were in group 2A (telephone assistance) and 52 in group 2B (low assistance). A total of 37 patients participated in group 3 (no integration IDM). In all, 107 users (49.8%) visited the platform at least once in the 15-month period. The mean number of sessions differed between the three groups (group 1: mean 10.5, SD 1.3; group 2: mean 8.8, SD 1.4; group 3: mean 3.7, SD 1.8; P=.01). The mean number of sessions differed between the high-assistance and low-assistance groups in groups 1 and 2 (high: mean 11.8, SD 1.3; low: mean 6.7, SD 1.4; F1,80=6.55, P=.01). High-assistance participants used more services

XSL•FO
RenderX

(mean 45.4, SD 6.2) than low-assistance participants (mean 21.2, SD 6.8; F1,80=6.82, *P*=.01). No association was found between educational level and usage and between GSES and usage.

Conclusions: Use of a self-management platform is higher when participants receive adequate personal assistance about how to use the platform. Blended care, where digital health and usual care are integrated, will likely lead to increased use of the online program. Future research should provide additional insights into the preferences of different patient groups.

Trial Registration: Nederlands Trial Register NTR4098; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4098 (Archived by WebCite at http://www.webcitation.org/6qO1hqiJ1)

(J Med Internet Res 2017;19(5):e185) doi:10.2196/jmir.7037

KEYWORDS

COPD; eHealth; self-management; integrated disease management; self-efficacy; Web-based platform; primary care; chronically ill; blended care

Introduction

Chronic obstructive pulmonary disease (COPD) is a slowly progressive lung disease and one of the main causes of morbidity and mortality in high-, middle-, and low-income countries [1]. Worldwide, nearly 3 million people die of COPD every year; in 2012, this was equal to approximately 6% of all deaths globally [2,3].

According to current COPD guidelines, symptoms and airflow obstruction should be monitored regularly to guide modification of treatment and for early identification of complications [4,5]. Routine monitoring should contribute to achieving management goals in COPD (ie, to delay disease progression and alleviate its manifestations). The most important primary care objective should be to improve quality of life (QoL) [6].

In the past decade, integrated disease management (IDM) was introduced as a means of improving quality of care. An IDM program for COPD consists of different components of care in which various health care providers cooperate on education, exercise, behavioral therapy, smoking cessation, medication, nutrition advice, and follow-up. The responsibility for the program lies largely with the health care professional, with a modest role for the patient. For people with COPD, IDM not only improves disease-specific QoL and exercise capacity, but it can also reduce hospital admissions and hospital days per person [7].

Self-Management of Chronic Obstructive Pulmonary Disease

Self-management of COPD has been introduced as an effective method to improve the quality and efficiency of IDM, and to reduce health care costs [8-10]. Interventions to support self-management have shown reductions in hospital admissions and fewer sick days as a result of exacerbations [11,12]. The core components of self-management include education, eliciting personalized goals, psychological coping strategies, formulating strategies to support adherence to treatment, and behavioral change, together with practical and social support [13,14]. Chronically ill patients who have experience with person-centered, high-quality chronic illness care that focuses on patient activation, decision support, goal setting, problem solving, and coordination of care are better self-managers [15]. Self-efficacy explores the emotional functioning and coping skills of an individual and is thought to be a strong predictor of health behavior of COPD patients; the General Self-Efficacy Scale (GSES) tool is a reliable and sensitive measure of self-efficacy for patients with COPD [16].

eHealth Interventions

Generally, eHealth interventions are effective in stimulating self-management because they allow patients to better cope with their illness at the time/place of their choosing, enabling them to adapt their lifestyle to their condition, while reducing medical staff consultations [17]. The deployment of eHealth apps facilitates accessibility to health care, which enhances patients' understanding of their disease, their sense of control, and willingness to engage in self-management [18,19]. Although patients' attitudes and receptiveness toward eHealth apps are promising in persons of a certain age and education level [20-22], large-scale adoption of self-management and eHealth in daily practice lags behind expectations [23].

Previous eHealth studies have revealed the challenges, barriers, and factors that make successful implementation difficult, yet many questions remain unanswered. Moreover, a major challenge of eHealth in care coordination is to make it beneficial and easy to use for health care providers *and* patients, otherwise neither will use it [24]. Also, online self-management support needs to be a fully integrated element of IDM. For example, in a Dutch study on adherence to an online self-management app for patients with COPD or asthma, patients tended to use the online application on a regular basis when the health care provider was involved, whereas patients without assistance used the app only sporadically [25]. For barriers related to clinicians, the eHealth evidence base needs strengthening, whereas for primary care practices a learning process (including staff training) needs to be instituted [26]. In addition, it is necessary to more adequately inform patients about the possibilities and consequences of eHealth [27]. Furthermore, poor user-friendliness of Web-based apps and the lack of "push" factors (eg, automated reminders or messages from health care professionals) are a common cause of low usage or decline in usage of Web-based apps [28]. In any eHealth study, a substantial proportion of users drop out before completion, or stop using the app; thus, attrition is a common problem and should be analyzed to provide data for real-life adoption problems [29]. Studies on the use of online self-management

XSL•FO
RenderX

show that attrition tends to start when users "get lost" in the intervention [28,30].

Preconditions for starting eHealth are (1) it must be well organized within usual care (organizational perspective), (2) it should be beneficial and easy to use for patients (human perspective), and (3) the apps have to be technically sound (technical perspective).

Design of e-Vita COPD

Because low usage of eHealth is an ongoing problem, we designed a multilevel study to investigate the implementation of a self-management Web platform to support patients with COPD in primary care [31]. Because the Web platform provides continuous education and contact with health care professionals, it is expected to help patients better recognize and self-manage exacerbations in an early phase, thereby helping to stabilize their health status.

This study, called "e-Vita COPD," compares three different approaches to incorporating eHealth via Web-based self-management platforms into the integrated disease management of COPD using a parallel cohort design . Also, participants are randomly allocated in two of the cohorts (1 and 2), using the same platform to different levels of personal assistance. All three cohorts incorporated the platforms to different levels; the two levels of personal assistance for patients were a group with high assistance and a group with low assistance. The main aim is to analyze the factors that successfully promote the use of a self-management Web platform for patients with COPD.

From an organizational perspective, our hypothesis is that a self-management Web platform will be better adopted if the platform is an integrated part of IDM, with trained health care professionals who encourage patients to use the platform. From a human perspective, our hypothesis is that a self-management platform will be better adopted by patients if they receive sufficient personal assistance in how to use the platform, and will be better adopted by patients with a higher level of self-efficacy (assessed by GSES) and a higher educational level. From a technical perspective, our hypothesis is that a self-management Web portal will be better adopted if the platform is easy to use and has practical content.

Methods

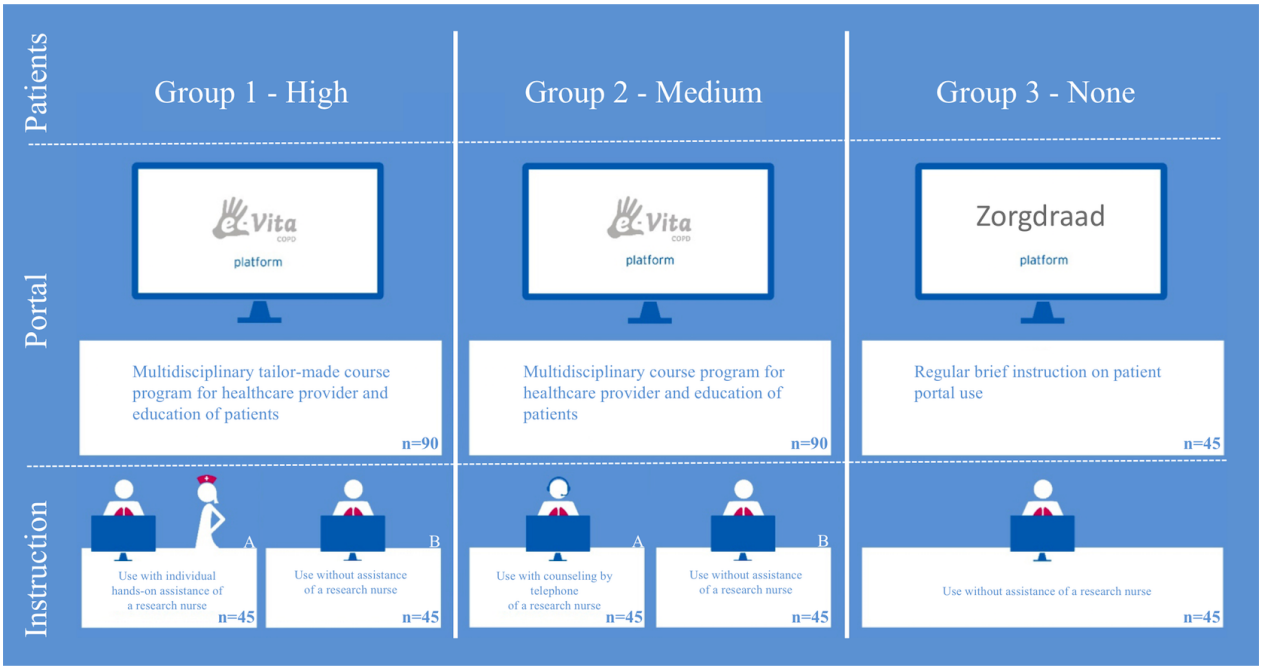
Study Design

For this study, we used the CONSORT-EHEALTH checklist that describes how the intervention should be reported [32].

We designed a quality improvement intervention and chose an implementation study [33]. We designed a method to promote the uptake of our research findings into routine primary health care; with this design, we aimed at studying the influences on health care professionals and patient behavior and at evaluating the process by which these effects are achieved.

This research combined different study methods to investigate organizational implementation methods and the net benefits of eHealth interventions from a human, organizational, and technical viewpoint. Full methodological details were reported previously [31]. Figure 1 presents an overview of the combined study design with organizational and technical differences.

Figure 1. Design of the e-Vita COPD Study. High, medium, and none refer to the level of integration of the web platform into the patient's integrated disease management program. A: high assistance; B: low assistance.



XSL•FO
RenderX

Included in the study were three different care groups (groups 1-3) and two Web portals (e-Vita and Zorgdraad). In group 1, the online e-Vita platform was offered as a highly integrated part of the COPD IDM with a tailor-made intensive course program on COPD and eHealth for health care providers and patients. Group 2 had a medium level of integration with a standard basic course program for health care providers and patients. The COPD patients in groups 1 and 2 who agreed to use the platform were invited by practice nurses for intake in which they defined a personal health goal together and discussed how and why to use the platform. In group 3, the online platform was offered without integration into disease management; health care providers and patients received instructions from the Web platform itself and received no training on COPD care.

Therefore, group 1 (high) had a high level of integration of the Web platform into their IDM program, group 2 (medium) had a medium level of integration into their IDM program, and group 3 (none) had no integration.

Two different levels of assistance for patients were distinguished within group 1 (high) and 2 (medium): one with a high level of personal assistance and one with a low level of personal assistance. Patients in groups 1 and 2 were randomly subdivided into two groups with high and a low levels of support. Randomization was carried out stratified on the care group (1:1) by computer. In group 1 (high integration-high assistance), a high level of support consisted of two home visits to patients by a research nurse who assisted in the use of the Web portal. In group 2 (medium integration-high assistance), a high level of support consisted of telephone consultation between the patient and a research nurse who explained use of the Web portal. In the low-assistance groups of groups 1 and 2, low-level support consisted of a primary care nurse showing the patient only one time how to use the Web platform, without any follow-up instruction. Patients in group 3 that used the online self-management platform (called “Zorgdraad”) had no active support from health care professionals or research nurses.

Both platforms were provided for the intervention period of 15 months.

Participants

Three health care groups participated in this study. Patients were eligible if they were diagnosed with COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria (postbronchodilator forced expiratory volume in first second/forced vital capacity <0.7) in accordance with the COPD Guidelines of Dutch general practitioners (GPs) [34] and they were being treated for COPD in primary care. The study aimed to be inclusive rather than exclusive, to achieve high external validity (applicability to daily practice). Patients were excluded if they were unable to fill in questionnaires, had no access to the Internet, had a terminal illness, were immobile, or were severe substance abusers .

Recruitment of Patients and Nonparticipation Analysis

We started by recruiting the primary care groups (groups 1-3). Healthcare professionals decided to participate in this study mainly because they wanted to join a project that offered possible health care improvement. In group 1, this was 19 of 170 GPs (11%); in group 2, 29 of 34 GPs (84%); and in group 3, all 27 GPs (100%).

Patients were invited to participate by letter. When they refused to participate, we defined them as nonparticipants. When participants in the e-Vita study logged in and used the Web platform at least once, we defined them as “users.” Patients were defined as “lost to follow-up” if they did not log on to the platform after signing informed consent and if they did not complete the whole intervention period.

Intervention

The interventions in the three groups consisted of a self-management program including different levels of education for health care professionals, different levels of integration in the COPD care program, and different levels of personal assistance for patients. We used two Web-based platforms (e-Vita in groups 1 and 2, Zorgdraad in group 3) that were very similar, with the same basic features and functionalities. The education, the care programs, and the platforms were specifically designed for COPD patients; their needs and wishes were processed.

The online self-management platform e-Vita is an initiative of the Dutch foundation “Care Within Reach” [35]. The content was created by experts guided by interviews with COPD patients about their thoughts/feelings related to living with COPD and its treatment; the experiences of health care professionals related to the treatment of patients with COPD were also integrated. The main content of the platform consists of insight into personal health data, self-monitoring of health values, education, and a coach for attaining personal goals. The first release of the platform was in January 2014 with an update in May 2015 (which was during the intervention period).

All patients in groups 1 and 2 that used the e-Vita platform had access to a telephonic and digital helpdesk to address any problems. Patients in groups 1 and 2 that used e-Vita received automated online reminders via email from our research team for the self-reported questionnaires and messages.

atients in group 3 used the online self-management platform Zorgdraad (an initiative of the Dutch foundation Zorgdraad and the diagnostic center Saltro). The content was created by experts guided by their experience in treating patients with COPD. The main content of the platform is basic and consists of insight into personal health data, self-monitoring of health values, and education. All patients that used Zorgdraad received automated online reminders via email for the self-reported questionnaires.

System and Services

When logging on (username plus password) for the first time to e-Vita or Zorgdraad, every user saw a pop-up with a brief explanation about e-Vita or Zorgdraad and the services that could be found on the website. After the pop-up, the user was directed to the home page. From there, users were able to access all functionalities of the platform. The log-on procedure of both platforms is based on Dutch security legislation and guidelines (the Dutch Personal Data Protection Act).

XSL•FO
RenderX

The e-Vita platform (Multimedia Appendices 1 and 2) consisted of the following set of interrelated services, which could be accessed via the home page:

1. An online coach for guidance when working on personal goals and planning of the personal actions.
2. Self-monitoring personal health values and self-reported questionnaires, in which users could register the values they measured for the disease-specific health status Clinical COPD Questionnaire (CCQ) [36], the modified Medical Research Council scale (MRC scale) for dyspnea [37], the GSES [38], sociodemographic characteristics, and EuroQol five-dimensions questionnaire on quality of life (EQ-5D) [39].
3. An education module with text and movies about COPD.
4. Extra information about COPD.
5. Information about the team of health care professionals
6. A module to send and receive messages to the health care professionals.

The Zorgdraad platform (Multimedia Appendix 3) has very basic usability and content, with the following set of interrelated services:

1. Self-monitoring personal health values and self-reported questionnaires, where users can register the same values as in the e-Vita platform.
2. A basic education module with text and movies about COPD.
3. Information about the team of health care professionals.
4. A module to send and receive messages to the health care professionals.

Outcome Measures

Full details on outcome measures were reported previously [31]. The primary outcome of this study was the usage of the online self-management platform: we defined usage as every visit to e-Vita and Zorgdraad that was tracked objectively by collecting log data in log files. We focused on the log data in the intervention period of 15 months. The usage was divided into sessions and services; a *session* was defined as the total period between logging in and logging out of the system and a *service* was defined as a focused action within the system, as described for both platforms previously. For every action in e-Vita and Zorgdraad (button clicks, page views, and database transactions), the following information was collected by the Web server and added to a log file: (1) the users' identification number, (2) time and day of the session, (3) the type of actions (services) taken, and (4) optional additional information about the actions (services). For these analyses, it was important to investigate not only the amount of use, but also the user's patterns.

Baseline characteristics were extracted from the log files of e-Vita and Zorgdraad, including age, gender, education level, and scores on the CCQ (range 0=very good health status to 6=extremely poor health status), MRC scale (range 1-5), EQ-5D, and GSES.

Data were collected that could plausibly be related to the study outcomes. In the analyses, the organizations of GPs (care groups PreventZorg, Zorg op Noord, and Leidsche Rijn Julius Gezondheidscentra), integration in IDM (integrated vs not integrated), and personal assistance for the participants (personal assistance vs no assistance) were used as determinants. Education was self-reported using eight response categories and converted into three levels based on the International Standard Classification of Education (ISCED): high (bachelor, master, doctor), medium (secondary and tertiary education), and low (no formal education, primary education) [40]. Self-efficacy was measured with the GSES in a self-reported 10-item questionnaire. Total scores ranged from 0 to 10.

An attrition curve was drawn with the nonusage attrition; the percentage of users who used the platforms were plotted over time.

Statistical Methods

For the nonparticipation analysis, differences in age and gender between participants and nonparticipants were compared using a chi-square test and a Mann-Whitney *U* test after normality tests.

Categorical baseline characteristics were reported as numbers and percentages, normally distributed continuous variables as means with standard deviations (SD), and nonnormally distributed variables as medians with interquartile ranges (IQRs). Characteristics between the three groups were explored using chi-square tests and Kruskal-Wallis tests.

To predict the use of the e-Vita and Zorgdraad platforms, multiple linear regression analyses (mean number of sessions/services/mean session time/mean number of services per session) for the different independent variables were performed:

1. Integration in integrated disease management including training by health care professionals by comparing groups 1 (high), 2 (medium), and 3 (none);
2. Personal assistance given to patients by comparing the high-assistance groups of groups 1 (high) and 2 (medium) versus the low-assistance groups of groups 1 (high) and 2 (medium);
3. Patients' scores on the GSES; and
4. Patients' educational levels.

We analyzed the main effects because, theoretically, we presumed no interaction between factors. All analyses were adjusted for age and gender.

Attrition was measured by logging and evaluating the percentage of users that used one of the platforms every month during the intervention period plus three months. The area under the curve was calculated for a period of 18 months; after this period, usage dropped to zero for two groups. We used a calculating program to measure the area under the curve using definite integrals.

All analyses were performed with SPSS version 22.0 (IBM Corporation, Armonk, NY, USA).

Results

In total, 942 diagnosed COPD patients from the medical files of the three care groups were selected to be eligible for the study (Figure 2). The GPs of these care groups excluded 240 COPD patients from participation due to (1) other diseases, (2)

XSL•FO
RenderX

treatment in hospital, and (3) probably incompetent to participate in the program. In the end, 702 COPD patients were invited to start with the e-Vita platform; of these, 215 (30.6%) agreed to register and provided informed consent. Reasons for declining to participate are presented in Figure 2.

Of the 215 COPD patients included at baseline, 82 were in group 1 (high), 96 in group 2 (medium), and 37 in group 3 (none).

After randomization, 36 patients in group 1 were allocated to high-assistance group, 46 to low-assistance group; in group 2 44 patients were randomized to high-assistance group, and 52 to low-assistance group.

The total number of patients lost to follow-up was 132. Figure 2 shows the reasons for drop out in groups 1 and 2; patients in group 3 were not asked for their reasons. Of the 215 participants, 107 (49.8%) patients were identified as platform users: 43 in group 1, 42 in group 2, and 22 users in group 3 (Figure 3).

Figure 2. Flowchart of inclusion of participantss in the e-Vita COPD study. High, medium, and none refer to the level of integration of the web platform into the patient's integrated disease management program. A: high assistance; B: low assistance.

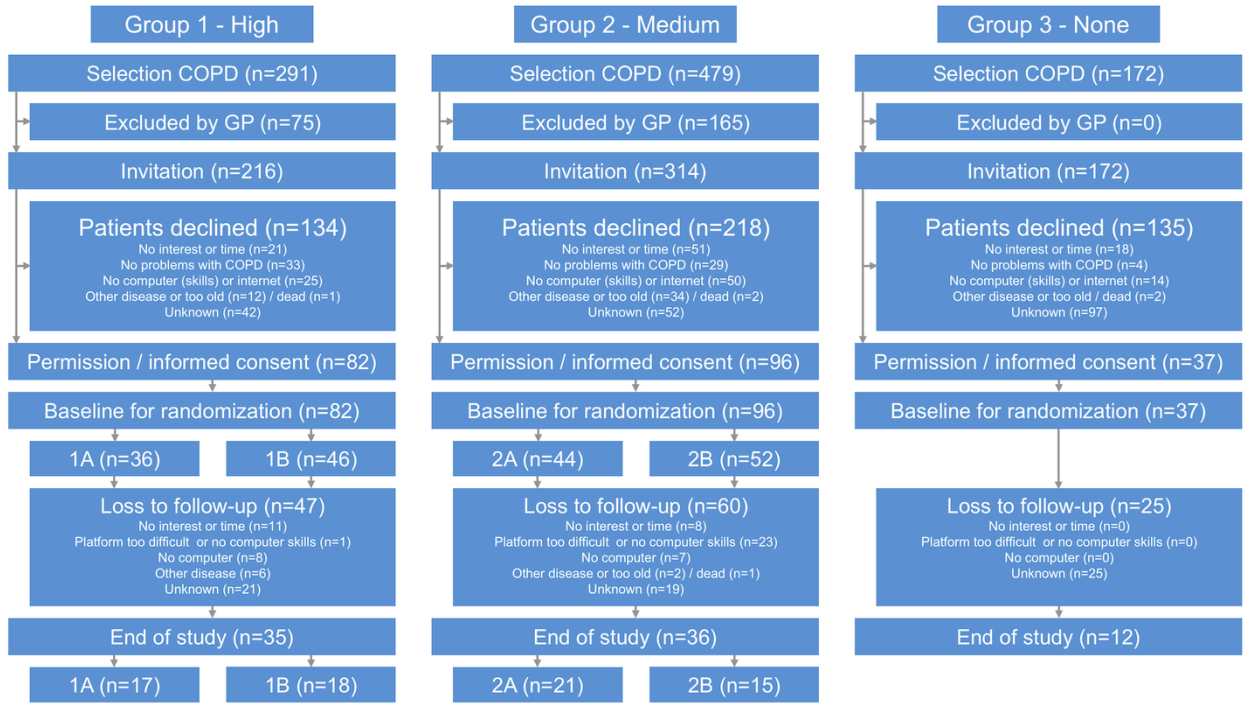
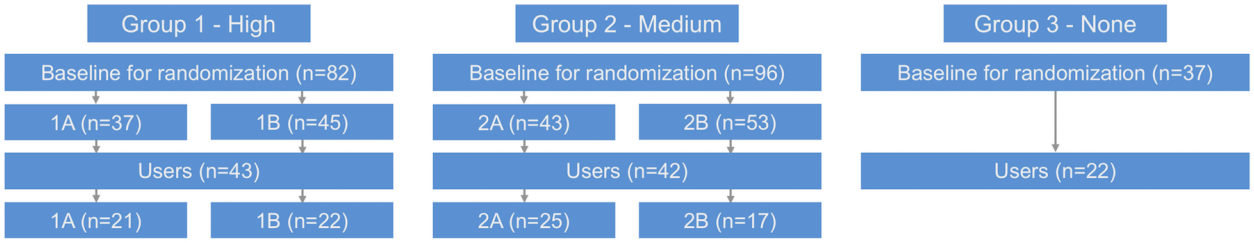


Figure 3. Flowchart of the platform users of the e-Vita COPD study. High, medium, and none refer to the level of integration of the web platform into the patient's integrated disease management program. A: high assistance; B: low assistance.



XSL•FO
RenderX

Table 1. Characteristics and comparison of participants and nonparticipants of the e-Vita study.

Nonparticipants/participants	Group 1 (high)	Group 2 (medium)	Group 3 (none)	Total
Nonparticipants				
n	209	383	135	727
Age (years), median (IQR)	69.3 (61.1-77.5)	69.8 (61.6-78.5)	66.3 (60.3-74.4)	68.5 (61.2-77.9)
Males, n (%)	108 (51.7)	167 (43.6)	61 (46.2)	336 (46.2)
Participants[a]				
n	82	96	37	215
Age (years), median (IQR)	66.3 (61.1-75.7)	67.3 (62.6-76.6)	64.1 (61.5-69.2)	66.6 (61.4-74.7)
Males, n (%)	51 (62.2)	41 (42.7)	20 (54.1)	112 (52.1)

[a]Baseline for randomization.

Table 2. Baseline demographic and clinical characteristics of the patients with COPD included in the e-Vita study.

Participants	Group 1 (high)		Group 2 (medium)		Group 3 (none)	Total
	High	Low	High	Low		
n	36	46	44	52	37	215
Age (years), (IQR)	66.3 (61.0-79.2)	65.6 (61.3-73.4)	68.7 (64.0-78.3)	66.8 (60.3-75.1)	64.1 (61.5-69.2)	66.6 (61.4-74.7)
Male, n (%)	19 (52.8)	32 (69.6)	17 (38.6)	24 (46.2)	20 (54.1)	112 (52.1)
Education level, n (%)						
Low	4 (28.6)	8 (38.1)	5 (22.7)	8 (42.1)	7 (53.8)	32 (36.0)
Medium	7 (50.0)	8 (38.1)	11 (50.0)	8 (42.1)	4 (30.8)	38 (42.7)
High	3 (21.4)	5 (23.8)	6 (27.3)	3 (15.8)	2 (15.8)	19 (21.3)
Questionnaire, median (IQR)						
CCQ	1.0 (0.6-1.9)	1.2 (0.8-1.6)	1.3 (0.9-2.1)	1.4 (1.1-2.1)	1.3 (0.6-1.8)	1.2 (0.8-1.9)
mMRC scale	1.0 (1.0-3.0)	1.0 (1.0-2.0)	2.0 (1.0-3.0)	2.0 (1.0-2.0)	1.0 (1.0-1.0)	1.0 (1.0-2.0)
GSES	3.4 (3.1-3.7)	3.3 (3.0-3.8)	3.3 (2.8-3.5)	3.3 (3.1-3.7)	3.4 (3.3-3.7)	3.3 (3.0-3.7)
EQ-5D	0.85 (0.7-1.0)	0.89 (0.81-1.0)	0.85 (0.72-1.0)	0.84 (0.71-1.0)	0.9 (0.84-1.0)	0.86 (0.78-1.0)

Nonparticipation Analysis

The age and gender of participants and nonparticipants are presented in Table 1. Participants and nonparticipants did not differ with regard to gender (52.1%, 112/215 male vs 46.2%, 336/727 male, *P*=.13) or age (median 66.6, IQR 61.4-74.7 vs median 68.5, IQR 61.2-77.9 years, *P*=.20). Because only a few nonparticipants filled in a questionnaire on CCQ, the mean CCQ could not be determined for nonparticipants.

Baseline Characteristics of Patients

Table 2 presents the baseline demographic and clinical characteristics of the included COPD population (median age 66.6 years; 52.1% was male). These patients had mildly symptomatic COPD which is reflected by a median MRC scale of 1.0 and a median CCQ of 1.2. Of all participants, 89 of 215 (41.4%) filled in the online questionnaire for education level; most participants had a middle education level (42.7%). The median GSES was 3.3 and the median EQ-5D 0.86. The characteristics age (χ^2_2=5.4, *P*=.07), education level (χ^2_4=2.2, *P*=.70), GSES (χ^2_2=1.74, *P*=.42), and EQ-5D (χ^2_2=2.4, *P*=.28) were similar in the three groups. There was a difference in the characteristics gender (χ^2_2=6.8, *P*=.03), with more male patients in group 1; and a difference in CCQ (χ^2_2=6.5, *P*=.04) and MRC scale (χ^2_2=11.3, *P*=.003) with a higher CCQ and MRC scale in group 2.

Use of the Online e-Vita and Zorgdraad Platforms

Table 3 presents the 15-month usage pattern by patients using the log files of e-Vita and Zorgdraad. In total, 107 users visited the platform at least once in the 15-month period. The helpdesk received 101 calls; most questions concerned problems with the log-on procedure.

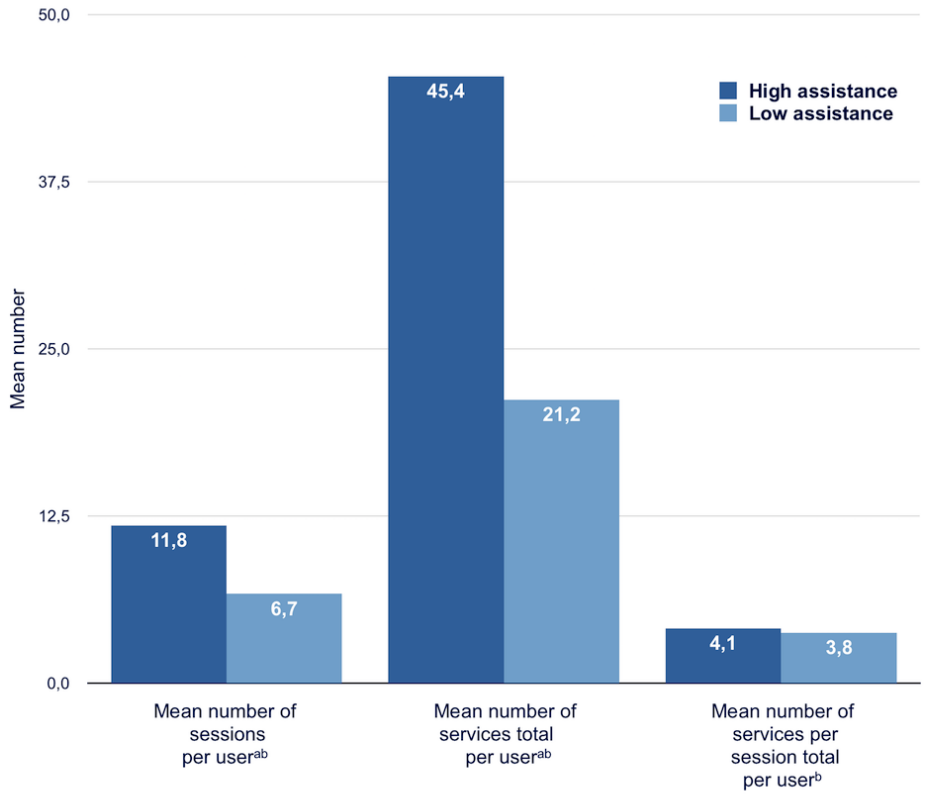
XSL•FO
RenderX

Table 3. Usage patterns for groups 1, 2, and 3.

Usage[a]	Group 1 (high)	Group 2 (medium)	Group 3 (none)	Total	*P*	$F_{2,100}$
n	43	42	22	107		
Sessions, mean (SD)	10.5 (1.3)	8.8 (1.4)	3.7 (1.8)	8.2 (8.7)	.01	4.68
Session time (minutes), mean (SD)	3.5 (0.7)	4.8 (0.7)	6.7 (0.9)	4.8 (4.3)	.03	3.83
Total services per user, mean (SD)	45.2 (6.1)	27.9 (6.2)	6.7 (8.3)	28.8 (41.0)	.001	7.18
Total services per session per user, mean (SD)	3.9 (0.4)	4.1 (0.4)	2.1 (0.6)	3.6 (2.8)	.02	3.97

[a] Adjusted for age and gender.

Figure 4. Usage patterns for high and low levels of assistance. a: mean number significantly higher in high assistance group b: adjusted for age and gender.



In the 15-month intervention period, the total number of sessions for the e-Vita platform was 830 (9.8 per user) and for the Zorgdraad platform 78 (3.5 per user). The mean number of sessions differed between the three groups ((group 1: mean 10.5, SD 1.3; group 2: mean 8.8, SD 1.4; group 3: mean 3.7, SD 1.8; *P*=.01) (Table 3). In groups 2 and 3, the mean session time was higher than in group 1; the difference between the three groups was significant (*P*=.03). In groups 1 and 2, the mean number of services in total was higher than in group 3 (group 1: mean 45.2, SD 6.1; group 2: mean 27.9, SD 6.2; group 3: mean 6.7, SD 8.3; *P*=.001) and the number of services per session in groups 1 and 2 was higher than in group 3 (group 1: mean 3.9, SD 0.4; group 2: mean 4.1, SD 0.4; group 3: mean 2.1, SD 0.6; *P*=.02).

Figure 4 shows use of the e-Vita for the two groups with high and low assistance; higher usage of the platform was related to a higher level of personal assistance.

The mean number of sessions differed between the high-assistance groups and the low-assistance groups in groups 1 and 2 (high assistance: mean 11.8, SD 1.3; low assistance: mean 6.7, SD 1.4; $F_{1,80}$=6.55, *P*=.01). Participants in the high-assistance groups used more services (mean 45.4, SD 6.2) than participants in the low-assistance groups (mean 21.2, SD 6.8; $F_{1,80}$=6.82, *P*=.01). In the high-assistance groups, the mean number of services per session did not differ from the low-assistance groups (mean 4.1, SD 0.4 vs mean 3.8, SD 0.5; $F_{1,80}$=0.36, *P*=.55).

XSL•FO
RenderX

An overview of the online platform services visited during the intervention period is provided in Figure 5; for every service, the mean number per user is depicted. The log files revealed that all services were mainly used by group 1. Furthermore, it revealed that the e-Vita and Zorgdraad Web platforms were predominantly used for online questionnaires, general information, and depicting wishes/goals related to their lifestyle, and to a lesser extent for online education, visiting the library, and looking for information about their health care professionals. Log files also showed that there was almost no interest in the measurement values CCQ and MRC scale. The email feature and an explanation of the test results of e-Vita were used to a moderate extent.

Figure 5. Usage patterns of the mean number of services per user in each group.

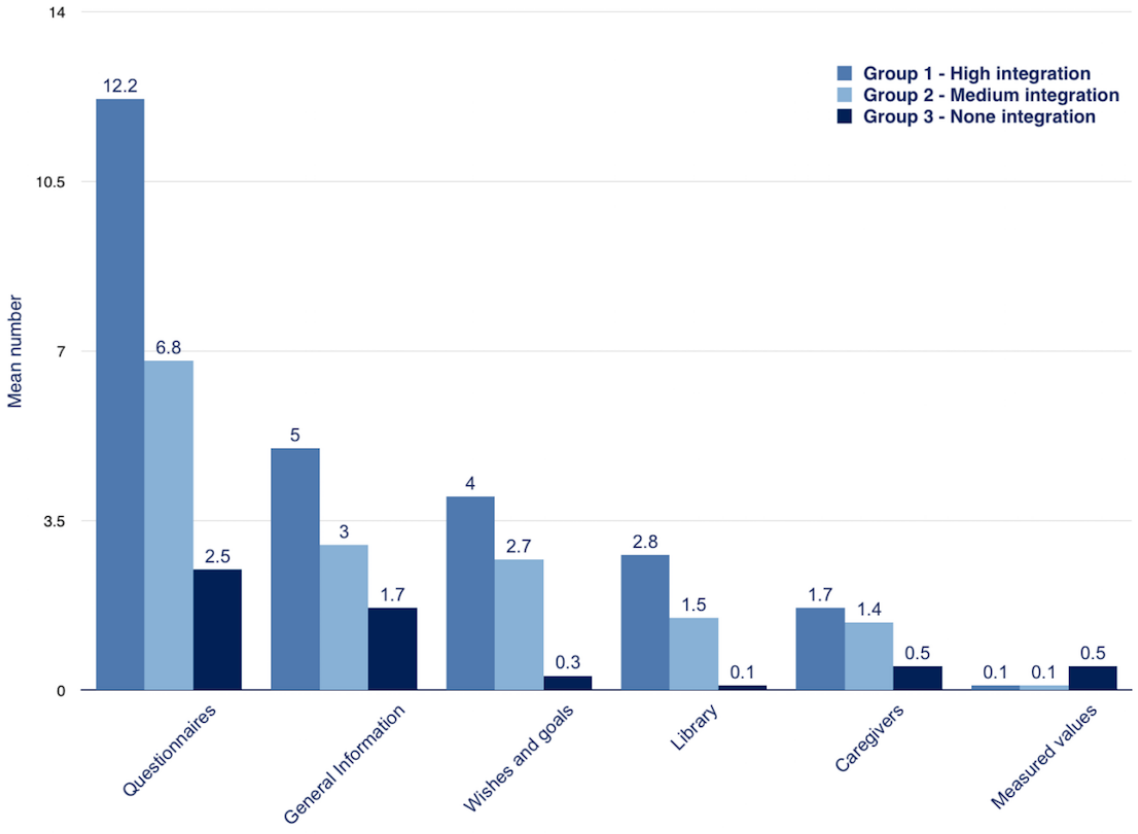
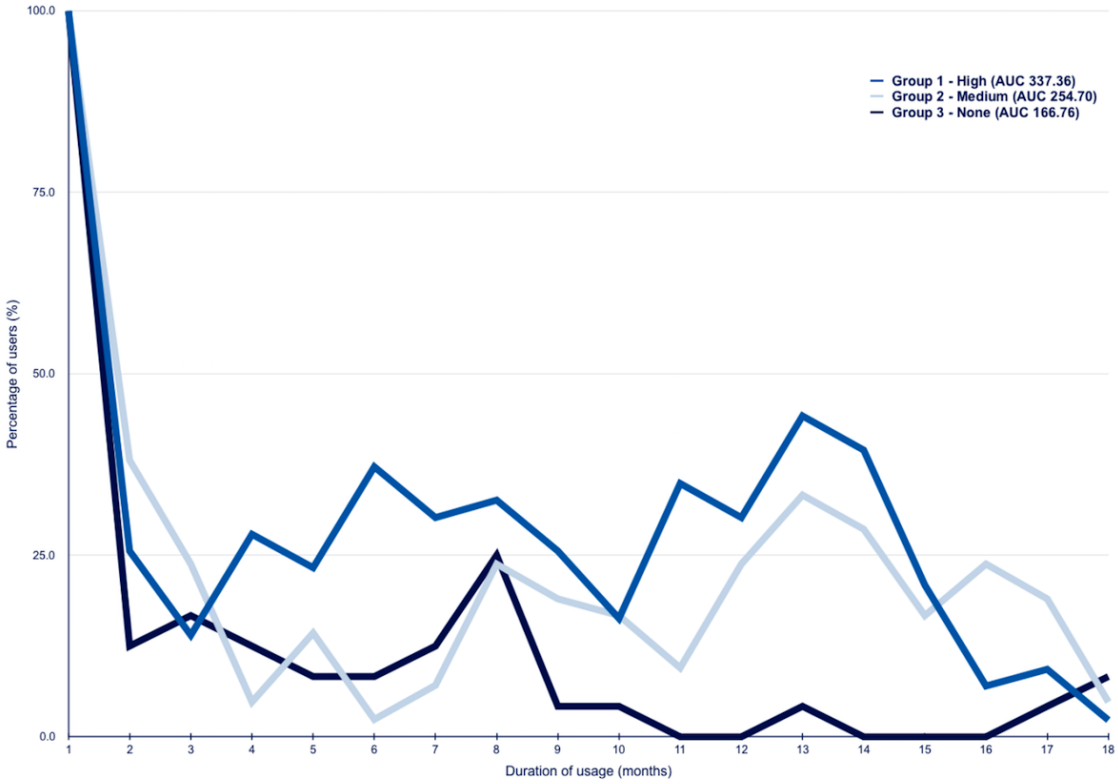


Table 4. Educational level and the General Self-Efficacy Scale (GSES) as predictors for Web platform usage.

Education/GSES	Unadjusted for age and gender		Adjusted for age and gender	
	B (95% CI)	*P*	B (95% CI)	*P*
Education		.20		.15
Low	–4.40 (–9.51 to 0.71)		–5.30 (–10.86 to 0.26)	
Medium	–1.27 (–5.90 to 3.39)		–1.85 (–6.61 to 2.90)	
High	0.00		0.00	
GSES	0.23 (–4.85 to 5.31)	.93	0.67 (–4.43 to 5.78)	.80

XSL•FO
RenderX

Figure 6. Attrition curve of group 1, 2 and 3.



Educational Level and the Generalized Self-Efficacy Scale

The association between the educational level and scores on the GSES and the mean number of sessions are presented in Table 4. Educational level was not associated with the number of sessions (*P*=.15). No association was found between the GSES and the mean number of sessions (*P*=.79).

Attrition

The log files revealed that a substantial proportion of the users did not continuously use the platforms before completion of the study. Figure 6 shows the patterns of use of the Web platforms in groups 1 to 3 during the intervention period, with the percentage of users on the y-axis, starting with 100% of the users, and the duration of usage on the x-axis. The area under the curve (AUC) until the 18th month for attrition in group 1 was 337.36, 254.70 in group 2, and 166.76 in group 3.

Discussion

Main Results

In this study, usage of the COPD self-management Web-based platform is higher when the platform is an integrated part of IDM with trained health care professionals who encourage patients to use the platform. Furthermore, usage of the e-Vita COPD platform is higher when patients receive more personal assistance in learning how to use the platform. Usage of the self-management Web-based platform e-Vita (high and medium level of integration in IDM) is higher than that of Zorgdraad (no integration in IDM).

Interpretation and Findings

Despite high expectations and numerous initiatives in the area of eHealth, implementation and use of eHealth apps are not yet common practice. Our primary aim was to analyze the factors that successfully promote the usage of two self-management Web platforms for COPD patients. We compared different organizational implementation methods. An implementation setup with greater personal support is expected to increase the use of an online program.

Our findings highlight the importance of integrating a Web-based platform into IDM; usage of the self-management Web platform is higher and more varied when the platform is an integrated part of IDM with appropriate personal coaching for patients. Patients in care groups with a high level of integration of the platform in IDM showed a higher number of sessions and a larger amount of visited services with more variation. Patients that received personal assistance also showed higher usage of the platform. Similar results were found in a study on COPD and asthma patients; the online app was used on a more regular basis with higher involvement of the health care provider and more assistance of the patients [25]. The e-Vita study on patients with diabetes mellitus showed minimal impact of implementing a personal health record including self-management support in primary diabetes care; recommendations were made to use additional strategies for patient motivation and engagement of professionals for a successful adoption of Web-based platforms [41,42].

In this study, we implemented extensive professional training of health care professionals on IDM and self-management

XSL•FO
RenderX

supported by platforms; we also offered personal assistance for the users to guide them through the platform as well as push factors (automated reminders or messages by health care professionals). Both strategies are essential elements to influence the use of platforms.

The self-efficacy of users (GSES) was not a predictor for use of the platforms. The construct of perceived self-efficacy reflects an optimistic self-belief [43]; a correlation can be understood based on the belief that one can change behavior, perform a novel or difficult task (eg, using a platform), or cope with adversity with a higher GSES [44,45]. In a healthy Dutch population, the mean GSES is reported to be approximately 3.1 [38,46]; in our study population, the median GSES was 3.3. Educational level was not a predictor of use of the e-Vita COPD platform. These predictors might be useful for future studies on and the development of platforms.

The EQ-5D values reflect the relative desirability of health states on a scale in which 1 refers to full health and 0 refers to death. In our study population, the median EQ-5D was 0.86 compared with 0.87 in a healthy Dutch population [47].

In the three groups, there was no bias regarding baseline age, gender, education level, CCQ, MRC scale, GSES, and EQ-5D.

Analysis of attrition provided insight into the decrease in usage (eg, after 1 month, 10%-45% of the participants were actively using the platform). The periodic steep rise in the percentage of users might be explained by the email reminders sent by the platform to fill in the questionnaires; all users received continuous reminders during the intervention period. In group 3, all users received urgent and repeated requests to fill in questionnaires at the end of the intervention period, which probably explains the steep rise in the percentage of users at the end of the study. The attrition curve depicts the “push factors” that are required to remind participants to use the platform. This “law of attrition” (the phenomenon of participants stopping usage) is a common finding in eHealth evaluations and one of the fundamental and methodological challenges in the evaluation of eHealth apps [29].

During this study, there were several lessons learned by the research team. First, it took a lot of effort to motivate health care professionals to work together with patients on self-management platforms; we experienced differences in communication skills among the health care professionals working with patients in a more modern, less hierarchic way. When patients started using the platform, it took great effort to stimulate the usage with several reminders, even though we established a high amount of attrition probably due to low usability of the platforms and logging problems.

Strengths and Limitations

This e-Vita COPD study has several strengths. To our knowledge, it is the first to combine different study designs thereby enabling simultaneous investigation of clinical effects, as well as the effects of different organizational implementation methods. Randomization was carried out for the level of assistance for patients. This study also adds evidence to the existing body of knowledge; this is important because local political and financial factors have a major impact on successful integration of eHealth in daily practice [48].

This study also has limitations. Although well-conducted randomized trials provide the most reliable evidence on the effectiveness of interventions, they are not feasible for our setting of an implementation design with organizational changes in a real-life health care system within three different care groups with different demands.

From a technical perspective, development of the Web-based platforms was difficult due to a lack of broad experience in the field. We used two different platforms with the same basic principles and functionalities. The platforms were new and the usability was not tested thoroughly before starting the study. The platform technique and the decisions made during the design phase were beyond the influence of our research group, but have affected our outcomes. From a human perspective, self-management skills imply behavioral changes. Behavioral changes require time, whereas the study period was restricted to 15 months. Furthermore, patients in a primary care setting have a low burden of disease (in this study, the mean CCQ score was 1.2) and motivation to use the platform might be negatively influenced by this fact. In respiratory medicine, there is a lack of research on patients with mild to moderate COPD despite that more than 80% of COPD patients suffer from this stage of disease and are often treated in primary care [49]. From an organizational perspective, other projects in primary care cooperatives can influence the speed and thoroughness of the implementation of our Web portal. Finally, this study also has typical limitations found in eHealth trials. The loss to follow-up is high, as in all eHealth studies. Because general practices, as well as patients, were free to volunteer, bias might have occurred in our research groups. Users were self-selected and were presumably motivated to use the Web-based platform as would be expected in a real-life setting. The patients that were invited by GPs/nurses to participate in the study might differ from other patient groups. Furthermore, GPs excluded 26% of the COPD patients from this study. Of the 702 eligible patients, 215 (30.6%) were willing to participate and provided informed consent, and 132 (61.4%) of the participants dropped out during follow-up. Even though nonparticipants did not differ in age and gender from participants, caution is required when generalizing these results to general practice as a whole. However, the practical applicability of our results for other primary care groups is positive (ie, the study provides practical insight into successful implementation of patient platforms). Nevertheless, primary care organizations should take into account the different aspects of good organization of blended care and good quality of implementation.

More studies are needed (preferably with larger sample groups and among the nonusers) to gain more insight into the preferences of various patient groups. The substantial workload generated by integrating a Web-based platform in IDM emphasizes the importance of piloting and assessing workforce implications for primary care groups during the planning and implementation phase. These results provide additional insight into the organizational aspects of the implementation of platforms, including the need to assist patients in the use of Web-based platforms integrated in IDM.

XSL•FO
RenderX

Conclusions

Use of a self-management Web-based platform is higher when the platform is an integrated part of IDM, with trained health care professionals encouraging patients to use the platform.

Use of a self-management platform is higher when participants receive adequate personal assistance about how to use the platform.

The self-management Web-based platform e-Vita had higher usage than Zorgdraad, which is probably due to the superior organizational conditions of integrated care and because e-Vita is, technically, better customized for use. An implementation setup with blended care through integration of the online platform in IDM, together with greater personal support of the users, will likely lead to increased use of the online program. Future research should provide additional insights into the preferences of different patient groups.

Acknowledgments

This research was supported by the Dutch foundation “Care Within Reach.” The authors thank Pim Assendelft for his valuable contribution to the design and Anneke Vass for her research assistance. The authors thank the Zorgdraad Foundation for their assistance on the Zorgdraad platform and the acquisition of data. The authors also thank Ron Wolterbeek for reviewing the statistical analysis. The authors wish to thank all coordinators, health care professionals, and participants within the care groups PreventZorg, Zorg op Noord, and Leidse Rijn Julius Gezondheidscentra.

Authors' Contributions

ET was the principal investigator and contributed to all aspects of the research. NV assisted on all aspects. MK assisted on the statistical analysis and was responsible for revising the manuscript several times. LH was responsible for revising the manuscript several times. IT was responsible for the acquisition of data and revising the manuscript several times. MN was responsible for revising the manuscript. NC was responsible for the concept, design, and for revising the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Homepage e-Vita.

[JPG File, 352KB - jmir_v19i5e185_app1.jpg]

Multimedia Appendix 2

Getting started with the e-Vita platform - https://www.youtube.com/watch?v=h3MDKgPCUJg.

[JPG File, 68KB - jmir_v19i5e185_app2.jpg]

Multimedia Appendix 3

Homepage Zorgdraad.

[JPG File, 735KB - jmir_v19i5e185_app3.jpg]

References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2095-2128. [doi: 10.1016/S0140-6736(12)61728-0] [Medline: 23245604]
2. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. Clin Chest Med 2014 Mar;35(1):7-16. [doi: 10.1016/j.ccm.2013.10.002] [Medline: 24507833]
3. World Health Organization. 2015 Mar. Chronic obstructive pulmonary disease (COPD): fact sheet URL: http://www.who.int/mediacentre/factsheets/fs315/en/ [accessed 2015-10-17] [WebCite Cache ID 6lKKYMcI5]
4. Global Initiative for Chronic Obstructive Pulmonary Disease. 2016. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease URL: http://goldcopd.org/ [accessed 2016-10-17] [WebCite Cache ID 6lKKjGL5q]
5. Bellamy D, Bouchard J, Henrichsen S, Johansson G, Langhammer A, Reid J, et al. International Primary Care Respiratory Group (IPCRG) Guidelines: management of chronic obstructive pulmonary disease (COPD). Prim Care Respir J 2006 Feb;15(1):48-57 [FREE Full text] [doi: 10.1016/j.pcrj.2005.11.003] [Medline: 16701758]

XSL•FO
RenderX

6. van den Bemt L, Schermer T, Smeele I, Bischoff E, Jacobs A, Grol R, et al. Monitoring of patients with COPD: a review of current guidelines' recommendations. Respir Med 2008 May;102(5):633-641 [FREE Full text] [doi: 10.1016/j.rmed.2007.12.014] [Medline: 18242067]
7. Kruis AL, Smidt N, Assendelft WJ, Gussekloo J, Boland MRS, van Rutten MM, et al. Integrated disease management interventions for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2013 Oct 10(10):CD009437. [doi: 10.1002/14651858.CD009437.pub2] [Medline: 24108523]
8. Kruis AL, Smidt N, Assendelft WJ, Gussekloo J, Boland MRS, van Rutten MM, et al. Cochrane corner: is integrated disease management for patients with COPD effective? Thorax 2014 Nov;69(11):1053-1055 [FREE Full text] [doi: 10.1136/thoraxjnl-2013-204974] [Medline: 24415716]
9. Effing T, Monninkhof EM, van der Valk PD, van der Palen J, van Herwaarden CL, Partidge MR, et al. Self-management education for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2007(4):CD002990. [doi: 10.1002/14651858.CD002990.pub2] [Medline: 17943778]
10. Bourbeau J, Collet J, Schwartzman K, Ducruet T, Nault D, Bradley C. Economic benefits of self-management education in COPD. Chest 2006 Dec;130(6):1704-1711. [doi: 10.1378/chest.130.6.1704] [Medline: 17166985]
11. Effing T, Kerstjens H, van der Valk P, Zielhuis G, van der Palen J. (Cost)-effectiveness of self-treatment of exacerbations on the severity of exacerbations in patients with COPD: the COPE II study. Thorax 2009 Nov;64(11):956-962. [doi: 10.1136/thx.2008.112243] [Medline: 19736179]
12. Maltais F, Bourbeau J, Shapiro S, Lacasse Y, Perrault H, Baltzan M, Chronic Obstructive Pulmonary Disease Axis of Respiratory Health Network, Fonds de recherche en santé du Québec. Effects of home-based pulmonary rehabilitation in patients with chronic obstructive pulmonary disease: a randomized trial. Ann Intern Med 2008 Dec 16;149(12):869-878. [Medline: 19075206]
13. Taylor SJC, Pinnock H, Epiphaniou E, Pearce G, Parke HL, Schwappach A. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS - Practical systematic Review of Self-Management Support for long-term conditions. Health Serv Deliv Res 2014;2(53):1. [Medline: 25642548]
14. Effing TW, Vercoulen JH, Bourbeau J, Trappenburg J, Lenferink A, Cafarella P, et al. Definition of a COPD self-management intervention: International Expert Group consensus. Eur Respir J 2016 Jul;48(1):46-54. [doi: 10.1183/13993003.00025-2016] [Medline: 27076595]
15. van Houtum L, Heijmans M, Rijken M, Groenewegen P. Perceived quality of chronic illness care is associated with self-management: Results of a nationwide study in the Netherlands. Health Policy 2016 Apr;120(4):431-439. [doi: 10.1016/j.healthpol.2015.11.006] [Medline: 27017047]
16. Vincent E, Sewell L, Wagg K, Deacon S, Williams J, Singh S. Measuring a change in self-efficacy following pulmonary rehabilitation: an evaluation of the PRAISE tool. Chest 2011 Dec;140(6):1534-1539. [doi: 10.1378/chest.10-2649] [Medline: 21737490]
17. McLean S, Protti D, Sheikh A. Telehealthcare for long term conditions. BMJ 2011;342:d120. [Medline: 21292710]
18. Finkelstein J, Friedman RH. Potential Role of Telecommunication Technologies in the Management of Chronic Health Conditions. Dis Manag Health Out 2000;8(2):57-63. [doi: 10.2165/00115677-200008020-00001]
19. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
20. Beenkens FH. Acceptance of e-Health Technology: A Patient Perspective. Alblasserdam: Haveka; 2011.
21. Bischoff EW, Hamd DH, Sedeno M, Benedetti A, Schermer TR, Bernard S, et al. Effects of written action plan adherence on COPD exacerbation recovery. Thorax 2011 Jan;66(1):26-31. [doi: 10.1136/thx.2009.127621] [Medline: 21037270]
22. Bucknall CE, Miller G, Lloyd SM, Cleland J, McCluskey S, Cotton M, et al. Glasgow supported self-management trial (GSuST) for patients with moderate to severe COPD: randomised controlled trial. BMJ 2012;344:e1060 [FREE Full text] [Medline: 22395923]
23. KPMG International Cooperative. 2012. Accelerating innovation: the power of the crowd. Global lessons in eHealth implementation URL: https://www.kpmg.com/Ca/en/IssuesAndInsights/ArticlesPublications/Documents/Accelerating-Innovation-the-Power-of-the-Crowd.pdf [accessed 2015-10-17] [WebCite Cache ID 6lKL9Pqyz]
24. Kasteleyn MJ, Bonten TN, Taube C, Chavannes NH. Coordination of care for patients with COPD: Clinical points of interest. Int J Care Coord 2015 Dec 08;18(4):67-71. [doi: 10.1177/2053434515620223]
25. In 't Veen JC, Mennema B, van Noort E. Online self-management in COPD or asthma: with or without the healthcare provider? Eur Respir J 2012;40(suppl 56):237.
26. Flynn D, Gregory P, Makki H, Gabbay M. Expectations and experiences of eHealth in primary care: a qualitative practice-based investigation. Int J Med Inform 2009 Sep;78(9):588-604. [doi: 10.1016/j.ijmedinf.2009.03.008] [Medline: 19482542]
27. Hofstede J, de Bie J, van Wijngaarden B, Heijmans M. Knowledge, use and attitude toward eHealth among patients with chronic lung diseases. Int J Med Inform 2014 Dec;83(12):967-974. [doi: 10.1016/j.ijmedinf.2014.08.011] [Medline: 25269992]

XSL•FO
RenderX

28. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
29. 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]
30. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013;15(8):e172 [FREE Full text] [doi: 10.2196/jmir.2258] [Medline: 23963284]
31. Talboom-Kamp EP, Verdijk NA, Blom CM, Harmans LM, Talboom IJ, Numans ME, et al. e-Vita: design of an innovative approach to COPD disease management in primary care through eHealth application. BMC Pulm Med 2016 Aug 16;16(1):121 [FREE Full text] [doi: 10.1186/s12890-016-0282-5] [Medline: 27530775]
32. 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]
33. van Bokhoven MA, Kok G, van der Weijden T. Designing a quality improvement intervention: a systematic approach. Qual Saf Health Care 2003 Jun;12(3):215-220 [FREE Full text] [Medline: 12792013]
34. Smeele IJ, van Weel C, van Schayck CP. Dutch College of General Practitioners Guideline for COPD Diagnosis. Huisarts Wet 2007;50(8):362-379.
35. VitalHealth News. 2012. Foundation Care Within Reach implements the interactive e-Vita care platform URL: http://www.vitalhealthsoftware.com/news/2012/10/29/foundation-care-within-reach-implements-the-interactive-e-vita-care-platform [accessed 2016-01-03] [WebCite Cache ID 6lKLIeisy]
36. van der Molen T, Willemse BW, Schokker S, ten Hacken NH, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. Health Qual Life Outcomes 2003 Apr 28;1:13 [FREE Full text] [Medline: 12773199]
37. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax 1999 Jul;54(7):581-586 [FREE Full text] [Medline: 10377201]
38. Scholz U, Gutiérrez-Doña B, Sud S, Schwarzer R. Is general self-efficacy a universal construct? Psychometric findings from 25 countries. Eur J Psychol Assess 2002;18:242-251.
39. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [Medline: 11491192]
40. International Standard Classification of Education: ISCED 2011. Montreal, QC: UNESCO Institute for Statistics URL: http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf [accessed 2017-04-05] [WebCite Cache ID 6pUidkdmJ]
41. van Vugt M, de Wit M, Sieverink F, Roelofsen Y, Hendriks SH, Bilo HJG, et al. Uptake and effects of the e-Vita personal health record with self-management support and coaching, for type 2 diabetes patients treated in primary care. J Diabetes Res 2016;2016:5027356 [FREE Full text] [doi: 10.1155/2016/5027356] [Medline: 26955640]
42. Sieverink F, Kelders SM, Braakman-Jansen LM, van Gemert-Pijnen JE. The added value of log file analyses of the use of a personal health record for patients with type 2 diabetes mellitus: preliminary results. J Diabetes Sci Technol 2014 Mar 4;8(2):247-255 [FREE Full text] [doi: 10.1177/1932296814525696] [Medline: 24876574]
43. Schwarzer R. Self-Efficacy: Thought Control Of Action. Oxford, UK: Taylor & Francis; 1992.
44. Luszczynska A, Mohamed NE, Schwarzer R. Self-efficacy and social support predict benefit finding 12 months after cancer surgery: The mediating role of coping strategies. Psychol Health Med 2005;10:365-375.
45. Scholz U, Sniehotta FF, Schwarzer R. Predicting physical exercise in cardiac rehabilitation: The role of phase-specific self-efficacy beliefs. J Sport Exerc Psychol 2005;27:135-151.
46. Luszczynska A, Gutiérrez - Doña B, Schwarzer R. General self - efficacy in various domains of human functioning: evidence from five countries. Int J Psychol 2005 Apr;40(2):80-89. [doi: 10.1080/00207590444000041]
47. Stolk E, Krabbe P, Busschbach J. Using the Internet to collect EQ-5D norm scores: a valid alternative. In: Proceedings of the 24th Scientific Plenary Meeting of the EuroQol Group. 2009 Presented at: 24th Scientific Plenary Meeting of the EuroQol Group; Sep 13-15, 2007; Kijkduin-The Hague, the Netherlands p. 153-165.
48. Nijland N. Grounding eHealth: Towards a Holistic Framework for Sustainable eHealth Technologies [thesis]. Enschede, the Netherlands: University of Twente; 2011.
49. Kruis AL, Boland MR, Schoonvelde CH, Assendelft WJ, Rutten-van Mölken MP, Gussekloo J, et al. RECODE: design and baseline results of a cluster randomized trial on cost-effectiveness of integrated COPD management in primary care. BMC Pulm Med 2013 Mar 23;13:17 [FREE Full text] [doi: 10.1186/1471-2466-13-17] [Medline: 23522095]

Abbreviations

CCQ: Clinical COPD Questionnaire
COPD: chronic obstructive pulmonary disease

XSL•FO
RenderX

EQ-5D: EuroQol 5D
GP: general practitioner
GSES: General Self-Efficacy Scale
IDM: integrated disease management
MRC scale: modified Medical Research Council dyspnea scale
QoL: quality of life

Edited by G Eysenbach; submitted 24.11.16; peer-reviewed by N Liu, A Farmer, M Stellefson, T Irizarry; comments to author 05.02.17; revised version received 28.02.17; accepted 24.04.17; published 31.05.17.

Please cite as:
Talboom-Kamp EPWA, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJSH, Numans ME, Chavannes NH
High Level of Integration in Integrated Disease Management Leads to Higher Usage in the e-Vita Study: Self-Management of Chronic Obstructive Pulmonary Disease With Web-Based Platforms in a Parallel Cohort Design
J Med Internet Res 2017;19(5):e185
URL: http://www.jmir.org/2017/5/e185/
doi:10.2196/jmir.7037
PMID:28566268

©Esther PWA Talboom-Kamp, Noortje A Verdijk, Marise J Kasteleyn, Lara M Harmans, Irvin JSH Talboom, Mattijs E Numans, Niels H Chavannes. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of Pollen Apps Forecasts: The Need for Quality Control in an eHealth Service

Katharina Bastl[1], Mag rer nat, Dr rer nat; Uwe Berger[1], MBA; Maximilian Kmenta[1,2], Mag rer nat

[1]Research group Aerobiology and pollen information, Department of Oto-Rhino-Laryngology, Medical University of Vienna, Wien, Austria

[2]Research group Paleobotany, Department of Paleontology, University of Vienna, Wien, Austria

Corresponding Author:
Katharina Bastl, Mag rer nat, Dr rer nat
Research group Aerobiology and pollen information
Department of Oto-Rhino-Laryngology
Medical University of Vienna
HNO-Klinik, 8J
Währinger Gürtel 18-20
Wien, 1090
Austria
Phone: 43 140400 ext 33380
Fax: 43 14040039040
Email: katharina.bastl@meduniwien.ac.at

Abstract

Background: Pollen forecasts are highly valuable for allergen avoidance and thus raising the quality of life of persons concerned by pollen allergies. They are considered as valuable free services for the public. Careful scientific evaluation of pollen forecasts in terms of accurateness and reliability has not been available till date.

Objective: The aim of this study was to analyze 9 mobile apps, which deliver pollen information and pollen forecasts, with a focus on their accurateness regarding the prediction of the pollen load in the grass pollen season 2016 to assess their usefulness for pollen allergy sufferers.

Methods: The following number of apps was evaluated for each location: 3 apps for Vienna (Austria), 4 apps for Berlin (Germany), and 1 app each for Basel (Switzerland) and London (United Kingdom). All mobile apps were freely available. Today's grass pollen forecast was compared throughout the defined grass pollen season at each respective location with measured grass pollen concentrations. Hit rates were calculated for the exact performance and for a tolerance in a range of ±2 and ±4 pollen per cubic meter.

Results: In general, for most apps, hit rates score around 50% (6 apps). It was found that 1 app showed better results, whereas 3 apps performed less well. Hit rates increased when calculated with tolerances for most apps. In contrast, the forecast for the "readiness to flower" for grasses was performed at a sufficiently accurate level, although only two apps provided such a forecast. The last of those forecasts coincided with the first moderate grass pollen load on the predicted day or 3 days after and performed even from about a month before well within the range of 3 days. Advertisement was present in 3 of the 9 analyzed apps, whereas an imprint mentioning institutions with experience in pollen forecasting was present in only three other apps.

Conclusions: The quality of pollen forecasts is in need of improvement, and quality control for pollen forecasts is recommended to avoid potential harm to pollen allergy sufferers due to inadequate forecasts. The inclusion of information on reliability of provided forecasts and a similar handling regarding probabilistic weather forecasts should be considered.

(J Med Internet Res 2017;19(5):e152) doi:10.2196/jmir.7426

KEYWORDS

pollen forecast; pollen information service; mobile app; pollen forecast quality; allergen avoidance

XSL•FO
RenderX

Introduction

Background

Pollen allergies are a global health concern [1]. They affect a considerable fraction of the population in various countries to a considerable extent [2-5]. Pollen forecasts have proven to be a highly valuable tool for allergen avoidance, handling, and treating pollen allergies [6,7]. Pollen information is consumed more often during the pollen season [8]—an indication regarding the need for such services. Today pollen forecasts are also distributed via mobile technology (mobile phones, tablets, any wireless devices) and mobile health (mHealth) has gained importance in terms of informing the public. However, eHealth is still in need for developing standards and the ethical framework [9] to avoid unintended consequences [10]. Sawand et al [11] provided a recent review on secure eHealth monitoring. In addition, a global survey on eHealth was performed by the World Health Organization [12] which reported an increase in eHealth or mHealth activities in countries with higher incomes.

Data quality is a concern [13] and can be judged by various criteria such as accuracy, precision, completeness, timeliness, relevance, legibility, accessibility, usefulness, currency or freshness, and confidentiality. The quality of data or information given in any eHealth service especially in terms of accuracy, precision, and completeness is not always examined; among them mobile apps for food composition [14], diabetes [15], complex chronic disease and disability [16], iron intake [17], lupus management [18], or an online portal for patients with dementia [19] are good examples. All these studies indicate a need for an improvement in quality and usability in eHealth or mHealth services.

Pollen Forecasts

A range of mobile apps relevant to pollen allergies which provide pollen forecasts are freely available, but users have not been informed about the accurateness of pollen forecasts till date. Simultaneously, the identity of the publishers of such apps with health-related content for pollen allergy sufferers has not been available. Also, studies on apps with a focus on pollen allergies and allergen prevention are still rare [20,21]. Five of the apps examined herein were included in a study by Berger et al [20] evaluated concerning their functionality and content. The criteria in the Berger et al study were different from this study and examined the app from a formal point of view, but did not consider the quality of the pollen forecast itself. Although parameters of good scientific practice for pollen forecasting were recently defined [22], it is not known how different apps perform concerning their pollen forecast quality. Questions regarding the extent to which pollen forecasts represent reality and how forecasts perform concerning their prediction of the readiness to flower remained unanswered.

Herein, we analyze the grass pollen forecasts of 9 mobile apps from four different countries during the grass pollen season of 2016. Moreover, the forecast dates for the readiness to flower for grasses in order to examine the accurateness and reliability of these forecasts was evaluated. Gaps and errors in the forecasts are also considered in order to complete the picture of the performance of specific apps.

Methods

Selection of Mobile Apps Included

The following free downloadable apps were included in this study: (1) “Pollen” [23], (2) “Biowetter,” (3) “Pollenwarner,” (4) “DWD Pollenflug-Index” (short: “DWD”), (5) “Allergiehelfer,” (6) “Pollenflug Vorhersage” (short: “Pollenflug”), (7) “Allergohelp Deutschland” (short: “Allergohelp”), (8) “Pollen News,” and (9) “Hayfever pollen forecast” (short: “Hayfever”). Apps 1-3 were used for Austria and their forecast for Vienna was evaluated. Apps 4-7 were used for Germany and their forecast for Berlin was evaluated. App 8 was used for Switzerland and its forecast for Basel was evaluated. App 9 was used for the United Kingdom and its forecast for London was evaluated. Every day a screenshot of each app forecast was made at the same time (1 PM) during the study period, so that always an up-to-date forecast (the day’s forecast) was used for the evaluation. Forecasts were not analyzed 2 or 3 days ahead in order to examine the base accurateness of the pollen forecasts. The forecasted load for grass pollen (Poaceae) was analyzed for the evaluation of the forecasts, since the grass pollen season is a long pollen season and a less fluctuating one, thus one of the more simple ones to forecast.

Datasets

The pollen data used was retrieved from the European Aeroallergen Network (EAN) database and originates from the national pollen monitoring networks (see “Acknowledgments”). Pollen data was attained by Hirst-type pollen traps [24]. Operation of those volumetric traps as well as the light microscopic evaluation of the resulting air samples follow the minimum requirements of the European aerobiology community [25] also required from the EAN database assuring high quality of the pollen data. The grass pollen season was defined on the EAN standard definition from 1% to 95% of the annual grass pollen index for each location based on local grass pollen data: Vienna (May 1, 2016 to August 5, 2016), Berlin (May 12, 2016 to July 23, 2016), Basel (April 30, 2016 to August 8, 2016), and London (May 28, 2016 to July 22, 2016).

Statistics

Descriptive statistics and R3.3.1 (R Foundation for Statistical Computing, Vienna, Austria) were used for the analysis. Daily grass pollen concentrations were compared within this defined grass pollen season with the forecasted loads. Used load classes and their respective pollen concentration ranges are shown in Table 1. The ranges used are adopted from the Austrian pollen information service and partly correspond to the ranges also used in Germany, Switzerland, and the United Kingdom. A tolerance between the load categories was introduced to allow flexibility around the boundaries (due to biogeographical differences or different pollen load definitions). A correct prediction in this setting means that the predicted load is the “exact” category or the category computed with the increased or decreased value. We define herein “hit rate” as the measure of the pollen forecast performance concerning accuracy of forecasted daily pollen concentrations, that is, the rate of correct forecasts within a defined time period. The calculations comprise

XSL•FO
RenderX

an exact hit rate (no tolerance), a hit rate of ±2 pollen, and a hit rate of ±4 pollen around the boundary of each category (Table 2). This is because a forecast should not be falsified automatically due to a small difference in overlap (eg, of one pollen grain). An adopted approach had to be developed for the "DWD" app since it also uses intermediate forecast loads in contrast to all other apps which should overlap with two neighboring categories. Numerical intervals were computed based on the mean between the load boundaries and a 20% interval (see Table 1). For the exact match, the ratio of correct predictions was computed (in the case of the "DWD" app, if the pollen concentration was in the overlapping interval, both categories was counted as correct).

In addition, the performance of the forecast "readiness to flower" for grasses was compared with the first-occurring moderate pollen load in the season. The first moderate grass pollen load was chosen as "threshold," as low grass pollen loads can occur in the air preceding the grass pollen season. However, only two pollen apps ("Pollen" and "Pollen News") provided such a service.

Results

Description of the Mobile Apps

The mobile apps under study herein are characterized concerning their main options, functional background, and noteworthy observations. All of them are available for free.

"Pollen" comprises pollen information (forecasts, countdown for the season, different pollen dispersal models, and forecast maps), services concerning pollen allergy (diary for allergic symptoms and burden, daily load based on symptom entries, personalized pollen information, search for allergologists), botanical information, push notification services, and an imprint (the Austrian pollen information service is the developer and operator). To the best of our knowledge, it is the only app described and published scientifically [23] among the apps studied herein.

"Biowetter" is an app focusing on weather and biorhythm, thus pollen information is only one part of the app. An imprint does not exist and advertisement appears regularly in the app. During the study period, it was the app with the most frequent crash reports and missing forecast information (Figure 1).

"Pollenwarner" includes pollen information, push notification services, a symptom diary, and general information (tips and tricks). An imprint is not available, but viewing the logos indicates that the companies Tempo and Otriven are connected with this app.

Table 1. The forecasted load classification is shown along with the respective pollen concentrations used in this study. The intermediate stages (no-low, low-moderate, moderate-high) are used only by one app ("DWD").

Forecast load	Daily pollen concentrations
No	0-0.99
No-low	0-2
Low	1-19.9
Low-moderate	15.6-23.4
Moderate	20-49.9
Moderate-high	39.6-59.4
High	50 and above

Table 2. The results of the hit rates are shown per app and per analyses (ie, exact hit rates and hit rates with tolerances of ±2 or ±4 pollen, respectively). Note the increasing hit rates in some apps (such as "Pollen," "Allergiehelfer," and "Pollenflug") versus the less improving apps (such as "Hayfever").

Mobile apps	Exact hit rates (%)	Hit rates with tolerance 2 (%)	Hit rates with tolerance 4 (%)
"Pollen"	62.9	77.3	80.4
"Biowetter"	31.8	39.4	42.2
"Pollenwarner"	34.0	39.2	42.3
"DWD"	41.1	46.6	52.1
"Allergiehelfer"	48.6	54.3	57.1
"Pollenflug"	50.7	56.2	58.9
"Allergohelp"	45.2	50.7	52.1
"Pollen News"	42.4	48.5	51.5
"Hayfever"	35.7	39.3	39.3

XSL•FO
RenderX

"DWD" provides pollen forecasts, a forecast map, and possesses an imprint, "Deutscher Wetterdienst," which is not only the official German weather service, but also a higher federal authority in the business area of the Federal Ministry of Transport and Digital Infrastructure that provides pollen forecasts based on the pollen measurements from the Foundation German Pollen Information (PID).

"Allergiehelfer" comprises pollen forecasts including push-service and general information. The pharmaceutical company GlaxoSmithKline GmbH & Co. KG is indicated in the imprint.

"Pollenflug" delivers pollen forecasts, a forecast map, push notification services, an allergy questionnaire, and general information. The pharmaceutical company Hexal AG is mentioned in the imprint and the app includes advertisement of allergy medication.

"Allergohelp" provides pollen forecasts, forecast maps, general information on allergy and therapy, therapy documentation, and a search for allergologists. The pharmaceutical company Allergopharma GmbH & Co. KG is mentioned in the imprint.

"Pollen-News" comprises pollen information (pollen forecasts, flowering start, maps) and general information (tips and tricks, information on aeroallergens). The aha! Allergiezentrum Schweiz and Bundesamt für Meteorologie und Klimatologie MeteoSchweiz are indicated as publisher in the imprint. The aha! Allergiezentrum Schweiz sees itself as patient organization, whereas the Bundesamt für Meteorologie und Klimatologie of MeteoSchweiz is the national weather and climate service in Switzerland.

"Hayfever" provides pollen forecasts and forecast maps. Advertisement is present and an imprint does not exist. However, the company A.Vogel (natural and herbal remedies) is present in the form of a logo.

Results of the Pollen Forecast Quality as Assessed by Hit Rates

The quality of the grass pollen forecast, measured by hit rates (Table 2) varies per app. Considering the exact hit rates, 5 apps attained a score of 40-50% ("DWD," "Allergiehelfer," "Pollenflug," "Allergohelp," and "Pollen News"), whereas 3 apps attained a lower score, below 40% ("Biowetter," "Pollenwarner," and "Hayfever") and only one app scored above 60% ("Pollen"). The hit rates improved when a tolerance range was calculated and most apps benefitted from an increasing tolerance. Only "Hayfever" showed a low improvement followed by stagnation from a tolerance from ±2 to ±4 pollen. A significant increase in exact hit rates to a tolerance of ±2 was observed in most apps, especially in "Pollen" which showed an increase of about 15%. "Pollen" is also the app with the highest hit rates in total (up to 80% with a tolerance of ±4) and the greatest improvement and thus the smallest error. The next best performances ranged from 50% to 59% ("Pollenflug," "Allergiehelfer," "DWD," "Allergohelp," and "Pollen News"). Even with the highest tolerance, there were apps with a score below 50% ("Biowetter," "Pollenwarner," and "Hayfever").

Boxplots of the forecasted pollen level for all apps analyzed revealed further insights (Figure 1). Apps with gaps or errors during the grass pollen season were "Biowetter," "Allergiehelfer," and "Pollen News." It is remarkable that "Biowetter" not only showed the highest frequency of missing forecasts, but also forecasted no or a low grass pollen load in contrast to all other apps. Low quality also became evident in a similar distribution of the mean of the forecasted load and the 25-75% quantiles (eg, "Biowetter," "Allergohelp," and "Hayfever").

Grasses were ready to flower based on the forecast countdown of "Pollen" in Vienna on May 9, 2016; and the first moderate grass pollen load was also occurred on the very same day. Grasses were ready to flower based on the forecast countdown of "Pollen News" in Basel on May 10, 2016; and the first moderate grass pollen load occurred 3 days earlier (ie, on May 7, 2016).

Discussion

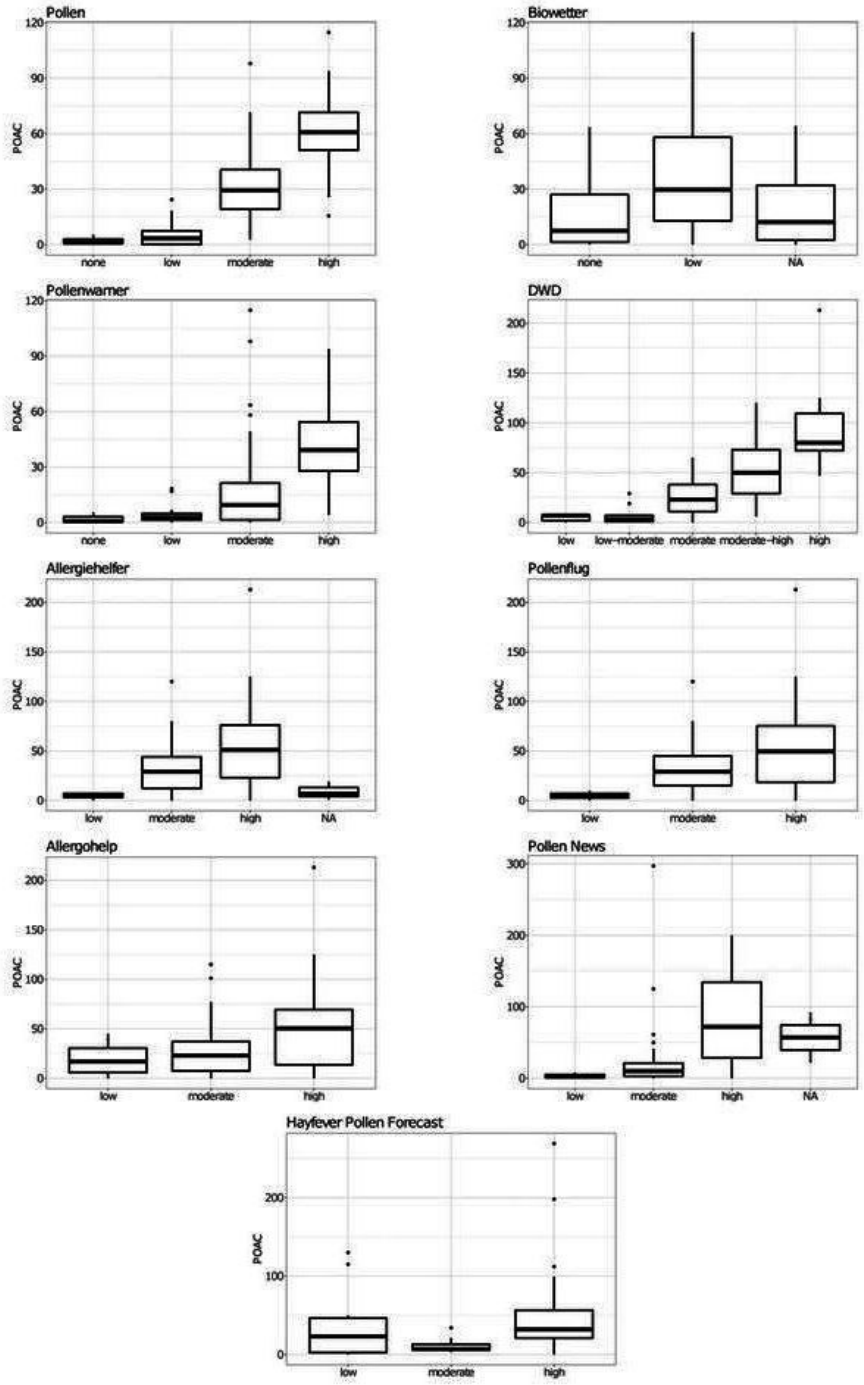
Principal Findings

Forecasting pollen concentrations or loads is not a trivial concern since many factors play a role in the development of the burden for pollen allergy sufferers. Several data sources and information are required including the biogeography of the region, continuous pollen monitoring and reliable pollen data, high quality weather forecasts, phenology, models (numerical simulations of pollen dispersal), medical-allergological expertise respective symptom data, and experience in the task of pollen forecasting [22]. The majority of apps in this study achieved only hit rates of 40-60% (8 apps) for pollen forecasts for the actual day when applying a tolerance range of ±4 pollen. The results of this study indicate that a certain forecast quality is already existent with room for improvement.

Although only two apps provided a forecast on the readiness to flower ("Pollen" and "Pollen News"), those forecasts had an excellent performance. The forecasted dates used here were the ones used when the information "ready to flower" occurred. Earlier forecast dates from mid-April (eg, April 15, 2016) announced for both apps May 6, 2016 as date for grasses ready to flower, which is still accurate enough considering it was a forecast from nearly a month before the start of the respective grass pollen season. The forecasted date coincided with the first moderate grass pollen load ("Pollen") or within a few days ("Pollen News") and may thus be recommended as a useful service for pollen allergy sufferers to prepare them for the grass pollen season.

XSL•FO
RenderX

Figure 1. Boxplots of the apps under study and their performance concerning forecasting grass pollen loads (none, low, moderate, high) and missing forecasts (NA).



Limitations

Certain facts and dependencies complicating the generation of an accurate pollen forecast have to be mentioned. Pollen forecasting activities depend on weather forecasts that are as accurate as possible. Although weather forecasting accuracy hit rates range from 85% to 95% for minimum and maximum temperatures on the same day, the forecasts of precipitation and sun hours are more complex and achieve lower rates [26]. Also, forecasts for environmental pollutants such as ozone are dependent on the accurateness of meteorological parameters. For example, the accuracy of ozone forecasts deviated from 6% to 25% in a 3-year evaluation due to meteorological forecast errors [27]. In addition, a consideration should be if pollen forecasts should be treated similarly to probabilistic weather forecasts. Hereby, we want to formulate the question if the scientific community and pollen information services should treat a forecast of a "low pollen load" as "high probability of a low pollen load" with all consequences ranging from defining probabilities and promoting and communicating it in an understandable way to the public which is already difficult for probabilistic weather forecasts [28]. Also, pollination varies with plant abundance and microclimate resulting in different pollen concentrations in the same locality or city [29-31]. This

XSL•FO
RenderX

fact was considered in this study regarding the analyses of hit rates with tolerances. We propose that analyses with tolerances, such as used herein, are advantageous because they compensate for minor discrepancies that might not be attributed to as errors. Missing pollen forecasts are a problem for some apps (4 apps) and a problem of apps published by companies. There is a need for improvement in reliability of the service since no information at all is as useless as inaccurate pollen forecasts for allergy sufferers concerned. It is obvious that a quality control of forecasts would be a benefit for pollen allergy sufferers to inform them about the reliability of the pollen forecasts of specific suppliers and to improve the quality of pollen forecasts.

Recommendations and Possible Gateways for the Future

Mobile apps dealing with allergen avoidance with the aim of supporting pollen allergy sufferers should fulfill certain criteria and functionalities, among them easily understandable pollen forecasts, a minimum of forecasted aeroallergens, botanical information, symptom diaries, allergy risk questionnaires, and an imprint with the publisher of the app stating the responsible institution, at best without conflict of interests [20]. Herein, it should be noted that advertisement is common in pollen information apps, especially if the publisher is a company and not directly involved in pollen forecasting. Apps with direct advertisement are “Biowetter,” “Pollenflug,” and “Hayfever.” Who the publisher of an app actually is represents crucial information. An institution involved in forecasting is in the background of the apps “Pollen,” “DWD,” and “Pollen News” in terms of occurrence in the imprint, but not in the six other apps. The need for an improvement and quality control of pollen forecasts is underlined by the outcome of this study.

A combination of data sources and methods will lead to an improvement of pollen forecasts and is already used in the mobile app with the best performance found in this study. Phenological routines assess the local progression of the pollen season. Pollen dispersal models and readiness to flower models support the person preparing the pollen forecast. Symptom data reveal the impact of the daily pollen concentrations on pollen allergy sufferers and allows for tailoring the forecast to the needs of persons concerned. These are possible gateways in the need for a tighter connection to a future pollen forecast. Specific recommendations concerning the improvement of pollen forecasts comprise (1) ongoing evaluation of pollen forecast quality at best also during the pollen season to raise quality as fast as possible, (2) comprehension of the probabilistic nature of pollen forecasts and implementation of this aspect in services visible for users (such as percentual information on the forecast accuracy), and (3) implementation of all information sources necessary such as symptom data and phenological routines besides pollen data to improve a pollen forecast [22].

Conclusions

Pollen forecasts are essential for pollen allergy sufferers in terms of allergen avoidance and thus the accuracy of such forecasts is a key factor for improving the quality of life. Most apps deliver forecasts with a hit rate of about 50%, which is a score that is too low for this purpose. Quality control of pollen forecasts should be introduced since wrong forecasts can be seen as potential physical injury and may harm persons concerned significantly. Pollen information and pollen forecasts should never be given out by pharmaceutical companies or be accompanied by advertisement [22] to ensure unbiased pollen forecasting. Finally, we have determined that proof of knowledge and necessary datasets for institutions developing or promoting apps with pollen forecasts is necessary and vitally important to ensure the safety of users and the quality of this eHealth service.

Acknowledgments

We thank the following institutions and persons for supplying pollen data: MeteoSwiss (for Basel), Foundation German Pollen Information (PID), Prof Karl-Christian Bergmann, Matthias Werchan (Berlin), and Met Office in Exeter (London). Pollen data from Vienna originates from the Austrian pollen information service and was supplied by one of the authors (MK). We are grateful to Alexander Kowarik for support in the statistical analyses. Gina Semprebon (Bay Path University) proofread the manuscript. It is our wish to express also our gratitude to two anonymous reviewers, who contributed with their insights to the manuscript.

Authors' Contributions

The study was designed by all authors (KB, UB, and MK). Data analyses were performed by KB and MK. Technical and scientific supervision was assured by UB. All authors were involved in data interpretation, drafting the manuscript, editing, and final approval.

Conflicts of Interest

KB, MK, and UB report to have taken part in the development of the app “Pollen” that is freely available and without advertisement and thus is of no financial interest.

References

1. Worldallergy. WAO White Book on Allergy: Update 2013 URL: http://www.worldallergy.org/UserFiles/file/WhiteBook2-2013-v8.pdf [accessed 2016-12-06] [WebCite Cache ID 6mY6o6Nan]
2. Asher MI, Montefort S, Björkstén B, Lai CK, Strachan DP, Weiland SK, ISAAC Phase Three Study Group. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC phases

XSL•FO
RenderX

one and three repeat multicountry cross-sectional surveys. Lancet 2006 Aug 26;368(9537):733-743. [doi: 10.1016/S0140-6736(06)69283-0] [Medline: 16935684]
3. D'Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H, et al. Allergenic pollen and pollen allergy in Europe. Allergy 2007 Sep;62(9):976-990 [FREE Full text] [doi: 10.1111/j.1398-9995.2007.01393.x] [Medline: 17521313]
4. Bergmann K, Heinrich J, Niemann H. Current status of allergy prevalence in Germany: position paper of the Environmental Medicine Commission of the Robert Koch Institute. Allergo J Int 2016;25:6-10 [FREE Full text] [doi: 10.1007/s40629-016-0092-6] [Medline: 27069844]
5. Bousquet J, Anto J, Auffray C, Akdis M, Cambon-Thomsen A, Keil T, et al. MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. Allergy 2011 May;66(5):596-604 [FREE Full text] [doi: 10.1111/j.1398-9995.2010.02534.x] [Medline: 21261657]
6. Kiotseridis H, Cilio CM, Bjermer L, Tunsäter A, Jacobsson H, Dahl A. Grass pollen allergy in children and adolescents-symptoms, health related quality of life and the value of pollen prognosis. Clin Transl Allergy 2013;3:19 [FREE Full text] [doi: 10.1186/2045-7022-3-19] [Medline: 23799882]
7. Schmid-Grendelmeier P. [Pollen allergy and immunotherapy]. Ther Umsch 2012 Apr;69(4):239-248. [doi: 10.1024/0040-5930/a000280] [Medline: 22477663]
8. Kmenta M, Zetter R, Berger U, Bastl K. Pollen information consumption as an indicator of pollen allergy burden. Wien Klin Wochenschr 2016 Jan;128(1-2):59-67. [doi: 10.1007/s00508-015-0855-y] [Medline: 26373744]
9. Svensson P. eHealth applications in health care management. eHealth Int 2002;1(1):5. [doi: 10.1186/1476-3591-1-5]
10. Charani E, Castro-Sánchez E, Moore LS, Holmes A. Do smartphone applications in healthcare require a governance and legal framework? It depends on the application!. BMC Med 2014;12:29 [FREE Full text] [doi: 10.1186/1741-7015-12-29] [Medline: 24524344]
11. Sawand A, Djahel S, Zhang Z, Nait-Abdesselam F. Toward energy-efficient and trustworthy eHealth monitoring system. China Commun 2015 Jan;12(1):46-65. [doi: 10.1109/CC.2015.7084383]
12. World Helath Organization. Apps.WHO. Management of patient information: trends and challenges in Member States: based on the findings of the second global survey on eHealth URL: http://apps.who.int/iris/bitstream/10665/76794/1/9789241504645_eng.pdf [WebCite Cache ID 6mQnK0fCe]
13. Gutierrez RC, Riveill M. nyx.unice. e-Health monitoring applications: What about Data Quality? URL: https://nyx.unice.fr/publis/gutierrezrodriguez-riveill:2011.pdf [accessed 2016-12-01] [WebCite Cache ID 6mQo6UlsH]
14. Koroušić SB. Web-based eHealth applications with reference to food composition data. Eur J Clin Nutr 2010 Nov;64(Suppl 3):S121-S127. [doi: 10.1038/ejcn.2010.222] [Medline: 21045840]
15. 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]
16. Steele GC, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
17. Mann D, Riddell L, Lim K, Byrne LK, Nowson C, Rigo M, et al. Mobile phone app aimed at improving iron intake and bioavailability in premenopausal women: a qualitative evaluation. JMIR Mhealth Uhealth 2015 Sep 28;3(3):e92 [FREE Full text] [doi: 10.2196/mhealth.4300] [Medline: 26416479]
18. Neville C, Da Costa D, Rochon M, Peschken CA, Pineau CA, Bernatsky S, et al. Development of the Lupus interactive navigator as an empowering web-based eHealth tool to facilitate Lupus management: users perspectives on usability and acceptability. JMIR Res Protoc 2016 May 30;5(2):e44 [FREE Full text] [doi: 10.2196/resprot.4219] [Medline: 27240666]
19. Hattink B, Droes R, Sikkes S, Oostra E, Lemstra AW. Evaluation of the Digital Alzheimer Center: testing usability and usefulness of an online portal for patients with dementia and their carers. JMIR Res Protoc 2016 Jul 21;5(3):e144 [FREE Full text] [doi: 10.2196/resprot.5040] [Medline: 27444209]
20. Berger U, Bastl K, Kmenta M, Berger M, Bergmann K. Polleninformations-Apps für Pollenallergiker? die Voraussetzungen für eine nützliche App und Vorteile für Betroffene. Atemwegs- und Lungenkrankheiten 2017;43(2):63-67 [FREE Full text] [doi: 10.5414/ATX02171]
21. Pla M, Lemus-Zúniga L, Montanana J, Pons J, Garza A. A review of mobile apps for improving quality of life of asthmaticpeople with allergies. In: Innovation in Medicine and Healthcare. Switzerland: Springer International Publishing; 2015:51-64.
22. Bastl K, Berger M, Bergmann K, Kmenta M, Berger U. The medical and scientific responsibility of pollen information services. Wien Klin Wochenschr 2017 Jan;129(1-2):70-74. [doi: 10.1007/s00508-016-1097-3] [Medline: 27761736]
23. Kmenta M, Bastl K, Jäger S, Berger U. Development of personal pollen information-the next generation of pollen information and a step forward for hay fever sufferers. Int J Biometeorol 2014 Oct;58(8):1721-1726. [doi: 10.1007/s00484-013-0776-2] [Medline: 24357491]
24. Hirst JM. An automatic volumetric spore trap. Ann Appl Biol 1952 Jun;39(2):257-265. [doi: 10.1111/j.1744-7348.1952.tb00904.x]

XSL•FO
RenderX

25. Galán C, Smith M, Thibaudon M, Frenguelli G, Oteros J, Gehrig R, et al. Pollen monitoring: minimum requirements and reproducibility of analysis. Aerobiologia 2014 Mar 20;30(4):385-395. [doi: 10.1007/s10453-014-9335-5]
26. MetOffice. How accurate are our public forecasts? URL: http://www.metoffice.gov.uk/about-us/who/accuracy/forecasts [accessed 2016-12-01] [WebCite Cache ID 6mQmsjx3K]
27. Honoré C, Rouïl L, Vautard R, Beekmann M, Bessagnet B, Dufour A, et al. Predictability of European air quality: assessment of 3 years of operational forecasts and analyses by the PREV'AIR system. J Geophys Res 2008 Feb 22;113(D4):-. [doi: 10.1029/2007JD008761]
28. Gigerenzer G, Hertwig R, van den Broek E, Fasolo B, Katsikopoulos KV. A 30% chance of rain tomorrow": how does the public understand probabilistic weather forecasts? Risk Anal 2005 Jun;25(3):623-629. [doi: 10.1111/j.1539-6924.2005.00608.x] [Medline: 16022695]
29. Bergmann K, Simoleit A, Wagener S, Mücke H, Werchan M, Zuberbier T. The distribution of pollen and particulate matter in an urban agglomeration using the city of Berlin as an example. Allergo J 2013 Oct 29;22(7):475-479. [doi: 10.1007/s15007-013-0430-7]
30. Cariñanos P, Sánchez-Mesa JA, Prieto-Baena JC, Lopez A, Guerra F, Moreno C, et al. Pollen allergy related to the area of residence in the city of Córdoba, south-west Spain. J Environ Monit 2002 Oct;4(5):734-738. [Medline: 12400923]
31. Emberlin J, Norris-Hill J. Spatial variation of pollen deposition in north London. Grana 1991 Jan;30(1):190-195. [doi: 10.1080/00173139109427798]

Abbreviations

EAN: European Aeroallergen Network

Edited by G Eysenbach; submitted 30.01.17; peer-reviewed by C Geller-Bernstein, M Thibaudon; comments to author 22.02.17; revised version received 01.03.17; accepted 01.03.17; published 08.05.17.

Please cite as:
Bastl K, Berger U, Kmenta M
Evaluation of Pollen Apps Forecasts: The Need for Quality Control in an eHealth Service
J Med Internet Res 2017;19(5):e152
URL: http://www.jmir.org/2017/5/e152/
doi:10.2196/jmir.7426
PMID:28483740

©Katharina Bastl, Uwe Berger, Maximilian Kmenta. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review

Rebecca Grist[1], PhD; Joanna Porter[2], BSc (Hons); Paul Stallard[1,2], PhD

[1]Child and Adolescent Mental Health Group, Department for Health, University of Bath, Bath, United Kingdom

[2]Child and Family Mental Health, Temple House, Oxford Health National Health Service Foundation Trust, Keynsham, United Kingdom

Corresponding Author:
Rebecca Grist, PhD
Child and Adolescent Mental Health Group
Department for Health
University of Bath
Wessex House
Claverton Down
Bath, BA2 7AY
United Kingdom
Phone: 44 2253850 ext 86
Fax: 44 1225 383282
Email: r.grist@bath.ac.uk

Abstract

Background: There are an increasing number of mobile apps available for adolescents with mental health problems and an increasing interest in assimilating mobile health (mHealth) into mental health services. Despite the growing number of apps available, the evidence base for their efficacy is unclear.

Objective: This review aimed to systematically appraise the available research evidence on the efficacy and acceptability of mobile apps for mental health in children and adolescents younger than 18 years.

Methods: The following were systematically searched for relevant publications between January 2008 and July 2016: APA PsychNet, ACM Digital Library, Cochrane Library, Community Care Inform-Children, EMBASE, Google Scholar, PubMed, Scopus, Social Policy and Practice, Web of Science, Journal of Medical Internet Research, Cyberpsychology, Behavior and Social Networking, and OpenGrey. Abstracts were included if they described mental health apps (targeting depression, bipolar disorder, anxiety disorders, self-harm, suicide prevention, conduct disorder, eating disorders and body image issues, schizophrenia, psychosis, and insomnia) for mobile devices and for use by adolescents younger than 18 years.

Results: A total of 24 publications met the inclusion criteria. These described 15 apps, two of which were available to download. Two small randomized trials and one case study failed to demonstrate a significant effect of three apps on intended mental health outcomes. Articles that analyzed the content of six apps for children and adolescents that were available to download established that none had undergone any research evaluation. Feasibility outcomes suggest acceptability of apps was good and app usage was moderate.

Conclusions: Overall, there is currently insufficient research evidence to support the effectiveness of apps for children, preadolescents, and adolescents with mental health problems. Given the number and pace at which mHealth apps are being released on app stores, methodologically robust research studies evaluating their safety, efficacy, and effectiveness is promptly needed.

(J Med Internet Res 2017;19(5):e176) doi:10.2196/jmir.7332

KEYWORDS

mobile apps; smartphone apps; mHealth; mental health; self-help; child; adolescent; preadolescent; review

XSL•FO
RenderX

Introduction

Mental health problems are common in children and young people. Prevalence data suggests that up to 20% of children and young people up to 18 years of age have a diagnosable mental health problem [1,2]. Mental health problems in children persist with 50% of mental illness in adulthood beginning before the age 15 years and 75% before the age of 18 years [3]. Mental health problems cause significant distress and negatively impact on social relationships, school and occupational attainment, and physical health [4]. They also increase the risk of developing other mental health problems in adulthood [5]. Although evidence-based interventions are available for many child mental health problems, treatment services are limited and many children and adolescents either cannot or do not access appropriate help [2,4].

Digital technology provides a way of increasing access to evidence-based interventions [6]. Computerized cognitive behavioral therapy (CBT), for example, offers a promising and acceptable way of delivering interventions for anxiety and depression for children and young people [6,7]. However, technology is constantly evolving and mobile technologies in particular are being adopted at an increasing rate; by 2020, it is estimated there will be 6.1 billion mobile phone users globally [8]. The majority of children and adolescents in 2017 have use of a mobile phone (72% of children aged 0-11 years and 96% of those aged 12-17 years) [9]. Mobile tablet use is also increasing with seven in 10 (71%) children aged 5 to 15 years having access to a tablet at home [10]. Therefore, mobile health (mHealth) offers a particularly powerful and ubiquitous platform for delivering mental health interventions to adolescents. mHealth uses the functions of a mobile device, but most commonly relies on the download of mobile apps to help support health care delivery [11]. More than 15,000 mobile apps for health care were identified in a recent survey with at least 29% designed for mental health [12]. These apps vary in function and may focus on symptom assessment, psychoeducation, promoting engagement with therapy homework (eg, a thought diary or activity schedule), practicing skills learned in therapy, or monitoring symptoms or mood [11].

Advantages of mHealth include constant availability, greater access, equity of mental health resources, immediate support, anonymity, tailored content, lower cost, and increasing service capacity and efficiency [13]. Apps may overcome geographical barriers to treatment and engage traditionally hard-to-reach groups. It has been suggested that technology-based approaches may be particularly suited for children and young people who may be more accepting of technology [14]. Apps may reduce barriers to face-to-face help-seeking, such as the stigma or discomfort about discussing one's own mental health [15]. Therefore, mental health apps may engage young people who typically would not seek help through traditional routes. Global and national organizations, such as the World Health Organization (WHO), the US Department of Health and Human Services, and the National Health Service (NHS), are generating initiatives for the integration of mHealth in services, including child and adolescent mental health [16,17].

Despite the large number of apps available, the evidence base is scarce, particularly for adolescents. A 2013 review of mobile mental health apps for all ages identified eight papers describing only five apps [18]. Four of the five apps demonstrated significant reductions in depression, stress, and substance use, although a number of issues with the quality of these studies suggest these conclusions needed to be interpreted cautiously. The review also highlighted how research has lagged behind app development. A review of mHealth apps for the most prevalent conditions identified by the WHO identified more than 1536 apps for depression, but only 32 associated published articles [19]. Content analysis of commercially available apps for depression [20] and bipolar disorder [21] demonstrate a concerning trend that downloadable apps may not necessarily reflect evidence-based treatment guidelines. The majority rarely cite source information and often lack privacy policies. This was also evidenced in the now offline NHS App Library, in which only four of the 27 apps included any evidence of patient-reported outcomes to corroborate their effectiveness [22]. As such, the majority of mental health apps available for download are not supported by evidence-based research and may not follow evidence-based treatment guidelines.

Few apps have been specifically developed for children and adolescents, and the benefit of mental health mobile apps for this population is unclear. Two systematic reviews exploring the evidence for digital health interventions (including computerized CBT, mobile phone apps, and wearable technologies) for children and young people with mental health problems in 2014 and 2016 [6,23,24] identified randomized controlled trials (RCTs) for only two apps (Mobiletype and FindMe). Results showed no significant benefits of these apps on depression or autism spectrum disorder symptoms. A scoping review of mHealth interventions for children and young people yielded similar results [25]. Only one app (Mayo Clinic Anxiety Coach) included outcomes using a standardized rating scale, whereas the other two apps identified (SmartCAT and Mobile Mood Diary) had feasibility outcomes, but no efficacy outcomes reported [25].

Although important additions to the literature, the systematic reviews only included RCTs and so did not include feasibility studies providing information on acceptability [6,23,24]. The scoping review was limited to three databases and focused exclusively on studies in which participants had a diagnosed mental health problem [25], therefore excluding any preventive or general mental well-being apps that may exist. This review aims to provide a contemporary appraisal of the available research evidence for the efficacy and acceptability of mobile apps to support the management of mental health in adolescents. A secondary aim was to collate the feedback from mental health professionals and adolescents involved in these studies. This review will focus on mobile phone apps only (as opposed to broad mHealth and eHealth interventions) and will include a wide remit of publication types. Given the increasing number of commercially available apps and the policy drivers toward integrating mHealth into mental health services [16,17], such a review is timely.

XSL•FO
RenderX

Methods

Study Identification and Selection

Fifteen electronic databases were searched for relevant publications between January 2008 and July 2016, including APA PsychNet, ACM Digital Library, Cochrane Library, Community Care Inform-Children, EMBASE, Google Scholar, PubMed, Scopus, Social Policy and Practice, and Web of Science. Publication databases of key journals were also searched. These included *Journal of Medical Internet Research*, *Cyberpsychology, Behavior and Social Networking*, and *Internet Interventions*. A grey literature search of OpenGrey, Index to Thesis, and ACM Digital Library was also conducted. Words pertaining to mobile apps and devices, mental health problems, and the age of the study population were used in a main search string (see Multimedia Appendix 1 for full search strings by database). Database-specific filters such as human population, English language, and age groups were applied where available. Authors of identified trial protocols were also contacted to determine the current status of these trials and whether any further data were available.

We included abstracts describing mental health apps for mobile devices (mobile phone or tablet) for use by children and adolescents younger than 18 years. Studies with participants older than 18 years were included if the sample included children younger than 18 years. Mental health problems included depression, bipolar disorder, anxiety disorders, self-harm, suicide prevention, conduct disorder, eating disorders and body image issues, schizophrenia, psychosis, and insomnia. To ensure we were capturing current and emerging evidence, we included conference proceedings, theses, case studies, RCTs, uncontrolled feasibility studies, qualitative studies, articles analyzing apps for adolescents available in app stores, and articles detailing app design and development.

We excluded abstracts if (1) the target population was exclusively adult (ie, older than 18 years); (2) the primary purpose of the app was ecological momentary assessment for research purposes as opposed to an intervention; (3) the app was designed for neurodevelopmental disorders (autism spectrum disorders, Asperger syndrome, and attention-deficit/hyperactivity disorder), for substance use, health behaviors, or medical problems; (4) the study described an Internet-based intervention accessed via a mobile device or an intervention delivered via mobile device functions (text messaging, multimedia messaging, calls, videoconferencing, sending content to Internet interventions); and (5) the paper was a trial protocol, trial registration, systematic or scoping review, or did not provide any extractable outcome or feasibility data.

Results

Study Inclusion

Of the 5562 abstracts initially identified, 5438 were excluded on the basis of title, abstract screening, and duplicate removal. The remaining 124 full-text articles were assessed for eligibility with a further 100 being excluded. A total of 24 full-text articles met the inclusion criteria. Figure 1 is a flow diagram detailing the review process and results at each stage.

Study Characteristics

The 24 publications included in this review consisted of 12 feasibility studies [26-37], five design and development papers [38-42], and two analyses articles of existing apps in app stores [21,43]. The remaining five reported mental health outcome data [44-48]; of these papers, three reported outcomes from the same RCT [44-46]. Only two studies randomized individuals to trial conditions (Mobiletype RCT [44-46]; Pretty [47]). Publication dates ranged from 2008 to 2016 with a notable increase in publications since 2014. Table 1 reports selected study characteristics.

Mobile App Characteristics

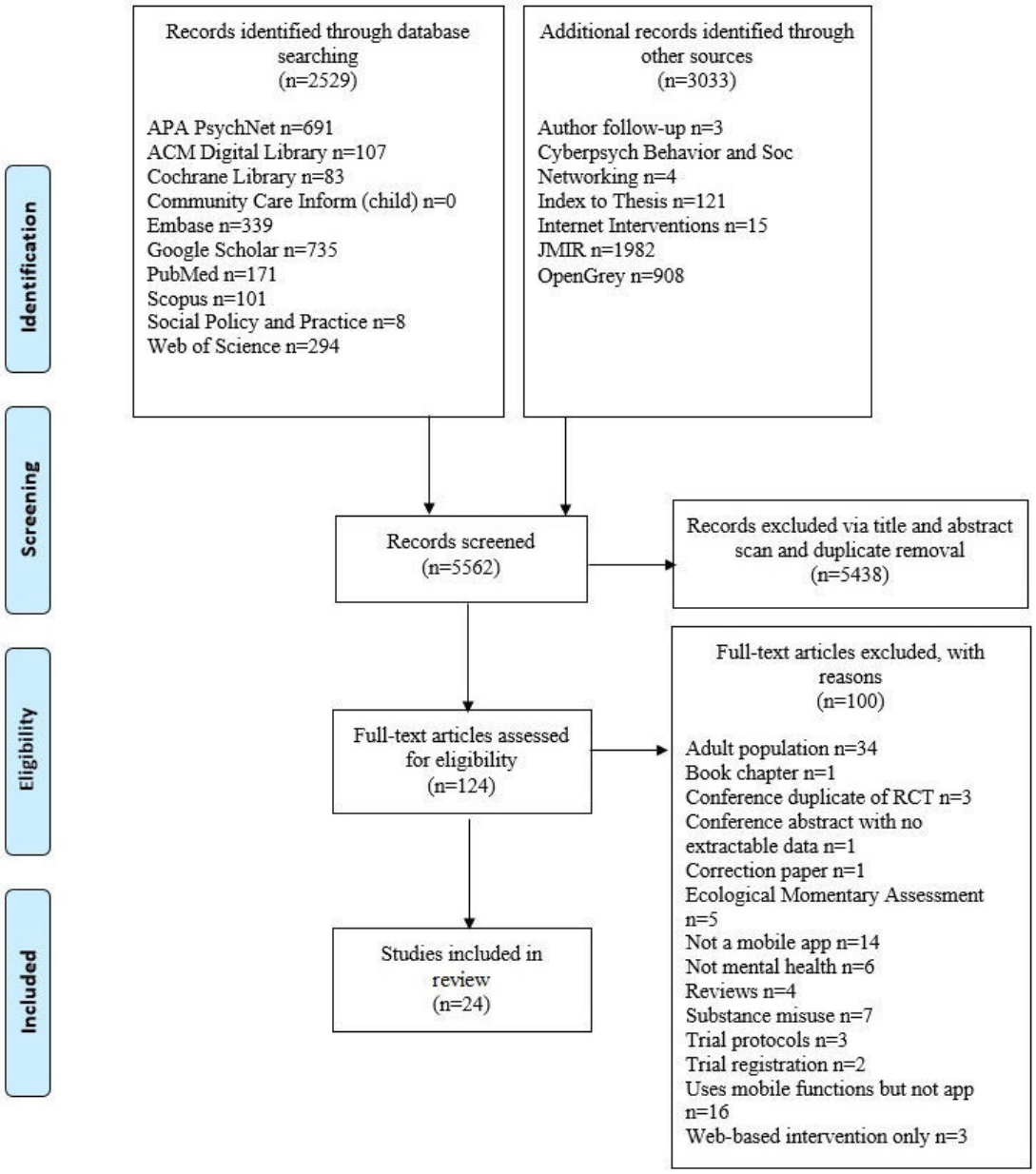
Table 2 summarizes the 15 apps identified in this review: CopeSmart [26,39], Crisis Care [30], Daybuilder [27], Mayo Clinic Anxiety Coach [42,48], Mobiletype [34,35,44-46], Mobile Mood Diary [28,29], Pretty [47], REACH app [32], Recovery Record [40], Safety Plan app [38], SmartCAT [33], The ACT app [41], and TickiT [37] (two apps had no name [31,36]).

Operating platforms included Android and iOS (n=3 [26,39,40,47]), Android only (n=4 [27,32,33,41]), iOS only (n=2 [37,42,48]), and multiple platforms (n=2 [28,29,34,35, 44-46]), with four being under development or not reporting the operating platform [30,31,36,38]. Note that CopeSmart, Mayo Clinic Anxiety Coach, Mobiletype, and Mobile Mood Diary were associated with multiple studies (Table 2). The primary focus of the apps were prevention and early intervention (n=4 [26,31,32,39,47]), assessment and screening (n=2 [34,35,37,44-46]), adjuncts to face-to-face mental health care (n=5 [28,29,33,36,38,41]), and standalone self-help interventions (n=4 [27,30,40,42,48]). The majority included some form of self-monitoring of symptoms, mood, emotions, behavior, or meals. The Mayo Clinic Anxiety Coach was the only app describing an active “treatment” component (ie, exposure and response prevention) although a further eight provided “coping strategies” and skills practice (eg, meditation, dialectical behavioral therapy [DBT] skills, and CBT techniques).

Two apps were available from Google Play or iTunes at the time of writing: Mayo Clinic Anxiety Coach (iTunes [42,48]) and Recovery Record (Google Play and iTunes [40]). Currently, Recovery Record has not published an evaluation of mental health outcomes but have RCTs registered to take place. As far as can be determined, none of these apps were specifically designed for use with children and young people.

XSL•FO
RenderX

Figure 1. PRISMA flow diagram of results and article selection.



Analysis Articles

A further six apps, targeted specifically at children or adolescents, were identified in two analysis articles of apps available from Google Play and iTunes [21,43]. The apps were Destructive Issues [43], Teen Depression Connect [43], Teen Hotline [43], Primary School Assessments [21], Preschooler Assessments [21], and Your Child's Social Health [21]. These analyses concluded that none of these apps have been subject to research evaluation and that the content of some did not reflect best practice guidelines. Further concerns included a lack of privacy policies and lack of resources for immediate help for those who are distressed.

Study Participants

Ages of those involved in studies ranged from 9 to 30 years with 13 articles including only children and adolescents 18 years or younger [26,28-30,33,34,36-39,42,47,48]. Demographic data including gender were sparsely reported. In total, 473 participants had used a mental health app as part of a feasibility or outcome study [26-28,31-37,44-48]. Of those younger than 18 years (n=316), only a small minority (22/316, 7.0%) had a recognized mental health problem identified by diagnostic interviews or screening questionnaires. In total, 95 adolescents had taken part in studies that evaluated their opinions on mHealth apps or prototypes without using the app itself [28,30,34,38,39]. Only 30 of 95 participants who took part in these studies (32%) had recognized mental health problems.

XSL•FO
RenderX

Table 1. Characteristics of publications of mental health mobile apps for preadolescents and adolescents included in review (N=24).

Study	Design[a]	Sample[a]	App
Aguirre et al (2013) [43]	App analysis: mobile apps for suicide prevention from Google Play and iOS	27 apps identified, 3 apps for children and young people	Destructive Issues, Teen Depression, and Teen Hotline
Kauer et al (2012) [44]	Outcome study: RCT of Mobiletype app vs abbreviated Mobiletype app	N=114 (68 intervention; 46 control) aged 14-24; GP-based recruitment	Mobiletype
Kennard et al (2015) [38]	App design: semistructured interviews gaining perspectives on a mobile safety plan for suicide prevention	N=10 teens aged 14-17 hospitalized for suicidality; n=10 parents	Safety Plan App
Kenny et al (2014) [39]	App design: focus groups gaining perspectives on mental health mobile apps and CopeSmart prototype	N=34, aged 15-16, school-based sample	CopeSmart
Kenny et al (2015) [26]	Feasibility: CopeSmart used to rate mood for 1 week	N=43, aged 15-17, school-based sample	CopeSmart
Løventoft et al (2012) [27]	Feasibility: describes design workshops and 4-week pilot trial	N=6 (aged 17-24); used psychotropic medication within last 2 years; community recruited	Daybuilder
Matthews & Doherty (2011) [28]	Feasibility: comprises 3 studies (1) initial design consultations, (2) nonclinical feasibility, (3) feasibility with clinical population	(1) n=6, (2) n=73 (21 app, 51 paper diary), (3) n=9 children seeing a therapist for a range of mental health problems	Mobile Mood Diary
Matthews et al (2008) [29]	Feasibility: app or paper-based mood charting; instructed to complete one mood entry every day for 2 weeks	N=73 (21 app, 51 paper diary), aged 13-17 years; school-recruited sample	Mobile Mood Diary
McManama et al (2016) [30]	Feasibility: pilot testing of prototype of app for suicide prevention following acute care discharge; think-aloud protocol	N=20 aged 13-18, history of suicidal thoughts and n=20 parents; outpatient psychiatry dept-recruited sample	Crisis Care
Nicolas et al (2015) [21]	App analysis: mobile apps (English language) for bipolar disorder from the Australian Google Play and iOS in 2014	82 apps identified 3 specifically for children and young people	Primary School Assessments, Preschooler Assessments & Your Child's Social Health
Niendam et al (2015) [31]	Feasibility: 4-month trial collecting medication adherence and clinical data using mobile phone app	N=36, aged 14-30; Early Psychosis participants recruited from early intervention programs	No name
Patwardhan et al (2015) [32]	Feasibility: pilot of REACH app; 30 minutes of app usage with researcher	N=22 (mean age=9.67 years); school-based recruitment	The REACH app
Pramana et al (2014) [33]	Feasibility: used for 8-16 alongside face-to-face CBT for anxiety	N=9 (aged 9-14), receiving face-to-face CBT for diagnosed anxiety disorder	SmartCat
Reid et al (2009) [34]	Feasibility: focus group and 1-week trial of Mobiletype; text prompt to complete diary 4 times a day	N=29 (n=11 in focus group, n=18 in study), aged 14-17; school-based recruitment	Mobiletype
Reid et al (2011) [45]	Outcome study: RCT of Mobiletype app vs abbreviated Mobiletype app	N=114 (68 intervention; 46 control) aged 14-24; GP-based recruitment	Mobiletype
Reid et al (2012) [35]	Feasibility: youth asked to self-monitor with app at least once a day for 2-4 weeks until next medical review	N=47 (aged 14-19), recruited from health clinic by pediatrician	Mobiletype
Reid et al (2013) [46]	Outcome study: RCT of Mobiletype app vs abbreviated Mobiletype app	N=114 (68 intervention; 46 control) aged 14-24; GP-based recruitment	Mobiletype
Scotti (2014) [36]	Feasibility: school-based DBT skills group + mobile or online tracking of skills usage	N=7 (aged 13-18), 2 of which used the app; had eating disorder or body image concerns; school-based recruitment	No name
Tregarthen et al (2015) [40]	App design: app made available to download and user information recorded	Ages ranged from 13-77 years	Recovery Record

XSL•FO
RenderX

Study	Design[a]	Sample[a]	App
Veldhuis (2014) [47]	Outcome study: app for body image or neutral app used in laboratory for 30 minutes	N=206 (aged 12-18); school-based recruitment	Pretty
Verstappen et al (2014) [41]	App design: development of ACT app for youth with depression learning ACT	Mentions "research clients" as a group of 15 "youth" undertaking 3-month ACT program at health center	The ACT app
Whitehouse et al (2013) [37]	Feasibility: piloting use of psychosocial screening app in a medical hospital setting	N=80 medical patients aged 12-18; recruited in medical clinics before appointments	TickIT
Whiteside et al (2014) [48]	Outcome study: case studies of two children with OCD using Mayo Clinic Anxiety Coach alongside face-to-face therapy for 3 months	N=2 (10 and 16 years) both diagnosed with OCD; mental health clinic-recruited	Mayo Clinic Anxiety Coach
Whiteside et al (2016) [42]	App design: user data from downloaders of Mayo Clinic Anxiety Coach	User data: children and adolescents 5-17 downloaded Mayo Clinic Anxiety Coach (likely with parents)	Mayo Clinic Anxiety Coach

[a] ACT: acceptance and commitment therapy; app analysis: article on app analysis; app design: article on app design and development; DBT: dialectical behavioral therapy; GP: general practitioner; OCD: obsessive-compulsive disorder; outcome study: study reporting mental health outcomes.

Mental Health Outcomes: Efficacy

As evident in Table 2, the included apps targeted a range of mental health areas. However, only five articles reported any mental health outcome data. These articles evaluated three apps targeting depression, stress, anxiety (Mobiletype [44-46]), body image and self-esteem (Pretty [47]), and obsessive-compulsive disorder (OCD; Mayo Clinic Anxiety Coach [48]). None of the other targeted areas had any outcome data associated with them.

Depression, Stress, and Anxiety

To date, Mobiletype is the only mobile app to have undergone a RCT [44-46]. Participants (N=118, aged 14-24 years) with emotional or mental health issues were recruited from general practitioner practices and randomly assigned to a full or an abbreviated version of Mobiletype (ie, no mental health self-monitoring). There were no significant differences between groups at posttest or follow-up (6 weeks) on depression (d= 0.09, P=.69), anxiety (d= 0.07, P=.76), or stress (d= 0.22, P=.32) as measured by the Depression, Anxiety and Stress Scale (DASS). Significant total group mean decreases on the DASS over time indicated significant reductions in depression, anxiety, and stress scores at follow-up regardless of group [44-46].

Body Image and Self-Esteem

Veldhuis [47] reported a laboratory-based, randomized trial of a mobile phone app (Pretty) to improve body image in a community sample of 206 adolescent girls (age 12-18 years, mean 13.88, SD 1.34 years). Participants were randomized to use Pretty or a comparison app for 30 minutes. Both apps presented pictures of models. Pretty asked users to rate the weight status of each model, whereas the comparison app asked neutral questions about a famous Dutch singing duo. There were no significant differences between apps on measures of self-esteem or body satisfaction postapp exposure (t_1=0.02, P=.90 and t_1=0.54, P=.46, respectively). Neither app improved body satisfaction; however, significant improvements in self-esteem postapp use were revealed, regardless of which app was used (t_{20}=–4.26, P=.001).

Obsessive-Compulsive Disorder

Whiteside [48] presented two case studies (ages 10 and 16 years) illustrating the treatment of pediatric OCD augmented with the Mayo Clinic Anxiety Coach App. Posttreatment assessments were at 4 and 3 months, respectively. Although symptoms on the Children's Yale-Brown Obsessive-Compulsive Scale were reduced, one child still met diagnostic criteria for OCD [48].

Feasibility and Acceptability

Feasibility outcomes of app usage and acceptability were extracted from studies assessing the following apps: CopeSmart [26,39], Crisis Care [30], Daybuilder [27], Mobiletype [34,35,44], Mobile Mood Diary [28,29], SmartCAT [33], REACH app [32], and TickiT [37]. Qualitative feedback from adolescents and therapists [26,28,29,39] was also reported.

App Usage

In the Mobiletype RCT [44,45], app users (N=68) were instructed to use the app at least twice a day for a minimum of 2 weeks. App use was good with participants completing a mean 3.3 (SD 1.4, range 1-8) Mobiletype entries each day and average app usage of mean 14.6 (SD 6.3, range 1-34) days. In a feasibility trial of CopeSmart [26], a nonclinical sample of adolescents (N=43) used the app for a mean 4.0 of 7.0 (SD 1.8) days. The "Rate My Mood" section was most frequently used (mean 3.5, SD 1.0 days), whereas use of the "Coping Tips" and "Resources" sections were low (mean 1.5, SD 1.0 days and mean 0.9, SD 1.0 days, respectively).

XSL•FO
RenderX

Table 2. Characteristics of mental health mobile apps for preadolescents and adolescents included in review (N=15).

App name	Description[a]	Main features[a]	OS[b]	Available to download[c]	Area targeted[d]
CopeSmart [26,39]	App to foster positive mental health in children and young people	Self-monitoring of mood, mood diary, coping tips, and contact details of mental health support services	Android & iOS	NA	Mental well-being
Crisis Care [30]	App for suicide prevention in children and young people to be downloaded on discharge from acute care	Coping skills (relaxation, behavioral activation, positive affect) and contact details of suicide hotline and adults they trust	Prototype /NR	NA	Suicide prevention
Daybuilder [27]	A "life management app" for people with depression	Symptom assessment, mood, appetite, and sleep self-monitoring, functions to let the user create events and reminders for what to do to prepare for that event, medication management	Android	NA	Depression
Mayo Clinic Anxiety Coach [42,48]	A self-help tool delivering CBT for a range of anxiety disorders	Self-monitoring, symptom assessment, psychoeducation, and treatment based on exposure therapy	iOS	Yes	OCD
Mobiletype [34,35,44-46]	A "mental health assessment and management app" for children and young people	Self-monitoring tool; prompts users 4 times a day to record mood, stressful events, alcohol use, cannabis use, quality and quantity of sleep, quantity and type of exercise, and diet	Cross-platform	NA	Mental health
Mobile Mood Diary [28,29]	App for children and young people in therapy to chart their mood	Self-monitoring of mood, sleep, and energy and a free text diary entry; no password protection or reminders	Cross-platform	NA	Mental health
Pretty [47]	Gamified app to prevent body image issues in children and young people	App is a series of pictures of models of various sizes and questions asking the user to rate each model's weight status to be either "extremely thin," "thin," "normal," "big," or "extremely big;" user gets feedback on whether their response was correct	Android & iOS	NA	Body image
REACH app [32]	App for anxiety prevention and early intervention in children and young people	Self-monitoring, resources, coping strategies, and CBT skills training	Android	NA	Anxiety
Recovery Record [40]	A CBT-based app for eating disorders self-monitoring	Self-monitoring of meals and symptoms, goal setting, coping tactics, meal plans, rewards and affirmations, social support, summative feedback	Android & iOS	Yes	Eating disorders
Safety Plan app [38]	Proposed app to support children and young people transitioning from inpatient to outpatient care	Intended to provide mobile access to pre-agreed safety plan for use in times of crisis and suicidal ideation	Prototype /NR	NA	Suicide prevention
SmartCAT [33]	App for children and young people with anxiety alongside brief CBT sessions	Skills coach, reward bank, media library, notifications, and secure messaging portal for use with therapist	Android	NA	Anxiety
The ACT app [41]	App for children and young people with depression attending therapy	Self-monitoring and symptom assessment, skills training, goal setting; based on acceptance and commitment therapy.	Android	NA	Depression
TickiT [37]	App-based psychosocial screening tool developed for children and young people attending hospital	Patients enter data in waiting room and the tool records response data, generating a report and alerts for clinicians, shifting clinical focus of the meeting	iOS	NA	Depression (screening)
No name [31]	App for recording medication adherence and symptoms in early psychosis care	Self-monitoring and symptom assessment; designed with daily and weekly surveys assessing symptoms, mood, medication adherence, and social contact	NR	NA	Early psychosis (medication adherence)

XSL•FO
RenderX

App name	Description[a]	Main features[a]	OS[b]	Available to download[c]	Area targeted[d]
No name [36]	App for recording behaviors and skills practice, adjunct to group DBT	Self-monitoring and tracking of DBT skills and ED behaviors via mobile app or online	NR	NA	Eating disorders

[a] CBT: cognitive behavioral therapy; DBT: dialectical behavioral therapy; ED: eating disorders.

[b] Cross-platform: article reports as JavaME app (Mobile Mood Diary) or "multiple models and firmware" (Mobiletype); NR: not reported; OS=operating system.

[c] NA: not available to download from Google Play, iTunes App Store, or Microsoft app store.

[d] Mental health: range of unspecified mental health problems.

In a feasibility trial of Mobile Mood Diary [28,29], a nonclinical sample of school children (N=73, aged 13-17 years) were asked to record at least one mood each day for 2 weeks. App users demonstrated significantly higher levels of compliance (entries: mean 8.12) compared to a control group who mood-charted with a pen and paper (mean 5.44). In a small pilot study [28] with a clinical population of children (N=9; age: mean 13.78, SD 2.63 years) attending therapy, mood diary adherence was 65% on average. All participants used Mobile Mood Diary for a minimum of two sessions and 8 of 9 (89%) used it for longer. In a SmartCAT feasibility trial [33], clinically anxious youth (N=9, aged 9-14 years) demonstrated good compliance, completing a mean 5.36 (SD 1.95) entries of 6.48 requests (5.36/6.48, 83% adherence rate) between each session. There was limited data about the longer-term use of apps. SmartCAT was highly utilized during week 1, but leveled off over time and almost halving by week 7 [33]. Similarly, in a feasibility trial of Mobiletype [35], participants (N = 47, age 14-24 years) completed 91% (47/51) of the Mobiletype entries every day in week 1, dropping to 58% (17/29) in week 4.

App Acceptability

Sample sizes were small, but overall app acceptability was good. The majority of CopeSmart users in the feasibility trial [26] found the app easy to use (40/43, 93%), 30 of 43 (70%) would use it in the future, 32 of 43 (74%) felt other young people would use it, and 30 of 43 (70%) would recommend it to a friend. Similarly, SmartCAT feasibility study participants [33] rated the app as highly usable (mean 1.7 on a scale of 0-7 with 1 indicating easy to use). All users reported being satisfied with SmartCAT and would recommend to others. The REACH app [32] (N = 22, age mean 9.67 years) was rated highly on ease of use, quality of support information, ease of learning, and system satisfaction with an overall mean usability score of 35.69 (SD 19.84) out of a possible score of 40. Participants who had no knowledge of the Android operating system rated the app worse. Users (N=21) of Mobile Mood Diary [29] also found it easy to use (mean 1.63, SD 0.76, where 1=very easy and 5=very difficult). Furthermore, 20 of 21 (95%) felt they had sufficient privacy and felt more privacy recording moods via the app compared to paper-based charting [29]. A feasibility study of TickiT (N=78, age range 12-18 years) demonstrated the app was easy to understand (72/78, 92%), easy to use (72/78, 92%), and efficient (63/79, 80%), with a completion rate of less than 10 minutes [37]. For participants in the Mobiletype feasibility study, 21 of 22 (95%) reported the feedback information reflected their actual experiences, was accurate (95%, 20/21), was helpful to them (71%, 17/24), and aided their doctor to understand them better (82%, 18/22) [35]. Usability of Crisis Care in a pilot study [30] was judged to be good (N=20). Mean scores on usability, utility in crisis, and content satisfaction ranged from 2 to 5 (5 being maximum score on subscales of System Usability Scale).

Adolescent Perspectives

Feedback from a focus group of nonclinical adolescents (N=34, age 15-16 years) highlighted the importance of apps being discrete and easy to conceal in order to avoid the stigma associated with mental health problems [39]. Privacy concerns were also highlighted by three participants from a clinical sample who declined to use Mobile Mood Diary in a pilot study because the title would be visible on their phone (eg, "one 16-year-old would not install the diary because her friends sometimes use her phone and she is afraid they will see an application named 'mood diary'" p 2954 [28]). Others report mobile apps offer increased privacy and discretion for activities, such as mood charting. A participant in the Mobile Mood Diary feasibility trial commented, "You can conceal more easily so there is more privacy" (p 123 [29]). On a practical level, adolescents would like apps to have password protection and to allow control over privacy settings [28,29,39]. Other feedback highlighted that apps should also be engaging, interactive, provide concise information, be esthetically attractive, allow for personalization, and provide reminders to use [26,28,29].

Therapist Perspectives

A survey [28] of therapist attitudes to mobile technology (N=28) revealed concerns about privacy and security. The "danger of someone else accessing confidential information" was a substantial therapist concern. Other concerns included increased responsibility, increased workloads, costs of implementation, need for training, setting clear boundaries between sessions, and a worry that clients would expect the therapist to continuously monitor their mood data [28]. Therapists (n=3) who used Mobile Mood Diary in a pilot study reported the app and printouts helped engage patients in therapeutic tasks, facilitated a less threatening disclosure of information, and broke down barriers in sessions. Mobile mood charting was perceived to be better than paper-based charting and printouts were useful for discussing clinical cases and saved therapist time inputting into a computer [28]. Lack of technical confidence was reported to be the greatest barrier to uptake of Mobile Mood Diary. Some therapists were incentivized to use the app when they saw others successfully using it.

XSL•FO
RenderX

Discussion

Principal Results and Comparisons With Previous Work

The aim of this review was to systematically examine the literature on mobile apps for mental health in children and young people. Our review identified 24 papers describing 15 apps or prototypes, two of which were available to download from Google Play or iTunes [40,42,48]. We identified only two small RCTs [44-47], one of which was a laboratory-based experimental study [47], and both failed to demonstrate a significant effect on their intended outcomes (depression or body image). Therefore, we conclude that currently there is no evidence to support the effectiveness of apps designed for adolescents with mental health problems. Because we did not identify any study participants younger than 9 years, there is no evidence to support the effectiveness of apps designed for children with mental health problems either.

Our conclusion is consistent with previous reviews and highlights that the evidence base has barely increased over the past 4 years [18,23-25]. The lack of empirical studies contrasts starkly with the commercial development of mobile apps. From October 2013 to June 2016, the number of apps available to download from the app store doubled to 2 million [49], of which 1.98% (39,600) were classed as medical. Given the significant increase in the availability of mobile apps, the lack of evidence to support their safety or effectiveness with vulnerable populations is concerning.

Although the evidence base is currently lacking, this does not rule out the fact that well-designed, adequately tested, evidenced-based mobile apps could be effective. The evidence base for the clinical effectiveness of mobile apps in adult mental health is slowly emerging [18,50-52]. Our review suggests that the comparative literature for children and adolescents is significantly lagging, a trend also noted within the literature on other forms of eHealth, such as computerized CBT [6,7].

In terms of acceptability, it has been suggested that apps and eHealth in general are particularly suited for adolescents who are familiar with and regular users of technology [23]. We identified 12 small pilot feasibility trials [26-37] that suggested, in the short term, adolescents may be favorably disposed to this form of delivery. Acceptability was generally positive with ratings of ease of use, satisfaction, and usability rated average to high [26,29,32,33,35,37]. The privacy and discretion afforded by well-designed apps were of importance to young people [28,29,39]. However, many participants were healthy, nonreferred adolescents and less is known about whether those with mental health problems would have similar views. It is important to also note that although adolescents may have positive attitudes toward mHealth, it does not necessarily mean they would prefer it over a face-to-face intervention [24]. It is also important to consider whether the affinity that people have for their mobile phones and the trust and expectations placed in them positively influence clinical outcomes and user satisfaction [53]. This "digital placebo effect" may account for why some people continue to download and use mobile apps for mental health even though the evidence base is largely absent [53]. Nonetheless, our findings support previous conclusions and suggest that apps may provide an acceptable way of supporting mental health interventions for some adolescents [18,54].

Therapist perspectives on mobile apps were mixed, with concerns relating to patient security, increased responsibility and workloads, and the need to set clear boundaries between sessions [28]. These are different concerns to those surrounding the use of other forms of eHealth, such as computerized CBT, in which clinicians were concerned about the effectiveness of computerized CBT with more severe mental health problems and the lack of a therapeutic relationship [55]. This may reflect differences in the purpose of these interventions (ie, computerized CBT being a therapeutic intervention compared to apps that are an adjunct to therapy). Interestingly, therapists who used the app Mobile Mood Diary in clinical practice reported benefits such as facilitation of client engagement [28]. Lack of technical confidence was the most common barrier to implementation. This lack of technical confidence may be addressed by improving the user-friendliness of the app, either by codesigning apps with therapists or providing training for therapists.

App usage, where reported, was moderate and adherence ranged from 65% to 83%, which is comparable to those seen in Internet interventions for depression and anxiety [56]. There was a suggestion that self-monitoring of mood via apps promoted higher adherence compared to paper self-monitoring [33,35]. Information on longer-term usage is scarce, but the included studies suggest app usage begins high and declines over time [33,35]. This "law of attrition" [57] is also a common challenge for computerized CBT and eHealth interventions [57]. As with these other technology-based interventions, using mHealth apps with support from a therapist offers one strategy for increasing longer-term engagement [24,58]. Indeed, the SmartCAT app used in conjunction with face-to-face support demonstrated an 83% completion rate [33], similar to completion rates demonstrated in face-to-face CBT (84%) and guided Internet CBT (81%) [59]. Making mHealth apps inherently more engaging by design is another strategy for increasing longer-term engagement. One promising proposition is the use of serious gaming, gamification principles, telepresence, and persuasive technology in eHealth (and by extension mHealth) design [58,60]. The evidence base for the benefits of these principles as applied to mHealth and eHealth is currently in its infancy, however, and is a burgeoning area of research [60].

This review highlights several methodological concerns about the quality of the research evidence for mental health mobile apps, especially those for adolescents. Sample sizes tend to be small and reporting of demographic data such as gender and age inadequate, particularly in pilot feasibility studies. Few participants have an identified mental health problem and, as such, little is known about the acceptability and use of apps with clinical groups. As far as can be determined, the youngest participant in these studies was 9-years-old, meaning there is no research evidence for the use of mobile apps in children younger than this age. Where reported, symptoms tended to be mild to moderate in severity and, as such, the appropriateness of mobile apps for complex or more severe problems is

XSL•FO
RenderX

unknown. Studies tend to be short in duration and there is sparse information on whether positive gains from using mobile apps are maintained. Finally, none of the apps in this review have been evaluated using a suitable RCT comparing a mobile app to an adequate control group. Future research should address these methodological concerns. Given the beneficial role that parent participation and engagement can have in adolescent mental health treatment [61], future research may also want to consider the role of parents/guardians in supporting adolescents using apps for mental health.

Our review has focused on the academic literature and of the apps identified, two of which were available to download. This contrasts starkly with the large number available from commercial sites and raises questions about the safety, quality, and efficacy of those that are available [11,22,62]. Content analyses [21,43] of six apps for children and adolescents available to download highlighted that none have been subject to any research evaluation. The authors also noted that the apps did not reflect best practice guidelines and lacked privacy policies [21,43]. Some of these apps claim to address a worrying number and type of complex problems, including "child abuse," "daughter's abusive relationship," and "teen suicide, depression, and stress." Ineffective or detrimental apps are a significant concern and incur costs to patient safety and care [11]. Therefore, our review adds to calls for better regulatory oversight to ensure app quality and safety [11,18,22,62].

Limitations

There are several limitations of this review. Firstly, the number of studies was small with generally limited sample sizes. Conclusions that mobile apps are acceptable for youth are therefore tentative. Secondly, the qualitative feedback is based on a small number of young people and therapists and generalizing their views to a wider population should be exercised with caution. This feedback is nevertheless informative and highlights the importance of involving young people in app design. Thirdly, we aimed to reduce publication bias, and although our inclusion criteria were broad, our search was limited to English-language papers. Fourthly, despite aiming for a precise overview of the literature on mobile apps for children and adolescents, a number of publications included adults. The majority of publications utilized teenage and young adult populations with only one study including a participant aged 9 years. As such, our results are limited to preadolescents and adolescents, rather than children. All the articles included in this review originated from work in North America, Northern Europe, and Australia; therefore, these results are limited to the experiences of adolescents in high-income countries. mHealth holds great promise for widening access to mental health treatment in low to upper-middle income countries where the challenges of meeting mental health needs are considerable [63]. This potential will not be realized unless future research is conducted in these contexts.

Conclusion

There is an urgent need for methodologically robust, adequately powered research evaluating the safety, efficacy, and effectiveness of mental health apps for children and young people with mental health problems. Well-designed RCTs with adequate power and control groups are needed to demonstrate whether mobile apps for mental health have any clinical benefit for children and young people. Because the development of apps is vastly outpacing the development of the evidence base, future research should also utilize quicker, good-quality designs [64]. This may require the inclusion of adolescents and therapists in the app design and development process to ensure apps are fit for purpose and user-centered [58], as well as continuous evaluation of evolving interventions [64]. At present, there is insufficient evidence to suggest that any mobile app for mental health can be used effectively with children and young people. Clinicians should be cautious about recommending mobile apps until there is sufficient evidence to support their safety and efficacy.

Authors' Contributions

RG codesigned the methodology, conducted the literature searches and data analysis, and drafted the manuscript. PS was responsible for review conception and methodology, data interpretation, and read and contributed to the manuscript. JP read and contributed to the manuscript. All authors read, contributed to, and approved the final manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of databases and search strings used for systematic review.

[PDF File (Adobe PDF File), 35KB - jmir_v19i5e176_app1.pdf]

References

1. Green H, McGinnity A, Meltzer H, Ford T, Goodman R. Mental Health of Children and Young People in Great Britain, 2004. Hampshire, UK: Palgrave MacMillan; 2005. URL: http://content.digital.nhs.uk/catalogue/PUB06116/ment-heal-chil-youn-peop-gb-2004-rep1.pdf [accessed 2017-01-10] [WebCite Cache ID 6nR8e7xCU]

XSL•FO
RenderX

2. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
3. Kim-Cohen J, Caspi A, Moffitt TE, Harrington H, Milne BJ, Poulton R. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. Arch Gen Psychiatry 2003 Jul;60(7):709-717. [doi: 10.1001/archpsyc.60.7.709] [Medline: 12860775]
4. Ford T, Mitrofan O, Wolpert M. Annual Report of the Chief Medical Officer 2013. Public Mental Health Priorities: Investing in the Evidence. London: Department of Health; 2014. Life course: children and young people's mental health URL: http://www.psymaptic.org/files/2014/09/CMO-Annual-Report-2013-Public-Mental-Health.pdf [accessed 2017-01-10] [WebCite Cache ID 6nPjWH9Wr]
5. Woodward LJ, Fergusson DM. Life course outcomes of young people with anxiety disorders in adolescence. J Am Acad Child Adolesc Psychiatry 2001 Sep;40(9):1086-1093. [doi: 10.1097/00004583-200109000-00018] [Medline: 11556633]
6. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
7. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
8. Ericsson. 2016 Nov. Ericsson mobility report: on the pulse of the networked society URL: https://www.ericsson.com/mobility-report [accessed 2017-01-10] [WebCite Cache ID 6nPXPKRxA]
9. eMarketer. 2013 May 22. Smartphones are increasingly UK kids' mobile phone of choice limited-function mobile phones may face a tough sell URL: https://www.emarketer.com/Article/Smartphones-Increasingly-UK-Kids-Mobile-Phone-of-Choice/1009910 [accessed 2017-01-10] [WebCite Cache ID 6nPXuF4e4]
10. Ofcom. 2014 Oct 08. Children and parents: media use and attitudes report URL: https://www.ofcom.org.uk/research-and-data/media-literacy-research/children/children-parents-oct-14 [accessed 2017-01-10] [WebCite Cache ID 6nPYLrMYE]
11. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
12. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 7;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
13. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [Medline: 25994025]
14. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
15. 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]
16. The National Institute for Health Research Horizon Scanning Research Intelligence Centre. 2015 Nov. New and emerging mobile health interventions that promote behavioural change URL: http://www.hsric.nihr.ac.uk/topics/review-new-and-emerging-mobile-health-interventions-that-promote-behavioural-change/ [accessed 2017-01-10] [WebCite Cache ID 6nPZ29P0g]
17. East ML, Havard BC. Mental health mobile apps: from infusion to diffusion in the mental health social system. JMIR Ment Health 2015;2(1):e10 [FREE Full text] [doi: 10.2196/mental.3954] [Medline: 26543907]
18. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
19. Martínez-Pérez B, De La Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the world health organization: Review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120. [doi: 10.2196/jmir.2600] [Medline: 23770578]
20. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
21. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
22. Leigh S, Flatt S. App-based psychological interventions: friend or foe? Evid Based Ment Health 2015 Nov;18(4):97-99. [doi: 10.1136/eb-2015-102203] [Medline: 26459466]
23. National Collaborating Centre for Mental Health. 2014 Mar. E-therapies systematic review for children and young people with mental health problems URL: http://www.e-lfh.org.uk/programmes/minded/general-resources/ [WebCite Cache ID 6nPea7xEv]

XSL•FO
RenderX

24. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems: a systematic and meta-review. J Child Psychol Psychiatry 2016 Dec 10:1. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
25. Archangeli C, Marti FA, Wobga-Pasiah EA, Zima B. Mobile health interventions for psychiatric conditions in children: a scoping review. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):13-31. [doi: 10.1016/j.chc.2016.07.009] [Medline: 27837939]
26. Kenny R, Dooley B, Fitzgerald A. Feasibility of “CopeSmart”: a telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
27. Løventoft P, Nørregaard LB, Frøkjær E. Designing daybuilder: an experimental app to support people with depression. In: Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases. 2012 Presented at: PDC'12; Aug 12-16, 2012; Roskilde, Denmark p. 1-4. [doi: 10.1145/2348144.2348146]
28. Matthews M, Doherty G. In the mood: engaging teenagers in psychotherapy using mobile phones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI 11; May 7-12, 2011; Vancouver, BC p. 2947-2956 URL: http://doi.org/10.1145/1978942.1979379 [doi: 10.1145/1978942.1979379]
29. Matthews M, Doherty G, Sharry J, Fitzpatrick C. Mobile phone mood charting for adolescents. Brit J Guid Couns 2008 May;36(2):113-129 [FREE Full text] [doi: 10.1080/03069880801926400]
30. McManama O'Brien KH, LeCloux M, Ross A, Gironda C, Wharff EA. A pilot study of the acceptability and usability of a smartphone application intervention for suicidal adolescents and their parents. Arch Suicide Res 2016 May 2;2:1-11. [doi: 10.1080/13811118.2016.1182094] [Medline: 27136115]
31. Niendam T, Iosif AM, Tully LM, Burch K, Carter C. Preliminary longitudinal study examining the clinical correlates of medication adherence assessed via a mobile health application in early psychosis care. In: Neuropsychopharmacology. 2015 Dec Presented at: ACNP 54th Annual Meeting; Dec 6-10, 2015; Hollywood, FL p. S473-S647. [doi: 10.1038/npp.2014.282]
32. Patwardhan M, Stoll R, Hamel D, Amresh A, Gary KA, Pina A. Designing a mobile application to support the indicated preventionearly intervention of childhood anxiety. In: Proceedings of the Conference on Wireless Health. 2015 Presented at: WH'15; Oct 14-16, 2015; Bethesda, MD URL: http://doi.org/10.1145/2811780.2811954 [doi: 10.1145/2811780.2811954]
33. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014 May;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
34. Reid SC, Kauer SD, Dudgeon P, Sanci LA, Shrier LA, Patton GC. A mobile phone program to track young people's experiences of mood, stress and coping. Development and testing of the mobiletype program. Soc Psychiatry Psychiatr Epidemiol 2009 Jun;44(6):501-507. [doi: 10.1007/s00127-008-0455-5] [Medline: 19011723]
35. Reid SC, Kauer SD, Khor AS, Hearps SJ, Sanci LA, Kennedy AD, et al. Using a mobile phone application in youth mental health-an evaluation study. Aust Fam Physician 2012 Sep;41(9):711-714. [Medline: 22962650]
36. Scotti J. ScholarWorks at University of Montana. 2014. School-based DBT skills groups for adolescent eating disorders and body image concerns: a pilot study [doctoral dissertation] URL: http://scholarworks.umt.edu/cgi/viewcontent.cgi?article=5482&context=etd [accessed 2017-01-10] [WebCite Cache ID 6pB2yfvXg]
37. Whitehouse SR, Lam P, Balka E, McLellan S, Deevska M, Penn D, et al. Co-creation with TickiT: designing and evaluating a clinical eHealth platform for youth. JMIR Res Protoc 2013 Oct 18;2(2):e42 [FREE Full text] [doi: 10.2196/resprot.2865] [Medline: 24140595]
38. Kennard BD, Biernesser C, Wolfe KL, Foxwell AA, Craddock LS, Rial KV, et al. Developing a brief suicide prevention intervention and mobile phone application: a qualitative report. J Technol Hum Serv 2015 Oct 01;33(4):345-357 [FREE Full text] [doi: 10.1080/15228835.2015.1106384] [Medline: 26977137]
39. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2014 Nov 10;22(2):265-275. [doi: 10.1177/1460458214555041] [Medline: 25385165]
40. Tregarthen JP, Lock J, Darcy AM. Development of a smartphone application for eating disorder self-monitoring. Int J Eat Disord 2015 Nov;48(7):972-982. [doi: 10.1002/eat.22386] [Medline: 26213130]
41. Verstappen M, Gardner P, Poon D, Bettridge T. Off the couch and out of the hospital, mobile applications for acceptance and commitment therapy. In: Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2014 Presented at: MobileHCI'14; Sept 23-26, 2014; Toronto, ON p. 431-434. [doi: 10.1145/2628363.2633573]
42. Whiteside SP. Mobile device-based applications for childhood anxiety disorders. J Child Adolesc Psychopharmacol 2016 Apr;26(3):246-251. [doi: 10.1089/cap.2015.0010] [Medline: 26244903]
43. Aguirre R, McCoy M, Roan M. Development guidelines from a study of suicide prevention mobile applications (apps). J Technol Hum Serv 2013 Jul;31(3):269-293 [FREE Full text] [doi: 10.1080/15228835.2013.814750]
44. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]

XSL•FO
RenderX

45. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011;12:131 [FREE Full text] [doi: 10.1186/1471-2296-12-131] [Medline: 22123031]
46. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: health service outcomes from a randomised controlled trial of mobiletype. BMC Fam Pract 2013;14:84 [FREE Full text] [doi: 10.1186/1471-2296-14-84] [Medline: 23782796]
47. Veldhuis J. Vrije Universiteit Amsterdam Faculteit der Social Wetenschappen. 2014. Media models matter in context: negotiated media effects of idealized body images [doctoral dissertation] URL: https://research.vu.nl/ws/portalfiles/portal/1125682 [WebCite Cache ID 6nPlGZs5W]
48. Whiteside S, Ale C, Vickers Douglas K, Tiede MS, Dammann JE. Case examples of enhancing pediatric OCD treatment with a smartphone application. Clin Case Stud 2014 Feb;13(1):80-94 [FREE Full text] [doi: 10.1177/1534650113504822]
49. Apple. 2016. Number of apps available in leading app stores as of June 2016 URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2017-01-10] [WebCite Cache ID 6nPmLiubY]
50. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
51. Dagöö J, Asplund RP, Bsenko HA, Hjerling S, Holmberg A, Westh S, et al. Cognitive behavior therapy versus interpersonal psychotherapy for social anxiety disorder delivered via smartphone and computer: a randomized controlled trial. J Anxiety Disord 2014 May;28(4):410-417. [doi: 10.1016/j.janxdis.2014.02.003] [Medline: 24731441]
52. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
53. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
54. Juarascio AS, Goldstein SP, Manasse SM, Forman EM, Butryn ML. Perceptions of the feasibility and acceptability of a smartphone application for the treatment of binge eating disorders: Qualitative feedback from a user population and clinicians. Int J Med Inform 2015 Oct;84(10):808-816 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.06.004] [Medline: 26113461]
55. Stallard P, Richardson T, Velleman S. Clinicians' attitudes towards the use of computerized cognitive behaviour therapy (cCBT) with children and adolescents. Behav Cogn Psychother 2010 Oct;38(5):545-560. [doi: 10.1017/S1352465810000421] [Medline: 20615273]
56. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
57. 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]
58. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-Therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016 Apr 18;7:7. [doi: 10.3389/fpsyt.2016.00065] [Medline: 4834305]
59. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
60. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
61. Haine-Schlagel R, Walsh NE. A review of parent participation engagement in child and family mental health treatment. Clin Child Fam Psychol Rev 2015 Jun;18(2):133-150 [FREE Full text] [doi: 10.1007/s10567-015-0182-x] [Medline: 25726421]
62. Wicks P, Chiauzzi E. 'Trust but verify'-five approaches to ensure safe medical apps. BMC Med 2015;13:205 [FREE Full text] [doi: 10.1186/s12916-015-0451-z] [Medline: 26404791]
63. Farrington C, Aristidou A, Ruggeri K. mHealth and global mental health: still waiting for the mH2 wedding? Global Health 2014;10:17 [FREE Full text] [doi: 10.1186/1744-8603-10-17] [Medline: 24670011]
64. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]

Abbreviations

CBT: cognitive behavioral therapy
DBT: dialectical behavioral therapy
DASS: Depression, Anxiety and Stress Scale
mHealth: mobile health

XSL•FO
RenderX

NHS: National Health Service
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.01.17; peer-reviewed by D Rickwood, TR Soron, S Aggarwal, E Dove, E Dogan, M Ashford; comments to author 09.03.17; revised version received 27.03.17; accepted 07.04.17; published 25.05.17.

Please cite as:
Grist R, Porter J, Stallard P
Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review
J Med Internet Res 2017;19(5):e176
URL: http://www.jmir.org/2017/5/e176/
doi:10.2196/jmir.7332
PMID:28546138

©Rebecca Grist, Joanna Porter, Paul Stallard. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 25.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Activity Recognition for Persons With Stroke Using Mobile Phone Technology: Toward Improved Performance in a Home Setting

Megan K O'Brien[1,2*], PhD; Nicholas Shawen[1*], MS; Chaithanya K Mummidisetty[1], MS; Saninder Kaur[1], MD; Xiao Bo[3], MS; Christian Poellabauer[3], PhD; Konrad Kording[2], PhD; Arun Jayaraman[1,2], PT, PhD

[1]Max Nader Lab for Rehabilitation Technologies and Outcomes Research, Rehabilitation Institute of Chicago, Chicago, IL, United States

[2]Department of Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL, United States

[3]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States

[*]these authors contributed equally

Corresponding Author:
Arun Jayaraman, PT, PhD
Max Nader Lab for Rehabilitation Technologies and Outcomes Research
Rehabilitation Institute of Chicago
345 E. Superior St.
Chicago, IL, 60611
United States
Phone: 1 312 238 6875
Fax: 1 312 238 2081
Email: a-jayaraman@northwestern.edu

Abstract

Background: Smartphones contain sensors that measure movement-related data, making them promising tools for monitoring physical activity after a stroke. Activity recognition (AR) systems are typically trained on movement data from healthy individuals collected in a laboratory setting. However, movement patterns change after a stroke (eg, gait impairment), and activities may be performed differently at home than in a lab. Thus, it is important to validate AR for gait-impaired stroke patients in a home setting for accurate clinical predictions.

Objective: In this study, we sought to evaluate AR performance in a home setting for individuals who had suffered a stroke, by using different sets of training activities. Specifically, we compared AR performance for persons with stroke while varying the origin of training data, based on either population (healthy persons or persons with stoke) or environment (laboratory or home setting).

Methods: Thirty individuals with stroke and fifteen healthy subjects performed a series of mobility-related activities, either in a laboratory or at home, while wearing a smartphone. A custom-built app collected signals from the phone's accelerometer, gyroscope, and barometer sensors, and subjects self-labeled the mobility activities. We trained a random forest AR model using either healthy or stroke activity data. Primary measures of AR performance were (1) the mean recall of activities and (2) the misclassification of stationary and ambulatory activities.

Results: A classifier trained on stroke activity data performed better than one trained on healthy activity data, improving average recall from 53% to 75%. The healthy-trained classifier performance declined with gait impairment severity, more often misclassifying ambulatory activities as stationary ones. The classifier trained on in-lab activities had a lower average recall for at-home activities (56%) than for in-lab activities collected on a different day (77%).

Conclusions: Stroke-based training data is needed for high quality AR among gait-impaired individuals with stroke. Additionally, AR systems for home and community monitoring would likely benefit from including at-home activities in the training data.

(J Med Internet Res 2017;19(5):e184) doi:10.2196/jmir.7385

KEYWORDS

smartphone; activities of daily living; ambulatory monitoring; machine learning; stroke rehabilitation

XSL•FO
RenderX

Introduction

Recovering independent mobility after a stroke, both at home and in the community, is a priority for most stroke survivors [1]. The development of therapeutic interventions to restore walking and functional recovery remains a major focus of rehabilitation for individuals after stroke [2]. However, it is difficult to know which rehabilitation strategies most improve functional mobility. As stroke units strive to optimize inpatient and outpatient care and shorten hospital stays, it becomes increasingly important to measure the impact of rehabilitation beyond the traditional clinical setting.

Monitoring daily physical activity is essential to understanding patient recovery. After all, community-dwelling stroke patients exhibit high levels of sedentary behavior [3,4], which has been identified as a risk factor for secondary cardiovascular disease and mortality [5]. Decreasing sedentary time by increasing ambulation or even standing time may reduce these risks [6,7] and prevent further health complications for persons with stroke. Wearable sensors, coupled with activity recognition (AR) models and machine learning techniques, can identify various mobility-related activities, such as sitting, lying, standing, walking, and stair use, in the clinic and in the home and community. This enables therapists to develop personalized, data-driven programs to advise patients and improve activity levels [8]. Thus, efficiently measuring mobility activities through a wearable AR system is a major quantitative outcome measure for studying new therapeutic interventions for stroke survivors.

Smartphones in particular have proven promising for unobtrusive health monitoring among patients [9], as they are now inexpensive, widely used, have an integrated system of movement-related sensors, and are able to transmit data continuously. Additionally, with access to the Global Positioning System (GPS), smartphones can allow clinicians and researchers to quantify outdoor community mobility [10-12] for tracking recovery and societal reintegration. Smartphones offer a favorable alternative to current monitoring devices such as pedometers, step activity monitors (SAMs), or other accelerometry-driven products [13], which focus exclusively on step counts and walking bouts.

A significant obstacle to the deployment of AR systems in rehabilitation is their reliability when applied to patients. Recent work has demonstrated that AR classifiers trained using in-lab data from young, able-bodied adults do not generalize to older adults [14], persons with Parkinson's disease [15], or patients with lower limb impairments [16]. Rather, using training data from the neurological population of interest notably improved AR accuracy, likely because such groups exhibit different movement patterns than a young, healthy cohort [17]. Despite some ongoing research for stroke-based AR [18], we know little about the generalization of AR classifiers from healthy subjects to stroke patients. In particular, the heterogeneous movement-related outcomes that accompany stroke, such as level of gait impairment, may affect AR performance. It remains to be seen whether activities from a healthy cohort provide sufficient training data to classify stroke activities.

Another obstacle for the deployment of AR systems in rehabilitation is their reliability when applied to at-home activities. That is, AR classifier training usually relies on activities performed in a laboratory, but the end goal for these classifiers is to detect activities performed elsewhere in the community. Considering that a laboratory is an unfamiliar environment under close researcher supervision, it stands to reason that activities performed in a lab may look different from those performed at home. It is thus critical to know whether in-lab activities provide sufficient training data to classify at-home activities or whether models should be tuned to the environment of interest.

Here, we set out to investigate AR for individuals with stroke—specifically, the dependence of AR performance on the type of training data (from stroke or healthy subjects) and the environment (laboratory or home setting). Young, healthy individuals and community ambulators with stroke wore a smartphone while performing and self-labeling various mobility-related activities. We compared the ability of AR models trained on either a healthy or stroke cohort to classify activities in stroke cohorts with different levels of gait impairment. We also compared the ability of AR models trained on activities collected in either a laboratory or a home setting to classify activities at home for a stroke cohort. This approach is an important first step in highlighting potential issues with home monitoring using traditional laboratory-based AR methods.

Methods

Participants

A total of 30 individuals with stroke (mean 60.7, SD 13.3 years; 18 males and 12 females) participated in this study. Of these subjects, 21 had ischemic strokes, and 9 had hemorrhagic strokes; 16 sustained right-side damage, and 14 sustained left-side damage. Median stroke latency was 4.6 years (range 1986-2015) at the time of the study. Exclusion criteria included severe cognitive impairment (scoring ≤17 points on the Mini–Mental State Examination [19] and physical impairment that would inhibit ability to use a smartphone. We determined gait impairment using preferred walking speed during a 10-meter walk test (10MWT). This method categorized 8 subjects with mild impairment (>0.8 m/s), 13 subjects with moderate impairment (≥0.4 and ≤0.8 m/s), and 9 subjects with severe impairment (<0.4 m/s). Fifteen healthy subjects (mean 31.1, SD 9.2 years; 4 males and 11 females) were also recruited for this study, using a sample of convenience.

All subjects provided written informed consent before participation. The study was approved by the Institutional Review Board of Northwestern University (Chicago, IL) in accordance with federal regulations, university policies, and ethical standards regarding research on human subjects.

Smartphone Sensing System

Subjects wore a Samsung Galaxy S4 running Android OS 4.4.4 on their waist in a belted pouch. The pouch was not restricted to any particular location on their waists (eg, right or left side) to make the home deployment as realistic as possible. A custom

XSL•FO
RenderX

app named CIMON [20,21] collected tri-axial accelerometer data at an average of 60 samples per second, gyroscope data at 60 samples per second, and barometer data at 6 samples per second. Subjects self-labeled various mobility-related activities through CIMON's user interface. Sensor data and labels were sent in real time via WiFi and LTE to a HIPAA-compliant (Health Insurance Portability and Accountability Act of 1996), secure server at the University of Notre Dame (Notre Dame, IN).

Activity Labeling

Subjects with stroke performed and labeled a sequence of six activities (Sitting, Lying, Standing, Stairs Up, Stairs Down, and Walking) during two in-lab sessions, and they performed another session independently at home. During their first visit to the lab, they completed the activity sequence three times between rest periods. They were then asked to perform each of the six activities at least twice at home on the following day. They returned to the lab on the third day to complete three additional sequences. The healthy subjects performed and labeled the same activities at their leisure over a two-week period, and we asked that they complete each activity at least twice per day. Before taking the phones home, all subjects were taught to use the app and labeling system, and they performed several activities under the supervision of the researchers to ensure understanding.

Labeling an activity involved removing the phone from its pouch, selecting an activity label from a dropdown menu on the user interface, replacing the phone in the pouch, and commencing with the activity. This approach generated noisy, high-frequency sensor signals unrelated to the movement of interest when removing and replacing the phone. We removed these extraneous signals using an activity-dependent threshold on sample entropy, supplemented by manual trimming, at the beginning and end of each label. We also removed any trials that the subjects had obviously mislabeled, which occasionally occurred in the Home sessions (eg, a brief "Walking" trial with nearly flat signals from the accelerometer and gyroscope sensors, suggesting that the subject selected the incorrect label and neglected to cancel the trial).

Activity Recognition

Data processing, activity recognition, and model analysis were implemented in MATLAB 2016b (MathWorks; Natick, MA). Accelerometer and gyroscope data were resampled to 50 Hz, and barometer data were resampled to 6 Hz to correct for any variability in the sensors' sampling frequencies. Each activity recording was then divided into 10-second data clips (instances) with 90% overlap.

Initially, 270 features were identified from the activity data clips. Of these features, 131 arose from the accelerometer, 131 arose from the gyroscope, and 8 arose from the barometer. Features included statistical measures of the sensor signal, its derivatives, and the frequency domain (Tables 1 and 2).

To reduce the complexity of the model, a reduced feature set was chosen using MATLAB's Out-of-Bag Predictor Importance method for a random undersampling (RUS) random forest (RF) model trained on the healthy subject data. RF models are often used in AR for their efficiency and flexibility [22,23], maintaining high accuracy for multiclass problems and large feature sets. The final feature set included 151 features, with 80 from the accelerometer, 63 from the gyroscope, and 8 from the barometer.

Table 1. Activity recognition model features from accelerometer and gyroscope signals.

Description	Number of features
Mean, range, and interquartile range	3 (per axis)
Moments: standard deviation, skew, and kurtosis	3 (per axis)
Histogram: bin counts of z-scores from -2 to 2	4 (per axis)
Moments of derivative: mean, standard deviation (SD), skew, and kurtosis	4 (per axis)
Mean of the squared norm	1
Sum of axial standard deviations	1
Pearson correlation coefficient, r(xy), r(xz), r(yz)	1 (per axis)
Mean cross products (raw and normalized), xy, xz, yz	2 (per axis)
Absolute mean of cross products (raw and normalized)	2 (per axis)
Power spectra: mean, standard deviation, skew, and kurtosis	4 (per axis)
Mean power in 0.5 Hz bins from 0-10 Hz	20 (per axis)

Table 2. Activity recognition model features from barometer signals.

Description	Number of features
Moments of derivative: mean, SD, skew, and kurtosis	4
SD, range, and interquartile range	3
Slope of linear regression	1

XSL•FO
RenderX

The same analysis was also performed using the stroke subject data, resulting in a set of 112 features, of which 102 were shared with the set of 151 features from the healthy subject data. Because of this similarity, we opted to use the set of 151 features throughout our models for simplicity. The chance of overfitting to the Healthy training set was negligible since less than half of the available features were discarded and the set overlapped substantially with features derived from the Stroke training set. Further potential for overfitting could have been removed if features were selected based on an external set not used for model training or testing, or by using an additional layer of cross-validation to the same effect.

RF tends to misclassify activities that are underrepresented in the data and can underperform on the Stairs Up and Stairs Down activities because of their rarity within the training data (respectively accounting for 1.68% (4138/246,283) and 1.51% (3723/246,823) of feature vectors). We implemented the RUSBoost algorithm [24] using decision trees (minimum leaf size=5, learn rate=1, number of trees=200) to improve model performance for an imbalanced class distribution (eg, underrepresented stair activity, overrepresented sitting activity). This number of trees was found to be sufficient for nearly full learning without overfitting. Using RUSBoost increased mean recall of Stairs Up by 6.5% and Stairs Down by 3.3% over an RF model, while showing little average change in recall over the remaining activities. Based on this finding, we used RUSBoost in the remainder of our analysis.

Model Analysis

We evaluated various AR classifiers for their performance in stroke activity recognition. We designated these classifiers as follows: (1) population models comparing the performance of a model trained using activity data from young, healthy subjects (Healthy) versus older stroke subjects (Stroke); (2) gait impairment models comparing the performance of a young, healthy training set on stroke subjects with mild versus moderate versus severe gait impairment; and (3) environmental models for stroke subjects, comparing the ability of the first in-lab training set (Lab 1) to recognize at-home activities (Home) versus the in-lab activities performed on a separate day (Lab 2). We also designed an environmental model trained and tested on only at-home activities. These model comparisons and their respective training and testing sets are depicted in Figure 1.

Population models were trained and tested on either the Healthy or Stroke subject data using leave-one-subject-out cross validation (Healthy-to-Healthy, Stroke-to-Stroke). The performance of the Healthy population model, trained on all 15 healthy subjects, was also evaluated for each Stroke subject (Healthy-to-Stroke). The Healthy-to-Healthy model yielded a baseline AR performance, against which the other models were compared.

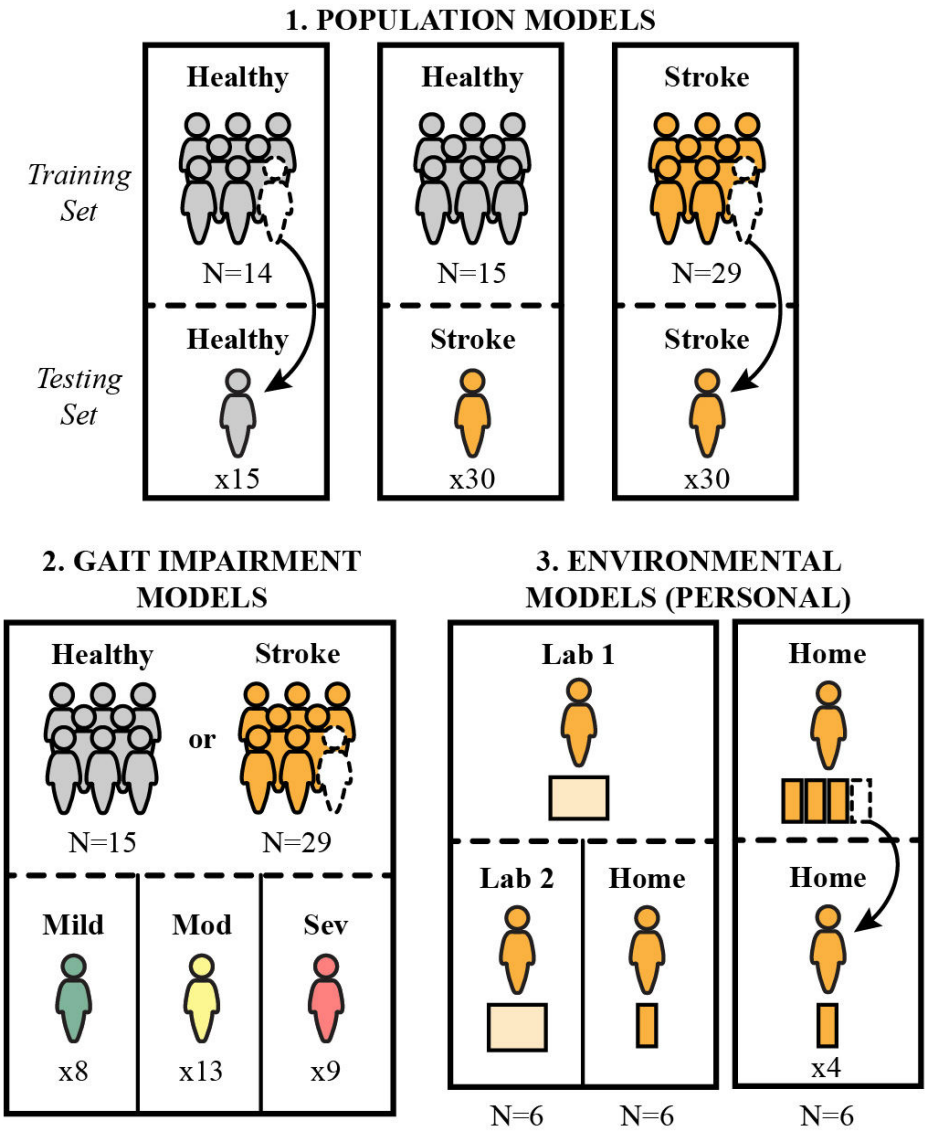
To further probe the efficacy of Healthy data in detecting activities for persons with stroke, the Healthy-to-Stroke model results were separated for different subgroups of the stroke population based on gait impairment (Mild, Moderate, and Severe). We hypothesized that AR models trained on Healthy activities would perform better for subjects with less gait impairment (ie, a Healthy-to-Mild model would perform better than a Healthy-to-Severe model) and that this improved performance would be more pronounced for ambulatory activities (Stairs Up, Stairs Down, and Walking). We performed a similar analysis using the Stroke population training data to determine whether the performance of the Stroke population model is affected by level of gait impairment. We hypothesized that the mixed-impairment Stroke population model would perform similarly across impairment subgroups.

Environmental models assessed the capacity of training data collected in a laboratory setting to classify activities performed at home. We used personal models trained on Lab 1 data and tested on Lab 2 data for a baseline comparison (Lab-1-to-Lab 2), as well as Home data (Lab 1-to-Home). For the same set of subjects, we designed personal models trained on Home data and tested on other Home data via a four-fold cross validation (Home-to-Home). In this model, the Home data were divided into four folds chronologically rather than randomly, selected to ensure no overlap between adjacent folds. We used only subjects who had at least 60 seconds of each activity in each environmental setting (Lab 1, Home, and Lab 2), which limited the pool to six subjects. With this smaller dataset, we reduced the number of classes to four (combining Sitting and Lying as well as Stairs Up and Stairs Down) to focus on the broader misclassification of stationary and ambulatory activities. We hypothesized that the Lab 1-to-Lab 2 model and the Home-to-Home model would perform better than the Lab 1-to-Home model, expecting differences to arise in subject behavior and in the relative distribution of activities when performing activities at home versus in the lab.

We focused on personal models for the most direct comparison of AR efficacy in the Lab and Home environments. In a practical implementation of AR, a global model—with a training set based on multiple subjects—would be used to classify activities from a new subject, as in the population models described above. In order to represent this use-case scenario, we also examined global models for the three environmental analyses: Lab 1-to-Lab 2, Lab 1-to-Home, and Home-to-Home. The environmental global models were implemented using leave-one-subject-out cross validation on the same group of six subjects evaluated in the personal models.

XSL•FO
RenderX

Figure 1. Efficacy of stroke AR was compared using three types of models.



Barometer Sensor Validation

The barometer is a relatively new feature in smartphone technology, and it is important to determine whether this sensor's signals provide useful information to AR algorithms. Much of AR to date has relied predominantly on movement signals from accelerometers and, more recently, gyroscopes to distinguish between daily mobility-related activities. We thus examined the Healthy-to-Healthy model with and without the 8 barometer features to assess whether this sensor contributes to overall AR performance.

We found that the barometer improved recognition of stairs activity, similar to the findings of Del Rosario et al [14]. Specifically, including the 8 barometer features reduced misclassification of Stairs Up as Walking (from a misclassification of 31.32% [1296/4138] to 13.12% [543/4238]) and reduced misclassification of Stairs Down as Walking (from a misclassification of 33.04% [1230/3723] to 14.07% [524/3723]). Including these features also decreased misclassification of Stairs Down as Stairs Up (from 17.67% [731/4138] to 9.59% [397/4138]). Any changes to Sitting, Lying, and Standing classification with and without the barometer were negligible. As the barometer proved beneficial for stair classification, we included barometer data in our main analyses.

Subject Sample Size for Training Data

We evaluated the effect of subject sample size on AR performance to determine the dependence of classifier accuracy on the number of subjects in the training set. For each population model and the Healthy-trained gait impairment model, we varied the number of subjects used in training from two to fourteen, selecting 1200 instances at random from the available training subjects. We chose 1200 instances because this was the maximum number available between the two subjects with the least amount of data. The number of instances was kept constant, though more subjects were added to the training set to determine the effect of intra-subject variability rather than simply the addition of more data. Each instance is a 10 second clip with 90% overlap between clips, so this corresponds to about 1200 seconds (20 minutes) of data. The models were evaluated on

XSL•FO
RenderX

the remaining test subjects to determine the mean recall. This process was repeated 1000 times for each subject sample size to provide the mean and 95% CIs.

Statistical Analysis

The primary measure of model performance was mean recall. Recall refers to the percentage of correct classifications for a single activity out of all instances of that activity. Recall is the multiclass version of sensitivity, which is only defined for problems with two classes. For the population and environmental models, mean recall was computed for each model as recall averaged across activity classes. Performance of the gait impairment models was additionally evaluated based on the misclassification between stationary (Sitting, Lying, and Standing) and ambulatory (Stairs Up, Stairs Down, and Walking) activities.

Paired *t* tests were used to compare mean recall between the Healthy-to-Stroke and the Stroke-to-Stroke population models, as well as to compare mean recall between the Lab 1-to-Home and the Lab 1-to-Lab 2 environmental models. A two-sample *t* test was used to compare mean recall between the Healthy-to-Healthy and the Stroke-to-Stroke population models. For the gait impairment models, analysis of variance (ANOVA) was used to examine variations in stationary and ambulatory recall among the three impairment groups, using the Tukey honest significant difference test for multiple comparisons. Pearson correlation coefficients were also computed to determine the association between model performance and gait impairment, including mean recall of ambulatory activities and misclassification between stationary and ambulatory activities. Statistical significance levels were set to alpha=0.05, and values are presented as mean (SD).

Results

Distribution of Activity Classes

The amount of data available for training and testing was affected by incidences of transmission-based data drop (poor 4G/LTE signal, leading to transmission backlog and data loss), noncompliance (subject did not label an expected activity), and mislabeling (incorrect activity selected). The percentages of activity labels affected are provided in Table 3. In conjunction with differences in the amount of time spent labeling, this produced class distributions that varied notably between subject groups (Figure 2). The Healthy population generated more than three times the number of instances than the Stroke population (246,283 vs 71,861). Stationary activities accounted for 82.88% (204,123/246,283) of the Healthy data and 53.99% (38,801/71,861) of the Stroke data. Walking was the most prevalent activity for Stroke subjects, accounting for 35.12% (25,234/71,861) of the instances for that population.

Of the six subjects included in the environmental models, three had mild gait impairment, one had moderate gait impairment, and two had severe gait impairment. The distribution of classes for each of these subjects is given in Figure 2.

Table 3. Data loss: average and 95% CIs for percentage of activity labels lost to transmission drop, noncompliance, and mislabeling for each population and environment.

Loss type		Healthy	Stroke
Transmission drop			
	Home	50.6 (40.8-60.4)	44.9 (28.9-61.4)
	Lab1	N/A[a]	11.9 (0.8-23.0)
	Lab 2	N/A	31.0 (14.2-47.9)
Noncompliance			
	Home	15.2 (4.9-25.4)	23.6 (13.3-33.9)
	Lab1	N/A	3.7 (0.4-7.1)
	Lab 2	N/A	1.4 (0-3.0)
Mislabeling			
	Home	2.5 (1.0-3.9)	12.4 (1.2-23.0)
	Lab1	N/A	12.1 (0.8-23.3)
	Lab 2	N/A	23.3 (8.9-37.7)

[a]N/A: not available.

XSL•FO
RenderX

Figure 2. (A) Prevalence of activity classes within the healthy population (gray) and within the stroke population (orange). (B) Prevalence of classes for each of the six Stroke subjects included in the personal environmental models. The total number of instances for each population or subject is shown in italics.

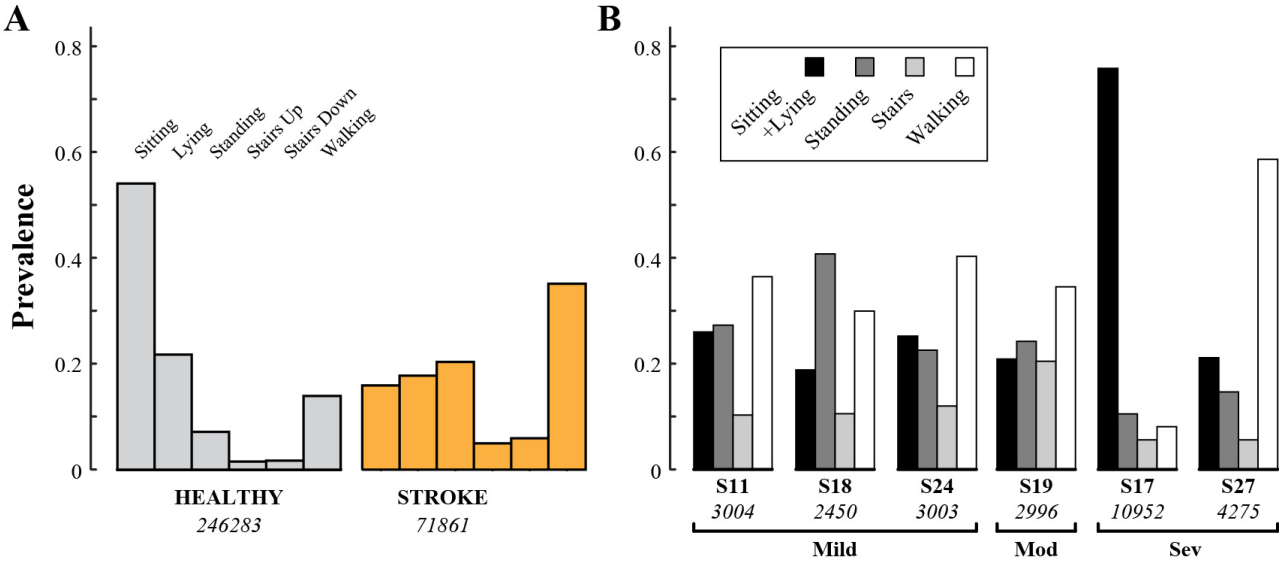
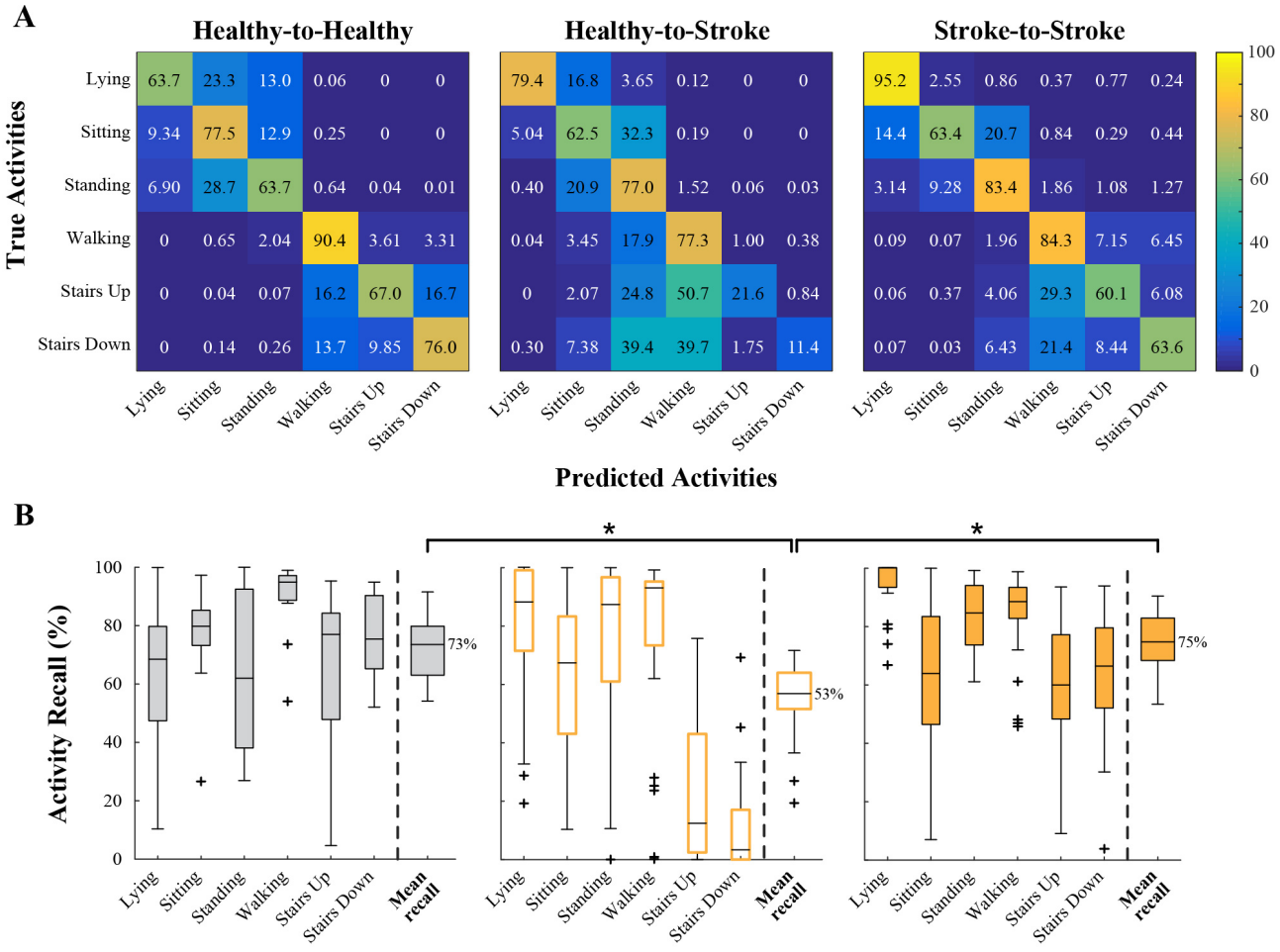


Figure 3. (A) Confusion matrices for each population model, showing average activity recall across test subjects. (B) Boxplots of activity recall for each population model.



Population Models

As expected, the Stroke-to-Stroke model had higher recall across activities than the Healthy-to-Stroke model (Figure 3), showing particular improvement when classifying the Stairs activities.

The mean recalls of the Healthy-to-Healthy, Healthy-to-Stroke, and Stroke-to-Stroke models were 73% (SD 11), 53% (SD 13), and 75% (SD 10), respectively (Figure 3). The Healthy-to-Stroke model performed significantly worse than the Healthy-to-Healthy model (two-sample t test, P<.001) and the

XSL•FO
RenderX

Stroke-to-Stroke model (paired *t* test, *P*<.001). There was no significant difference between the Healthy-to-Healthy and the Stroke-to-Stroke models (two-sample *t* test, *P*=.52).

Average misclassification of ambulatory activities as stationary, and vice versa, was low for the Healthy-to-Healthy (<1.1%) and the Stroke-to-Stroke models (<4.4%). For the Healthy-to-Stroke model, average misclassification of stationary activities as ambulatory was also low (0.77%, 297/38,801), but much more pronounced for ambulatory activities mistaken as stationary ones (30.88%, 10,210/33,060). That is, using a Healthy training set to test on post-stroke activities underestimated ambulation and overestimated less mobile activities such as sitting and standing.

Gait Impairment Models

Confusion matrices for the three gait impairment models using a Healthy training set are shown in Figure 4. For stationary activities, the mean recalls of the Healthy-to-Mild, Healthy-to-Moderate, and Healthy-to-Severe models were 59% (SD 7), 74% (SD 16), and 74% (SD 13), respectively. An ANOVA on the mean stationary recall for each stroke subject yielded no significant variation between the three models ($F_{2,26}$=1.83, *P*=.18). For ambulatory activities, the respective mean recalls of the three models were 56% (SD 13), 35% (SD 19), and 23% (SD 20). An ANOVA on the mean ambulatory recall for each subject revealed significant variation between the three models ($F_{2,26}$=6.96, *P*=.004).

A post-hoc Tukey test confirmed that the Healthy-to-Mild and Healthy-to-Severe mean recall was significantly different (*P*=.003); the Healthy-to-Moderate model did not have significantly different mean recall from Healthy-to-Mild (*P*=.11) or Healthy-to-Severe (*P*=.12).

In fact, across all stroke participants in this study, there was a significant, moderate correlation between mean recall in the ambulatory activities and walking speed in the 10MWT (Figure 4; *r*=.641, *P*<.001). That is, the performance of the Healthy training set to classify stairs and walking activity declined as gait impairment increased. On the other hand, there was no association between mean recall in stationary activities and walking speed (*r*=−.262, *P*=.17).

The effect of gait impairment appeared most pronounced in the classification of ambulatory activities. Average misclassification of stationary activities decreased slightly with gait impairment, from 1.07% (97/9075) in Mild to 0.12% (19/16,077) in Severe. Average misclassification of ambulatory activities as stationary activities increased substantially with gait impairment, from 10.21% (818/8014) in Mild to 30.02% (3354/11,172) in Moderate, and to 51.01% (7077/13,874) in the Severe subjects. The Healthy-to-Mild model more accurately distinguished between stationary and ambulatory activities, with mean recalls of 99% (SD 1) and 90% (SD 5), respectively.

There was a moderate negative correlation between misclassification of ambulatory activities and walking speed in the 10MWT (Figure 4; *r*=−.634, *P*<.001). There was no association between misclassification of stationary activities and walking speed (*r*=−.307, *P*=.11). Thus, gait impairment hindered the classification of ambulatory activities, without impacting stationary activities.

The mixed-impairment training data from the Stroke-to-Stroke model recognized activities more accurately than the Healthy-to-Stroke model, across impairment levels. The mean recalls of the Stroke-to-Mild, Stroke-to-Moderate, and Stroke-to-Severe models were 72% (SD 16), 78% (SD 11), and 74% (SD 16), respectively (Figure 4). An ANOVA on the mean recall for each subject yielded no significant variation between the three models ($F_{2,26}$=1.33, *P*=.28).

Average misclassification of stationary activities was similar across gait impairment groups for the Stroke-to-Stroke model, with 2.7% of instances misclassified in Mild, 3.7% in Moderate, and 1.8% in Severe (Figure 4). Average misclassification of ambulatory activities as stationary activities occurred more frequently for persons with severe gait impairment, with 6.9% of instances misclassified in Severe versus 2.4% in Mild and 1.3% in Moderate. Most of these errors in the Severe subjects resulted from confusing stairs activity with standing, presumably due to the slower speeds and longer pauses on the stairs.

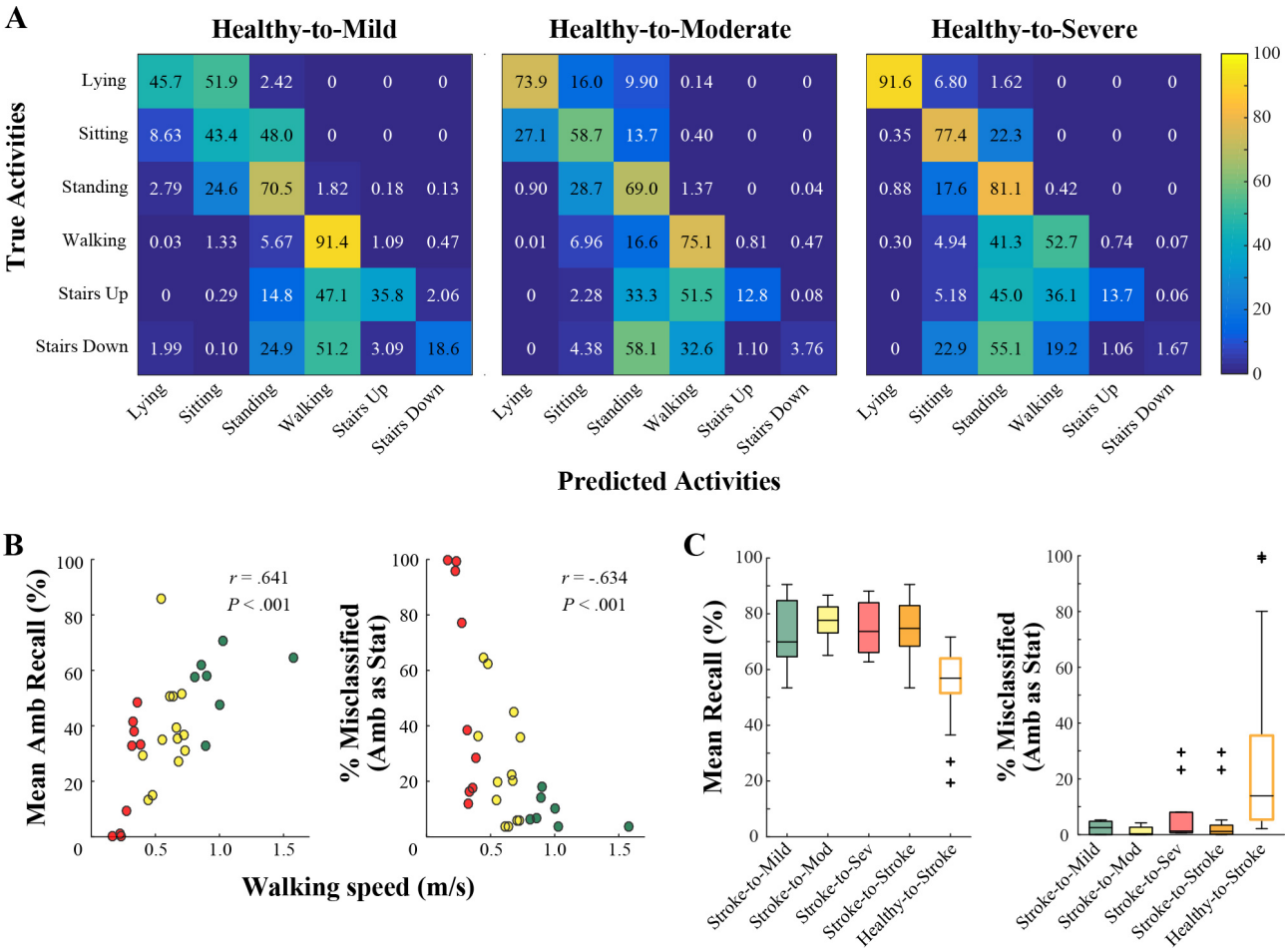
Environmental Models

The mean recall values of the Lab 1-to-Lab 2, Lab 1-to-Home, and Home-to-Home models were 72% (SD 18), 52% (SD 12), and 67% (SD 7), respectively (Figure 5). The Lab 1-to-Home model performed significantly worse than then Lab 1-to-Lab 2 model (*P*=.024) and the Home-to-Home model (*P*<.001). There was no significant difference between the Lab 1-to-Lab 2 and the Home-to-Home models (*P*=.43). In the Lab 1-to-Home model, average misclassification of stationary activities as ambulatory occurred in 17.86% (2,075/11,618) of instances, compared with 0.54% (7/1,296) in the Lab 1-to-Lab 2 model and 1.58% (179/11,316) in the Home-to-Home model. More drastically, the Lab 1-to-Home model misclassified ambulatory activities as stationary on an average of 40.34% (2,273/5,635), compared with 6.57% (68/1,035) for the Lab 1-to-Lab 2 model and 4.10% (218/5,320) for the Home-to-Home model.

To compare the effects of environment in a practical AR implementation, we also examined global models for Lab 1-to-Lab 2, Lab 1-to-Home, and Home-to-Home. The mean recall values of these models were 80% (SD 11), 65% (SD 14), and 61% (SD 9), respectively. There was no significant difference between the Lab 1-to-Lab 2 and the Lab 1-to-Home global models (*P*=.07), nor between the Lab 1-to-Home and the Home-to-Home global models (*P*=.53). The Home-to-Home global model performed significantly worse than the Lab 1-to-Lab 2 global model (*P*=.004). In the Lab 1-to-Home global model, average misclassification of stationary activities as ambulatory occurred in 0.84% (98/11,618) of instances, compared with 0.54% (7/1,296) in the Lab 1-to-Lab 2 global model and 1.34% (156/11,618) in the Home-to-Home global model. More drastically, the Lab 1-to-Home model misclassified ambulatory activities as stationary on an average of 12.49% (704/5,635) of the time, compared with 3.57% (37/1,035) for the Lab 1-to-Lab 2 model and 3.07% (173/5,635) for the Home-to-Home model.

XSL•FO
RenderX

Figure 4. (A) Confusion matrices for each gait impairment in the Healthy-to-Stroke model. (B) For a Healthy training set, walking speed in the 10MWT is positively correlated with mean recall across ambulatory activities (left) and negatively correlated with misclassification of ambulatory activities as stationary (right) when using a Healthy training set. (C) For a Stroke training set, mean recall and misclassifications of ambulatory activities are similar between gait impairment groups.



In summary, personal models exhibited a marked improvement in mean recall when using a Home training set rather than a Lab training set to classify Home activities. Global models that were run on six subjects did not show such improvement. Both types of AR models showed similar trends in reduced misclassification of Home ambulatory activities when using a Home rather than Lab training set.

Sample Size of Training Subjects Achieves Near-Optimal Performance

Mean recall increased with training set sample size in our population models (Figure 6). The Healthy-to-Healthy model performance plateaued after about 10 training subjects. The Healthy-to-Stroke performance plateaued at about 12 subjects, regardless of the impairment severity group used for testing (Figure 6). The Stroke-to-Stroke model began to plateau at 14 subjects. Therefore, we recommend using a training pool of at least 14 subjects for optimal AR performance for persons with stroke. More subjects may be necessary for AR involving more activity classes than the six used in this study.

This suggests that our models, using 14 healthy and 29 stroke subjects in training, achieved near-optimal performance. Note that the analyses presented in the main text included all available instances for each subject. Including more training instances would produce higher mean recall than those given for our population models. For a larger number of instances, fewer subjects may be necessary to achieve the point of marginal performance gains.

Healthy training data yielded only marginal improvements using more than 10-12 subjects. Stroke training data yielded continued improvements using up to 14 subjects.

XSL•FO
RenderX

Figure 5. (A) Confusion matrices and (B) boxplots of activity recall for the personal environmental models.

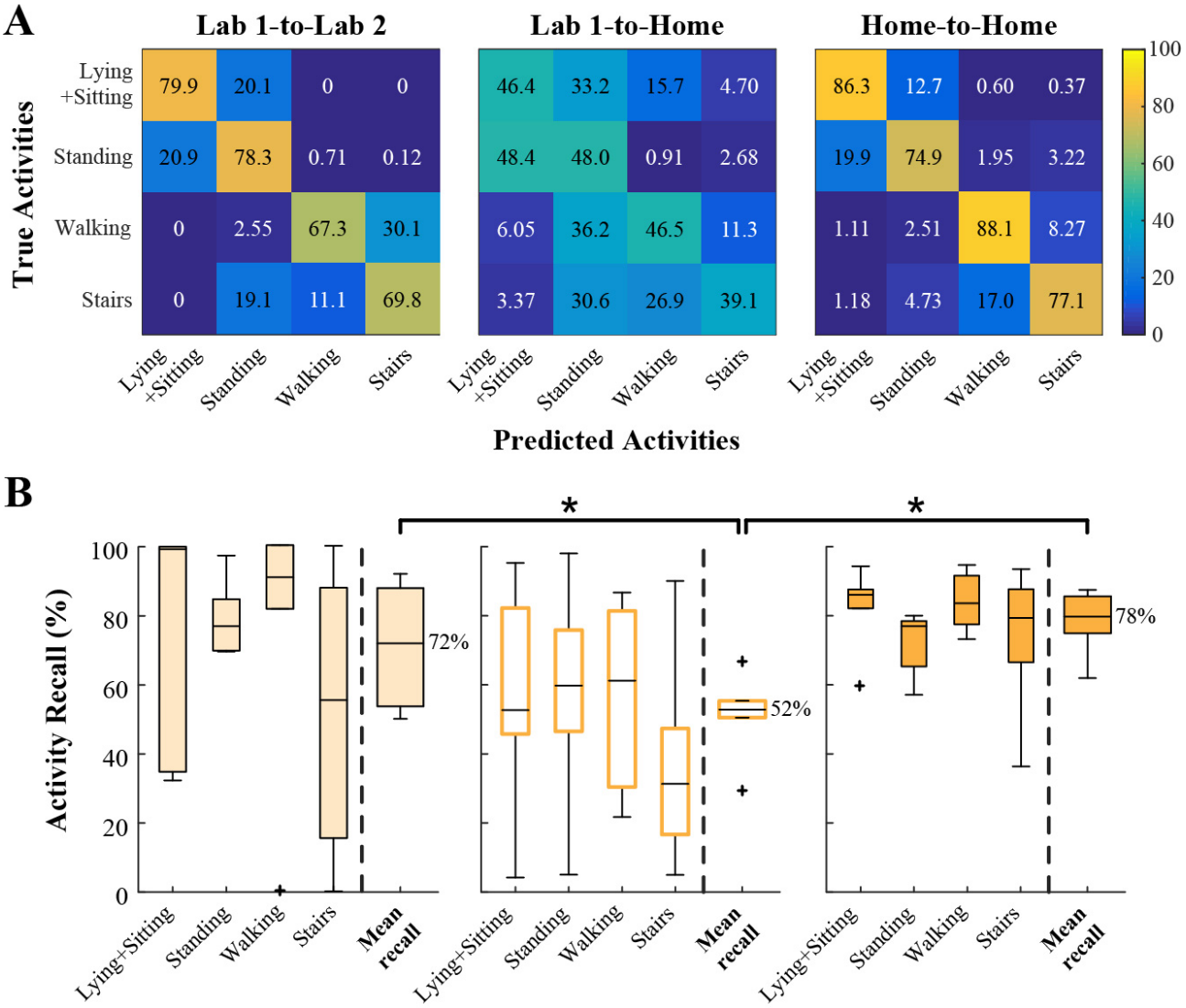
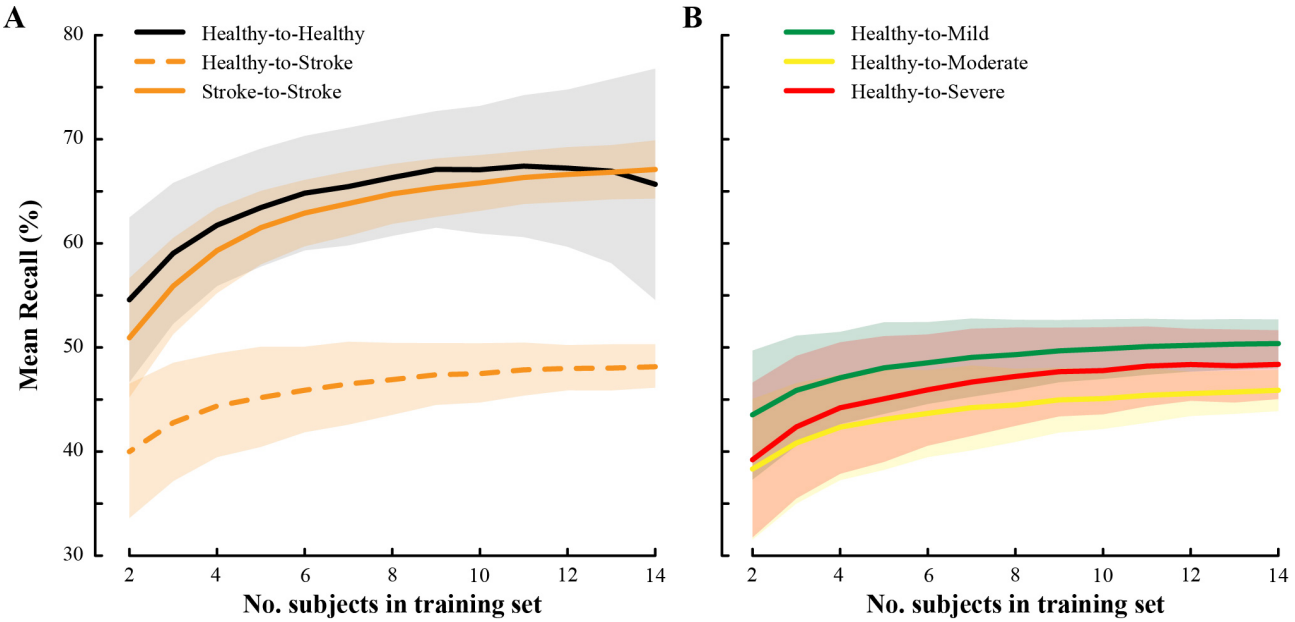


Figure 6. Average and SD of mean recall for (A) population models and (B) gait impairment models using different subject sample sizes in training data (1200 random instances pulled, repeated 1000 times).



XSL•FO
RenderX

Discussion

Principal Findings

We have shown that training AR models on a mixed-impairment stroke population improves activity recognition for persons with stroke compared with training on a healthy population, increasing mean recall from 53% to 75%. Models trained on either the healthy population or patients with mild gait impairment performed poorly when classifying ambulatory activities in patients with severe gait impairment, lowering mean recall to 23% by increasingly misclassifying ambulatory activities as stationary ones.

Finally, personal models trained on in-lab activities performed poorly when tested on at-home activities, with 56% recall. Global models, which account for inter-subject variability, may be less susceptible to variability in activities between environmental contexts. However, the effect of environment on global models remains to be demonstrated with a sufficient number of subjects in the training set; using only six subjects in the environmental global model analysis is likely insufficient to achieve optimal model performance. Taken together, our results suggest that future community monitoring systems for persons with stroke should incorporate activity training data collected outside the laboratory using cohorts with similar gait impairment or a wide range of gait impairments.

Comparison With Prior Work

Our results for the stroke population align with previous work validating AR reliability for persons with neurological injury when using training data from a young, healthy cohort [14-16,18]. We have extended this analysis to stroke by comparing AR across levels of gait impairment, finding that the healthy-trained model increasingly underestimated ambulation with impairment severity. This is especially problematic as it affects a population that already exhibits reduced ambulation. Conversely, misclassification was reduced across impairment groups using a model trained on a mixed-impairment stroke cohort (Figure 4). Our results indicate that training sets incorporating a broad range of gait impairments may be generally sufficient to classify the activities of a stroke subject.

Our study also offers several contributions to community monitoring research for persons with stroke. We have presented an activity recognition system using smartphone technology that allows users to independently label activity data outside the laboratory. Using this system, we have evaluated the validity of various activity recognition models when classifying activities in persons with stroke. Our findings agree with a recent study by Albert et al [25], in which support vector machine models trained on home activity data outperformed models trained on lab data when classifying mobility-related activities for patients with incomplete spinal cord injury. Our results further demonstrate the need to train AR models using data representative of home activities as best as possible for the population of interest. We expect this would also hold for activity data collected from other movement sensors, such as inertial measurement units (IMUs).

Limitations

While this work is an important first step in highlighting potential issues in AR for community-based monitoring, our findings should be considered in the context of several limitations. For one, the healthy cohort was not age-matched to the stroke cohort, which may underestimate the performance of the Healthy-to-Stroke population model. Nevertheless, patients with stroke have reduced walking speeds relative to age-matched controls [26], and we do not expect that age-matching would negate our overall findings in the population models. Indeed, age and neurological injury both appear to negatively impact AR performance when training with activities from a young, healthy population.

The positioning of the mobile phone may present another limitation. Other studies using mobile phone data have placed the phone in the pocket and achieved better resolution of stationary activities than the models presented here. For example, Kwapisz et al [27] achieved high sensitivities for both sitting and standing (>93%) with a mobile phone placed in the pocket. Sensor data collected from the waist may not be adequate to distinguish between sitting and standing postures. We decided to keep the phone in a belted pouch at the waist to minimize discomfort, avoid the need for large-pocketed clothing, and maintain consistent phone access for all subjects during labeling.

A third limitation of our study was the reduced amount of home data available for AR validation in the stroke population, resulting from data loss and protocol design. Most of the data loss resulted from transmission-based data drops, which happened in cases of poor 4G/LTE signal. Specifically, continuously poor signal would lead to a data backlog and prevent new data from being captured. This was a problem with the CIMON app that has since been improved for future studies. Furthermore, because stroke subjects only completed one session of at-home labeling, we expect the Home-to-Home personal models to have slightly overestimated performance, though this is likely negligible relative to the poor performance of the Lab-to-Home models. Follow-up studies should collect activity data on different days and in more varied community environments to better capture variance in behavior.

Future work should consider the amount of training data needed for satisfactory AR performance in a stroke cohort. Model accuracy changes with the sample size of healthy and stroke subjects included in the training set. We found that about 10 healthy subjects and 14 stroke subjects achieved satisfactory AR accuracy, using 20 minutes of total activity data for model training (Figure 6). However, the amount of training data is not the sole determinant of model quality. As we have demonstrated in this study, consideration of target population and environment are crucial to maximize AR performance.

Conclusions

Improving the reliability of AR algorithms for persons with stroke has significant benefits for home and community monitoring. Wearable technology paired with AR will allow clinicians to construct and supervise remote/home rehabilitation programs, utilizing data-driven feedback about patient activities.

XSL•FO
RenderX

Our recommendation for future stroke-based AR models is to use training data from a balanced distribution of gait impairment levels, thereby including as much variety as possible to improve performance across all levels. Future studies should further examine real-world activity labeling to improve the ability of AR models to generalize across multiple environmental contexts, not just in the laboratory and at home. These findings may help guide the construction of future AR models for persons with stroke and other clinical populations.

Acknowledgments

This work was supported by the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR 90RE5014-02-00).

Authors' Contributions

MKO and NS designed and performed data and statistical analyses and wrote and edited the manuscript. CKM developed materials and methods, collected data, and edited the manuscript. SK developed materials and methods and collected data. XB developed materials and methods (CIMON). CP developed materials and methods (CIMON) and edited the manuscript. KK designed and reviewed the data analysis and edited the manuscript. AJ conceived and designed the experiments, developed materials and methods, designed and reviewed the data and statistical analyses, and wrote and edited the manuscript.

Conflicts of Interest

None declared.

References

1. Lord SE, McPherson K, McNaughton HK, Rochester L, Weatherall M. Community ambulation after stroke: how important and obtainable is it and what measures appear predictive? Arch Phys Med Rehabil 2004 Feb;85(2):234-239. [Medline: 14966707]
2. Dobkin BH, Dorsch A. New evidence for therapies in stroke rehabilitation. Curr Atheroscler Rep 2013 Jun;15(6):331 [FREE Full text] [doi: 10.1007/s11883-013-0331-y] [Medline: 23591673]
3. Michael KM, Allen JK, Macko RF. Reduced ambulatory activity after stroke: the role of balance, gait, and cardiovascular fitness. Arch Phys Med Rehabil 2005 Aug;86(8):1552-1556. [doi: 10.1016/j.apmr.2004.12.026] [Medline: 16084807]
4. Tieges Z, Mead G, Allerhand M, Duncan F, van WF, Fitzsimons C, et al. Sedentary behavior in the first year after stroke: a longitudinal cohort study with objective measures. Arch Phys Med Rehabil 2015 Jan;96(1):15-23. [doi: 10.1016/j.apmr.2014.08.015] [Medline: 25220942]
5. Koster A, Caserotti P, Patel KV, Matthews CE, Berrigan D, Van Domelen DR, et al. Association of sedentary time with mortality independent of moderate to vigorous physical activity. PLoS One 2012;7(6):e37696 [FREE Full text] [doi: 10.1371/journal.pone.0037696] [Medline: 22719846]
6. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N. Too little exercise and too much sitting: inactivity physiology and the need for new recommendations on sedentary behavior. Curr Cardiovasc Risk Rep 2008 Jul;2(4):292-298 [FREE Full text] [doi: 10.1007/s12170-008-0054-8] [Medline: 22905272]
7. Tremblay MS, Colley RC, Saunders TJ, Healy GN, Owen N. Physiological and health implications of a sedentary lifestyle. Appl Physiol Nutr Metab 2010 Dec;35(6):725-740. [doi: 10.1139/H10-079] [Medline: 21164543]
8. Dobkin BH, Carmichael ST. The specific requirements of neural repair trials for stroke. Neurorehabil Neural Repair 2016 Jun;30(5):470-478. [doi: 10.1177/1545968315604400] [Medline: 26359342]
9. Dobkin BH, Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. Neurorehabil Neural Repair 2011;25(9):788-798 [FREE Full text] [doi: 10.1177/1545968311425908] [Medline: 21989632]
10. Evans CC, Hanke TA, Zielke D, Keller S, Ruroede K. Monitoring community mobility with global positioning system technology after a stroke: a case study. J Neurol Phys Ther 2012 Jun;36(2):68-78. [doi: 10.1097/NPT.0b013e318256511a] [Medline: 22592062]
11. Jayaraman A, Deeny S, Eisenberg Y, Mathur G, Kuiken T. Global position sensing and step activity as outcome measures of community mobility and social interaction for an individual with a transfemoral amputation due to dysvascular disease. Phys Ther 2014 Mar;94(3):401-410 [FREE Full text] [doi: 10.2522/ptj.20120527] [Medline: 24092905]
12. Mahendran N, Kuys SS, Brauer SG. Accelerometer and global positioning system measurement of recovery of community ambulation across the first 6 months after stroke: an exploratory prospective study. Arch Phys Med Rehabil 2016 Sep;97(9):1465-1472. [doi: 10.1016/j.apmr.2016.04.013] [Medline: 27181183]
13. Fini NA, Holland AE, Keating J, Simek J, Bernhardt J. How is physical activity monitored in people following stroke? Disabil Rehabil 2015;37(19):1717-1731. [doi: 10.3109/09638288.2014.978508] [Medline: 25374044]
14. Del Rosario MB, Wang K, Wang J, Liu Y, Brodie M, Delbaere K. A comparison of activity classification in younger and older cohorts using a smartphone. Physiol Meas 2014;35:2269-2286. [doi: 10.1371/journal.pone.0124414]

XSL•FO
RenderX

15. Albert MV, Toledo S, Shapiro M, Kording K. Using mobile phones for activity recognition in Parkinson's patients. Front Neurol 2012;3:158 [FREE Full text] [doi: 10.3389/fneur.2012.00158] [Medline: 23162528]
16. Lonini L, Gupta A, Kording K, Jayaraman A. Activity recognition in patients with lower limb impairments: do we need training data from each patient? Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:3265-3268. [doi: 10.1109/EMBC.2016.7591425] [Medline: 28269004]
17. Mizuike C, Ohgi S, Morita S. Analysis of stroke patient walking dynamics using a tri-axial accelerometer. Gait Posture 2009 Jul;30(1):60-64. [doi: 10.1016/j.gaitpost.2009.02.017] [Medline: 19349181]
18. Capela NA, Lemaire ED, Baddour N, Rudolf M, Goljar N, Burger H. Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. J Neuroeng Rehabil 2016 Jan 20;13:5 [FREE Full text] [doi: 10.1186/s12984-016-0114-0] [Medline: 26792670]
19. Folstein M, Folstein S, McHugh P. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]
20. Miller C, Poellabauer C. Configurable Integrated Monitoring System for Mobile Devices. In: Procedia Computer Science. 2014 Apr Presented at: Proc of the 11th International Conference on Mobile Systems and Pervasive Computing; 2014; Niagara Falls, Canada p. 410-417 URL: http://dx.doi.org/10.1002/iub.1366 [doi: 10.1016/j.procs.2014.07.046]
21. Hossain A, Poellabauer C. Challenges in building continuous smartphone sensing applications. In: Proc of the 9th IEEE International Workshop on Selected Topics in WirelessMobile Computing. 2016 Presented at: IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications; October 17-19, 2016; New York, NY.
22. Ellis K, Godbole S, Marshall S, Lanckriet G, Staudenmayer J, Kerr J. Identifying active travel behaviors in challenging environments using GPS, accelerometers, and machine learning algorithms. Front Public Health 2014 Apr;2:36 [FREE Full text] [doi: 10.3389/fpubh.2014.00036] [Medline: 24795875]
23. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. J Sci Med Sport 2017 Jan;20(1):75-80. [doi: 10.1016/j.jsams.2016.06.003] [Medline: 27372275]
24. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: a hybrid approach to alleviating class imbalance. IEEE Trans Syst, Man, Cybern A 2010 Jan;40(1):185-197. [doi: 10.1109/TSMCA.2009.2029559]
25. Albert MV, Azeze Y, Courtois M, Jayaraman A. In-lab versus at-home activity recognition in ambulatory subjects with incomplete spinal cord injury. J Neuroeng Rehabil 2017 Feb 06;14(1):10 [FREE Full text] [doi: 10.1186/s12984-017-0222-5] [Medline: 28166824]
26. von Schroeder HP, Coutts RD, Lyden PD, Billings E, Nickel VL. Gait parameters following stroke: a practical assessment. J Rehabil Res Dev 1995 Feb;32(1):25-31. [Medline: 7760264]
27. Kwapisz J, Weiss G, Moore SA. Activity recognition using cell phone accelerometers. SIGKDD Explor Newsl 2011;12:74-82.

Abbreviations

10MWT: 10-meter walk test
AR: activity recognition
HIPAA: Health Insurance Portability and Accountability Act of 1996
RF: random forest
RUS: random undersampling
SD: standard deviation

Edited by G Eysenbach; submitted 23.01.17; peer-reviewed by B Dobkin, M Albert; comments to author 16.03.17; revised version received 03.04.17; accepted 07.04.17; published 25.05.17.

Please cite as:
O'Brien MK, Shawen N, Mummidisetty CK, Kaur S, Bo X, Poellabauer C, Kording K, Jayaraman A
Activity Recognition for Persons With Stroke Using Mobile Phone Technology: Toward Improved Performance in a Home Setting
J Med Internet Res 2017;19(5):e184
URL: http://www.jmir.org/2017/5/e184/
doi:10.2196/jmir.7385
PMID:28546137

©Megan K O'Brien, Nicholas Shawen, Chaithanya K Mummidisetty, Saninder Kaur, Xiao Bo, Christian Poellabauer, Konrad Kording, Arun Jayaraman. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 25.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic

XSL•FO
RenderX

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

XSL•FO
RenderX

Original Paper

Harnessing Facebook for Smoking Reduction and Cessation Interventions: Facebook User Engagement and Social Support Predict Smoking Reduction

Sunny Jung Kim[1], PhD; Lisa A Marsch[1], PhD; Mary F Brunette[2], MD; Jesse Dallery[3], PhD

[1]Center for Technology and Behavioral Health, Department of Biomedical Data Science, Department of Psychiatry, Geisel School of Medicine at Dartmouth, Dartmouth College, Lebanon, NH, United States

[2]Center for Technology and Behavioral Health, Department of Psychiatry, Geisel School of Medicine at Dartmouth, Dartmouth-Hitchcock, Lebanon, NH, United States

[3]Department of Psychology, University of Florida, Gainesville, FL, United States

Corresponding Author:
Sunny Jung Kim, PhD
Center for Technology and Behavioral Health
Department of Biomedical Data Science, Department of Psychiatry
Geisel School of Medicine at Dartmouth, Dartmouth College
46 Centerra Parkway Suite 301
Lebanon, NH, 03766
United States
Phone: 1 603 646 7041
Fax: 1 603 646 7000
Email: sunny.j.kim@dartmouth.edu

Abstract

Background: Social media technologies offer a novel opportunity for scalable health interventions that can facilitate user engagement and social support, which in turn may reinforce positive processes for behavior change.

Objective: By using principles from health communication and social support literature, we implemented a Facebook group–based intervention that targeted smoking reduction and cessation. This study hypothesized that participants' engagement with and perceived social support from our Facebook group intervention would predict smoking reduction.

Methods: We recruited 16 regular smokers who live in the United States and who were motivated in quitting smoking at screening. We promoted message exposure as well as engagement and social support systems throughout the intervention. For message exposure, we posted prevalidated, antismoking messages (such as national antismoking campaigns) on our smoking reduction and cessation Facebook group. For engagement and social support systems, we delivered a high degree of engagement and social support systems during the second and third week of the intervention and a low degree of engagement and social support systems during the first and fourth week. A total of six surveys were conducted via Amazon Mechanical Turk (MTurk) at baseline on a weekly basis and at a 2-week follow-up.

Results: Of the total 16 participants, most were female (n=13, 81%), white (n=15, 94%), and between 25 and 50 years of age (mean 34.75, SD 8.15). There was no study attrition throughout the 6-time-point baseline, weekly, and follow-up surveys. We generated Facebook engagement and social support composite scores (mean 19.19, SD 24.35) by combining the number of likes each participant received and the number of comments or wall posts each participant posted on our smoking reduction and cessation Facebook group during the intervention period. The primary outcome was smoking reduction in the past 7 days measured at baseline and at the two-week follow-up. Compared with the baseline, participants reported smoking an average of 60.56 fewer cigarettes per week (SD 38.83) at the follow-up, and 4 participants out of 16 (25%) reported 7-day point prevalence smoking abstinence at the follow-up. Adjusted linear regression models revealed that a one-unit increase in the Facebook engagement and social support composite scores predicted a 0.56-unit decrease in cigarettes smoked per week (standard error =.24, P=.04, 95% CI 0.024-1.09) when baseline readiness to quit, gender, and baseline smoking status were controlled ($F_{4, 11}$=8.85, P=.002).

Conclusions: This study is the first Facebook group–based intervention that systemically implemented health communication strategies and engagement and social support systems to promote smoking reduction and cessation. Our findings imply that

XSL•FO
RenderX

receiving one like or posting on the Facebook-based intervention platform predicted smoking approximately one less cigarette in the past 7 days, and that interventions should facilitate user interactions to foster user engagement and social support.

(J Med Internet Res 2017;19(5):e168) doi:10.2196/jmir.6681

KEYWORDS

social media; social support; behavior and behavior mechanisms; smoking cessation; persuasive communication; social networking; technology; health promotion

Introduction

Background

Tobacco use is the primary cause of premature mortality and is responsible for almost half a million deaths every year in the United States and nearly 5 million deaths globally [1]. Although mass media–based (eg, TV, radio) health campaigns have been moderately effective for health promotion—influencing 4% to 8% of the population to change health behaviors [2]—several considerable drawbacks remain. First, mass media–based approaches for public health promotion, such as tobacco control and prevention campaigns, fall short in enabling a frequent and durable delivery system that can disseminate individually tailored messages. Second, a message exposure tracking system is often nonexistent or costly in traditional mass media settings. Thus, it is difficult to directly gauge the effects of mass media exposure to health messages on health behavior. Third, traditional media platforms offer a limited to nonexistent interactive engagement system that could enhance social support and user-centered engagement.

Significant advances in social media technologies and their ubiquity offer novel opportunities to provide geographically distant users with easily accessible, cost-effective, personalized health content, and social network-based support. For example, Facebook, one of the most widely adopted social media platforms, hosts approximately 1.22 billion daily active users [3] and has an enormous quantity of user-initiated virtual communities that are highly relevant to people seeking social support for health problems [4]. The magnitude of this social network platform and its popularity can considerably extend the reach of evidence-based health messages to the public and scale-up user-centered social support to the population level to address public health problems. In fact, a growing volume of research is leveraging social media to facilitate health behavior changes such as increasing physical activities [5], enabling addiction recovery support [6], providing support for cancer survivors [7,8], and reducing sexual risk behaviors among youth [9]. More evidence has shown that participants perceive social media approaches for health promotion as appealing, acceptable, and convenient [10].

In this regard, social media such as Facebook provide a range of communication features for putative processes of behavior change that are important to individuals with health problems. Those social media features and related processes include “posting” features for self-disclosure [11], search functions via hashtags for information-seeking [12], “share” features for social sharing [13], and using “comment” and “reaction” features for engagement and social support [14-16]. Strategic use of social media features that facilitates these processes may foster desirable health outcomes [17].

Facebook groups, in particular, can be used as a designated online social support community for members with similar health concerns [18]. They provide various social interaction features such as “likes” and “comments” for group members. Feedback “comments” from group members and clinicians as well as reciprocated interactivities can create a supportive environment for achieving health promotion [4]. Researchers can also track whether the target audiences view the intervention content and observe social dynamics among group members in adopting health attitudes and behaviors.

In this study, we utilized Facebook group features to effectively disseminate prevalidated antismoking messages with high frequency and longitudinal exposure. Message exposure frequency and exposure duration are pivotal factors for successful health campaigns [19] that are often limited in traditional mass media environments. In addition, we harnessed multidirectional communication processes among intervention target participants (smokers) to foster user-centered engagement and social support.

Despite the potential benefits of harnessing social media for health interventions, a critical gap in knowledge persists in terms of how to best utilize social media features to achieve positive health intervention outcomes. The intent of this study is to strategically leverage communication features that are available on Facebook groups to implement a smoking reduction and cessation intervention among regular smokers who are interested in quitting. For our intervention, we promoted smoking cessation as the optimal outcome to achieve, but we also accepted smoking reduction as a positive change for those who could not immediately quit smoking, as smoking reduction is a common step toward eventual cessation [20]. In doing so, we designed a smoking reduction and cessation intervention through a Facebook group with two primarily theory-guided intervention components: (1) message exposure to antismoking media content in a systematic manner while strategically changing messaging frequency (high vs low), and (2) social support and engagement systems with different levels (high vs low).

Exposure to Antismoking Messages

Prior studies on health promotions and health behavior models have demonstrated that exposure to health communications can enhance one’s health behavior by changing core beliefs and attitudes about expected health outcomes [21], by providing education on the skills needed to change health behavior [22], and by disseminating knowledge about the target health behavior [23]. In realizing these intermediate factors for behavior change, acquiring exposure to intervention messages is the most

XSL•FO
RenderX

important requisite condition for successful health promotion [24]. This study targeted systematic, frequent exposure to antismoking messages as a key factor. We posited that strategic exposure to antismoking messages could influence the target audience to form positive beliefs and behaviors about smoking reduction and cessation [25]. Different from traditional media environments, leveraging social media as a communication platform for a health intervention enables investigators to predesign and administer the frequency and delivery schedule of message exposure in a systematic manner (eg, one-time message exposure at the same time of the day). Social media also enable interventionists to disseminate campaign messages to geographically distant audiences. For our Facebook group–based intervention messages, we aimed to use preexisting, well-received antismoking campaign messages (eg, the Tips From Former Smokers campaign) rather than generating a new set of antismoking messages. In order to reduce the cost, time, and risks associated with developing new campaign content, the CDC's (Centers for Disease Control and Prevention) best practice guidelines for tobacco control and prevention campaigns recommend reuse of existing campaign messages that have shown positive campaign outcomes [1].

Social Support and User Engagement

Social support and engagement systems were additional theoretical components that were applied in our Facebook group–based smoking reduction and cessation intervention. Social support is defined as informational, emotional, reassuring, or tangible resources [26] provided by professional or nonprofessional social capital. Social support has been shown to be one of the most considerable coping resources for achieving desirable health outcomes [27], including behavioral health [28], physical health [5], and mental health [29]. Social media can be an excellent outlet for active user engagement and peer-to-peer health support [30,31]. Social media technologies offer various communication and social networking features that individuals can use to share their health concerns and engage their social networks for support [30]. In the intervention, we implemented health communication strategies to deliver social support and user engagement systems, which in turn may foster smoking reduction and cessation [32]

Based on the four dimensions of the social support conceptual framework [33], we operationalized three types of social support constructs—emotional support, informational support, and reassuring support—through comments and wall postings on our smoking reduction and cessation Facebook group. For emotional support, we showed an empathic understanding of the issues of participants (eg, We know how hard it is to quit smoking, and understand what you are going through). To operationalize informational support, we appraised the participants' circumstances and posted advising comments on topics ranging from nicotine craving symptoms to possible solutions (eg, Here are some tips for managing cravings), and we also provided reassuring support by assuring self-confidence to the participants through affirmative comments (eg, You can do it!).

One's perceived social support can help them enhance their self-efficacy beliefs in order to overcome barriers to adopting the health behavior being promoted. To deliver social support and engagement in relation to promoting smoking reduction and cessation during the 4-week intervention period, we manipulated the level of engagement and social support systems (high vs low) and juxtaposed it with high versus low message exposure.

Specific Objectives

We examined the feasibility of a Facebook group–based smoking reduction and cessation intervention. Additionally, the preliminary efficacy on smoking reduction (the reduced number of cigarettes consumed per week) and on 7-day point smoking abstinence at the follow-up was tested. We also tested whether the intervention components (social support and engagement systems) predict smoking reduction.

Methods

Recruitment

Recruitment messages and preliminary screening questions were disseminated through Amazon Mechanical Turk (MTurk) and social media platforms. MTurk is an anonymous Web-based labor market with over 500,000 registered workers worldwide. MTurk workers complete tasks distributed by requesters for small financial rewards. MTurk has been used as a recruitment pool in various fields of research for an array of tasks, including decision-making [34-36], health literacy [37], and natural language processing tasks [38].

Inclusion Criteria

Over 200 applicants who were interested in our four-week smoking reduction and cessation interventions were screened based on their self-reported characteristics. The inclusion criteria were regular smokers (smoking 5 days per week) who were between the ages of 18 and 65 years and living in the United States. To be eligible, participants had to have no chronic disease interfering with their daily lives, no use of illicit drugs, and be motivated to quit smoking (> 80, on a 100-point motivation to quit smoking scale [39]). To ensure that access to and use of Facebook were not barriers to participate in the intervention, participants had to have Internet access and had to use Facebook through their mobile or computer devices on a regular basis.

Qualified applicants (N=132) were invited to participate in our study. The eligible participants who responded to our invitation (N=46) were randomly assigned to one of the following conditions: email condition, MTurk-only condition, or the Facebook condition. Participants were introduced to their intervention and coached on how to participate in it. This report focuses on the subjects who were randomized to the Facebook condition (n=16) in order to give special attention to the findings that are unique due to the social media features exclusively available on Facebook (eg, comments, share, likes, and wall postings). Primary outcomes from all three conditions will be published in a separate report.

Intervention Guidelines

Participants were first provided with an electronic informed consent form. Before the start date of the intervention, researchers contacted consenting individuals through an

XSL•FO
RenderX

individual phone call meeting and provided guidelines in greater detail on how to participate in the Facebook intervention. For example, participants were encouraged to share their thoughts, progress, and peer support. We also provided practical methods on engaging with the intervention materials on a daily basis by leaving comments, liking posts, and interacting with other peers in the group throughout the four-week intervention period. The participants were informed that there is no incentive for intervention engagement in our smoking reduction and cessation Facebook group. We informed participants that our research members would post different antismoking messages throughout the intervention period and provide social support to keep participants motivated to engage in action for smoking reduction and cessation. In addition, on the start date, the research team greeted all participants on the Facebook group wall by posting encouraging statements such as "…If you are having a hard time quitting, let us hear. We are here to support you and encourage you to achieve your goal." This greeting statement was used to set a positive tone and an atmosphere inclusive of all participants. All procedures, materials, and study protocols were reviewed and approved by the university's Institutional Review Board.

Research Design of the Smoking Reduction and Cessation Facebook Group

We implemented and targeted different levels of message exposure and engagement and social support systems over four weeks. During week 1 (high message exposure combined with low engagement and social support), we posted antismoking messages three times per day without directly encouraging people to respond to the materials or share their thoughts. During week 2 (low message exposure combined with high engagement and social support), we posted antismoking materials once per day and delivered supportive comments and fostered user engagement by directly asking participants to share their motivating factors, their thoughts on posted antismoking messages, and their progress on quitting smoking with the group. During week 3 (high message exposure combined with high engagement and social support), we posted antismoking messages three times per day. In addition, a professional clinical expert joined the Facebook group and provided guidance on smoking reduction and cessation as well as methods to cope with nicotine withdrawal. We also continued our targeted engagement and social support communications by asking people to share their thoughts toward the guidance. During the last week (low message exposure and low engagement and social support), we posted antismoking materials once per day that focused on mindfulness, self-regulatory tips, and resources for smoking reduction and cessation. Participants were blinded from the intention of the intervention designs regarding message exposure frequency and levels of engagement and social support systems.

Stimulus Materials for Antismoking Message Exposure

In order to prepare intervention materials to be disseminated on our smoking reduction and cessation Facebook group for four weeks, a total of 80 different antismoking advertisements, campaign messages, and news articles were collected from publicly available online sources, such as smokefree.gov, cancer.gov, and the CDC's Media Campaign Resource Center (MCRC), a rich database with more than 10,000 antismoking ads produced by different states and federal agencies. The collected antismoking materials were either video-based or text-and-image-based materials that have shown population-level success or promising evidence on promoting tobacco control and prevention (eg, the "Tips From Former Smokers" campaign). To select the final set of intervention materials, in a separate MTurk-based randomized experiment, we evaluated the relative effectiveness of 80 antismoking materials among 1288 smokers prior to the interventions. Based on composite scores of message effectiveness and post-antismoking attitudes toward randomly assigned antismoking material, a total of 56 antismoking messages out of the 80 units were selected as intervention materials (3 messages × 7 days for the first week, 1 message × 7 days for the second week, 3 messages × 7 days for the third week, and 1 message × 7 days for the last week).

The 56 units of antismoking messages were posted in a random order on our smoking reduction and cessation Facebook group. Based on the ongoing feedback from our participants and weekly surveys, 5 message units of these 56 (approximately 9%) were replaced with other antismoking materials to correspond to the needs of participants (eg, asking for more information on smoking cessation tips).

Systematic Delivery of Evidence-Based Antismoking Materials

Our smoking reduction and cessation Facebook group intervention started in late November 2015 and ended in early January 2016. We delivered antismoking materials with different frequencies across four intervention weeks (as described above) but with fixed time schedules: 8:00 AM, 12:00 PM, and 5:00 PM (Pacific Time) for the first and third week; and 11:00 AM (Pacific Time) for the second and fourth week.

Engagement and Social Support Systems via Social Media Features

In addition to using social media as an intervention modality where participants were exposed to antismoking messages frequently, the research team utilized communication features on the Facebook group, such as pressing the "like" button to express support and affective responses toward users' wall postings and comments and leaving "comments" to provide constructive feedback. These activities were implemented to synchronously reciprocate them with information and foster social support and user engagement.

Research Assessments

A baseline survey, all weekly surveys administered during the four-week intervention period, and a two-week follow-up survey were conducted to participants via MTurk by using the "qualification type" function on MTurk. This function made the survey available only to our intervention participants. Participants were compensated with US $8 for each baseline and weekly survey and US $15 for the two-week follow-up survey for a total of US $55 over the study period. The median values of the time spent by participants on survey assessments throughout the entire intervention period ranged between 4.21 minutes and 13.96 minutes.

XSL•FO
RenderX

Demographic and Smoking Characteristics at Baseline

Demographic information such as age, gender, marital status, ethnicity, and race, self-reported smoking status (the average number of cigarettes participants smoked in the past 7 days), motivation to quit [39], and behavioral intention to quit smoking [40] were assessed at baseline. The readiness to quit smoking was also measured on a 10-point Likert scale [41], where scores between 1 and 3 reflect low readiness to quit (eg, "I don't want to quit. Tobacco is not a problem for me."); the range between 4 and 7 indicates moderate readiness to quit (eg, "I know quitting would be good for my health. I am interested in advice about quitting."); and the scores between 7 and 10 indicate high readiness to quit (eg, "I am ready to quit using tobacco. I would like help to quit using tobacco.").

Data Analysis: Facebook ESSC Scores, Smoking Reduction and Cessation

Primary Outcomes

The primary outcome was self-reported smoking reduction reported at baseline and the last follow-up (adopted from [42,43] and modified for the study). The number of cigarettes participants smoked per week measured at follow-up was subtracted from that measured at baseline to compute the reduced number of weekly cigarettes consumed per participant. Another primary outcome was smoking cessation (7-day point smoking abstinence at the follow-up).

Predictor Variable (Facebook ESSC [/ esit/] Scores)

For the key independent variable, we constructed individual-level Facebook engagement and social support composite scores (referred to as "Facebook ESSC Scores" hereafter) to capture user engagement and the social support received from our Facebook group. The Facebook ESSC score was aggregated for each participant by combining the number of postings each participant generated (both wall postings and comments) and the number of "likes" each participant received during the intervention period. Two trained coders verified the number of likes each participant received and the number of comments or wall posts each participant made. The two coders reached a consensus on these results.

Secondary Outcomes

Secondary constructs measured at baseline, weekly, and follow-up surveys include the antismoking attitudes scale on a 7-point semantic differential scale [44], readiness to quit on a 10-point Likert scale [41,45], motivation to quit [39], self-efficacy beliefs on a 7-point Likert scale (adopted from Wei et al study [46] and modified for the study), and perceived social support on a 5-point Likert scale [47]. In all of these scales, higher values indicate positive attitudes toward smoking, greater readiness to quit smoking, greater belief in self-efficacy, and greater perceived social support, respectively.

Facebook Intervention Feasibility Inventory

We generated the Facebook Intervention Feasibility Inventory by adopting and modifying questionnaires from usability and acceptability scales that were validated in the mHealth intervention context [48-50]. The Facebook Intervention Feasibility Inventory was designed to measure the perceived feasibility of using a Facebook group for smoking reduction and cessation interventions with 24 randomly ordered items on a 7-point Likert scale at follow-up. Examples of responses include statements such as "I thought the anti-smoking Facebook Group was easy to use." and "I felt very confident using the Anti-smoking Facebook Group."

User Engagement Patterns and Message Exposure Tracking

We monitored participants' engagement with the intervention content and other members by looking at the frequency of postings and "likes" that participants generated on the Facebook group. Due to limited access to extract user data specific to "seen by" activities, our trained research members counted the number of "seen by" activities per post. We also unobtrusively observed participants' exposure to intervention messages by checking the "seen by" feature on a daily basis, which enabled the research team to track whether each user had seen the materials posted on the wall of our Facebook group.

Statistical Analysis

R package version 3.2.5 [51] and IBM SPSS statistics version 22 [52] were used for statistical analyses. Descriptive analyses were first performed on the demographic variables, predictors, and primary and secondary outcomes to summarize distributions and patterns of each construct. For the primary outcome, we ran adjusted regression analyses to examine the relationship between the predictor and the primary outcome variable while controlling for baseline covariates (gender, readiness to quit, and smoking status, ie, the number of cigarettes participants smoked in the past 7 days at baseline). For secondary outcome variables, we performed General Linear Model (GLM) analysis with repeated measures to explore the main effects of Facebook ESSC scores on repeated secondary outcomes while controlling for the same baseline covariates. We performed a confirmatory factor analysis with an oblimin rotation on the Facebook intervention feasibility questionnaires. A bivariate correlation matrix was computed to examine reliability coefficients between predictors, smoking reduction, and Facebook feasibility subfactors. We also used various R packages and publicly available online software for data visualization to demonstrate participants' engagement trends based on the total number of postings aggregated throughout the intervention period.

Results

Demographic and Smoking Characteristics at Baseline

The majority of participants were white (n=15, 94%), female (n=13, 81%), and between 25 and 50 years old (mean 34.75, SD 8.15). On average, participants smoked 11.31 cigarettes per day (SD 6.81) and 6.93 days per week (SD 0.25) at baseline. The degree of readiness scores for smoking cessation was 7.50 (SD 1.59), which indicates high readiness to quit smoking. During the past 12 months prior to baseline, participants stopped smoking 1.81 times (SD 1.47 times) for at least one day or longer. The 7-item antismoking attitude scale at baseline had a high inter-reliability (Cronbach alpha=.91) and loaded on a single confirmatory factor (with an eigenvalue=4.965 with 70% variances being explained by this one factor). Thus, we created

XSL•FO
RenderX

a composite score (mean 2.37, SD 1.22): lower values indicate strong antismoking attitudes, such as smoking cigarettes is "bad, useless, and harmful for my health"; and higher values indicate positive attitudes toward smoking, such as smoking cigarettes is "good, beneficial, and useful."

Facebook ESSC Scores, Smoking Reduction and Cessation

All participants (N=16) completed all 6-time-point surveys. Descriptive statistics of predictors (Facebook ESSC scores), primary and secondary outcomes, and Facebook intervention feasibility questionnaires are presented in Table 1. Compared with the baseline, participants reported smoking an average of 60.56 fewer cigarettes per week (SD 38.83) at the follow-up, and 25% of the participants reported 7-day point smoking abstinence at the follow-up. Facebook engagement and social support composite scores (Facebook ESSC scores) were generated (mean 19.19, SD 24.35). The final adjusted linear regression model revealed that a one-unit increase in Facebook ESSC scores predicted a 0.56 unit decrease in cigarettes consumed in the past 7 days (standard error, SE=.24, *P*=.04, 95% CI 0.024-1.09) when baseline covariate characteristics were controlled, $F_{4, 11}$=8.85, *P*=.002, adjusted R^2=.68.

Composite Scores of Secondary Outcomes and Exploratory Analyses

Seven-item antismoking attitudes showed a high reliability for each time-point survey, ranging from Cronbach alpha=.91 to .99. Thus, we created a composite score for each time-point. Similarly, 5-item self-efficacy questionnaires had a high reliability for each time-point survey, ranging from Cronbach alpha=.88 to .98. We generated a composite score for self-efficacy beliefs measured at each time-point of the surveys. For perceived social support questionnaires, five items were averaged to a single factor for each time-point survey after verifying a good reliability score for each survey (Cronbach alpha scores ranged from .84 to .92). The descriptive statistics of secondary outcomes are reported in Table 1. Exploratory GLM analysis with repeated measures revealed that Facebook ESSC scores were not significant predictors of secondary outcomes when the baseline covariates were adjusted (*Fs*<1).

Table 1. Predictors, primary and secondary outcomes, and Facebook feasibility.

Variables	Statistics					
Predictors	Mean (SD)					
Number of Facebook "likes" received	13.25 (17.67)					
Number of Facebook "comments/wall posts" generated	5.94 (6.96)					
Number of Facebook engagement scores	19.19 (24.35)					
	Baseline	Week 1	Week 2	Week 3	Week 4	Follow-up
Primary outcomes						
Mean number of cigarettes smoked per week among smokers mean (SD)	79.19 (47.66)	44.38 (60.09)	32.44 (44.42)	29.81 (41.84)	20.44 (36.51)	18.63 (35.33)
Number of people who quit smoking in the past 7 days, n (%)	0 (0)	0 (0)	2 (13)	3 (19)	3 (19)	4 (25)
Secondary outcomes, mean (SD)						
Mean Antismoking attitude scale score	2.37 (1.22)	1.76 (1.33)	1.73 (1.34)	1.66 (1.32)	1.60 (1.51)	1.60 (1.53)
Mean self-efficacy for smoking cessation scale score	–	6.06 (0.70)	5.83 (0.83)	6.18 (0.75)	6.40 (0.67)	6.05 (1.50)
Mean score on readiness to quit item	7.50 (1.59)	8.19 (1.22)	7.88 (1.59)	7.94 (1.53)	8.63 (1.20)	8.56 (1.41)
Mean perceived social support scale score	–	3.95 (0.87)	4.10 (0.64)	4.03 (0.79)	4.18 (0.84)	3.89 (0.88)
Facebook intervention feasibility questionnaires, mean (SD)						
Response efficacy (alpha =.96)	–	–	–	–	–	5.47 (1.20)
Perceived technology barriers (alpha=.97)	–	–	–	–	–	2.23 (1.45)
Easiness to use (alpha=.96)	–	–	–	–	–	6.02 (0.98)

XSL•FO
RenderX

Table 2. Correlations across predictors, smoking reduction, and Facebook feasibility subfactors. Facebook engagement and social support composite scores (Facebook ESSC scores) are combined values of the number of Facebook "likes" one received (1 in the table) and the number of Facebook "comments" and "wall postings" each person generated (2 in the table).

Variables	1	2	3	4	5	6
1 Number of Facebook "likes" received	–					
2 Number of Facebook "comments/wall posts" generated	.95[a]	–				
3 Facebook ESSC scores	.996[a]	.97[a]	–			
4 Reduced number of cigarettes smoked	.49[c]	.48	.49[c]	–		
5 Facebook response efficacy	−.19	.01	−.01	.14	–	
6 Perceived technology barriers	−.34	−.35	−.34	−.56[b]	.12	–
7 Easiness to use	.24	.20	.23	.35	.62[b]	−.15

[a]$P<.01$.

[b]$P<.05$.

[c]$P=.05$.

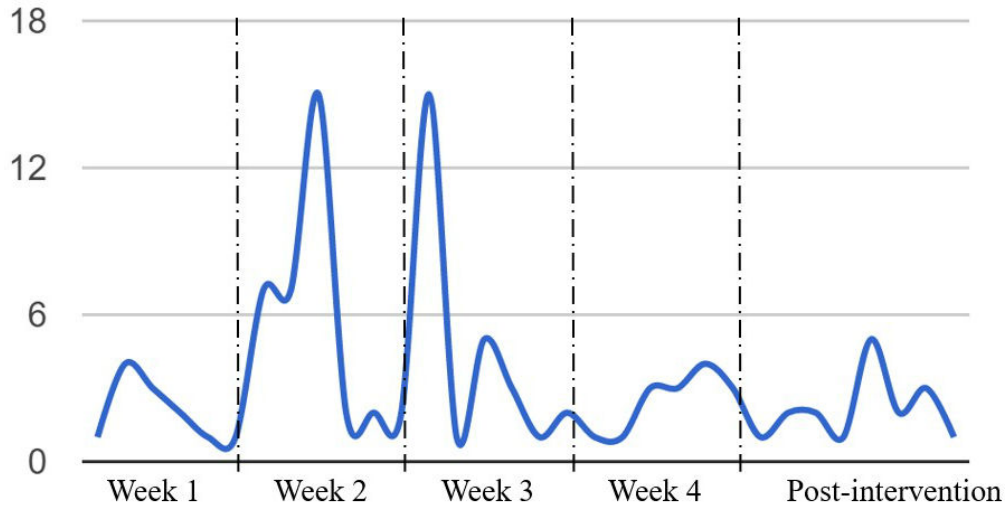
Factor Analysis on Facebook Intervention Feasibility Questionnaires

A confirmatory factor analysis with a direct oblimin rotation was performed on the 24-item Facebook intervention feasibility questionnaires. Items within each subfactor with loading scores greater than 0.6 were averaged to compose three subconcepts under the umbrella concept of perceived Facebook feasibility for health interventions. Those subfactors represent Facebook response efficacy (alpha=.96), perceived barriers of using Facebook (alpha=.97), and easiness of using Facebook for smoking reduction and cessation interventions (alpha=.96), respectively (Table 1 for descriptive analyses). Bivariate correlation coefficients between these feasibility subfactors and predictors and primary outcomes are presented in Table 2. Participants with low perceived barriers of using the Facebook Group tend to have greater reduction in weekly cigarette smoking compared to those with high perceived barriers of using the Facebook group intervention (Pearson $r=-.56$, $P=.02$).

User Engagement Patterns and Message Exposure Tracking

Figure 1 presents the aggregated number of postings from participants across the four-week intervention period. The second and third weeks, in which a high level of engagement and social support systems were targeted, demonstrate an increased generation of comments and wall posts by the participants during these two weeks. We proposed that the number of comments and wall posts indicated active user engagement. Note that week 1 and week 4 (low engagement and social support systems) revealed a decreased number of comments and wall posts in the graphic pattern, indicating relatively passive user engagement. After the intervention was over, we received a few posts from participants reporting their success in maintaining smoking cessation, as well as thanking the researchers for their help and requesting further assistance about tips for completely quitting smoking.

Figure 1. User engagement during the four-week intervention period. The y-axis indicates the number of wall posts and comments participants generated within the smoking reduction and cessation Facebook group. The values are indicative of user engagement.



XSL•FO
RenderX

Discussion

Principal Findings and Comparison With Prior Work

Using a closed Facebook group, we developed and delivered a smoking reduction and cessation intervention in a cost-effective manner while overcoming geolocational barriers and time constraints. We found that Facebook was highly feasible and we demonstrated 100% study retention and survey completion rates. In this study, 25% of participants reported 7-day smoking abstinence at the follow-up, and those who continued had dramatically reduced the number of cigarettes they smoked weekly. These outcomes are consistent with and comparable with previous studies of Facebook-delivered interventions for smoking [53].

Although we promoted smoking cessation as an optimal outcome for our intervention, we also accepted smoking reduction as a positive form of behavior change. Helping smokers to cut down on cigarette use has been attempted in many controlled trials [54]. However, we acknowledge that smoking reduction should be considered as an intermediate precursor to quitting rather than a goal of the intervention because reduced- and light-smoking still carry a considerable level of health risk for cardiovascular disease and cancer, as observed among heavy daily smokers [55]. That said, smoking reduction can have a substantial health impact in the long run [56,57]. For example, in a 30-year longitudinal follow-up cohort study, Godtfredsen and colleagues [58] found that a 50% smoking reduction significantly decreased lung cancer risk among heavy regular smokers.

Results also suggest that engagement and receipt of social support within this Facebook health communications intervention predicted smoking reduction among motivated smokers. Specifically, a one-unit increase in Facebook ESSC scores predicted a 0.56 unit decrease in cigarettes consumed in the past 7 days. That is, participants who received more “likes” and those who posted more content on our Facebook group, indicative of social support and user engagement, were more likely to reduce their weekly smoking ($F_{4,11}$=8.85, P=.002, ΔR^2=.68), suggesting a potential mechanism of action for the intervention.

We generated the term “Facebook engagement and social support composite scores” (ESSC [/ esit/] scores) in this study and tested the predictive validity of the ESSC scores on smoking reduction. The composite score was based on our conceptualization that “writing comments and wall posts” is an indicator of user engagement, as also defined by Facebook [59]. We used the “like” feature on Facebook to express our positive reaction toward participants and form perceived social support. The number of likes participants received from the interventionists and other participants within the intervention Facebook group was conceptualized as an index for receiving social support. However, we acknowledge that receiving comments (reciprocity) is not only an indicator of user engagement, but also an indicator of social support (as perceived social support increases when self-disclosed information is reciprocated [60]). That is, these two constructs (user engagement and social support) are covarying and correlated constructs as we found in our study (Pearson r=.95).

Given this conceptualization, we were interested in how the user engagement and social support systems worked synergistically to enhance intervention outcomes. Determining whether user engagement is exclusively more important than perceived social support or vice versa for predicting smoking reduction was beyond the scope of our research. Thus, we proposed a composite score by combining them to serve our conceptual approach. In addition, our approach was aligned with the principle from test theory that composite scores are more reliable than individual items [61]. Note that the number of likes participants received from peers and interventionists (beta coefficient=.72, SE=.33, P=.05) as well as the number of comments and wall posts each participant generated (beta coefficient=2.05, SE=.78, P=.02) were independently significant predictors of the reduced number of cigarettes consumed in the past 7 days. We suggest that future work disentangle the unique value of these two constructs and build a prediction model focusing on each construct as a single predictor. We also encourage future research to validate and examine the replicability of this composite score.

Throughout the intervention period, on average, participants generated six comments during the four-week intervention period (SD 6.96, median 3.50). Thrul and colleague [62] found that 79 participants made a total of 718 individual comments during the three-month intervention period, which are about nine comments per participant for three months. When averaged by month, their engagement level is three comments per participant per month. We consider our user engagement level (six comments or wall posts per participant for four-week) is comparable with other social media–based interventions for smoking. In fact, we found that the average number of comments and wall posts were negatively skewed due to observers who did not generate any comments or wall posts (n=6, 37.5%). This proportion of observers is relatively low compared with that reported in Thrul et al’s study [62]. On social media health forums, lurking or observing is a common practice [63,64]. In future studies, we hope to develop and examine engagement strategies specific to targeting these intervention observers.

To understand how and for whom social media–based interventions work, future work may examine potential moderating factors that impact the relationship between user engagement and intervention outcomes. An array of baseline characteristics have predicted technology-based intervention outcomes [65]. These characteristics may include demographic characteristics, personalities and traits (eg, self-regulation [66]), and even the stage of change for smoking cessation [62]. By investigating moderating factors in future research, researchers may identify subgroups of smokers that may benefit the most from social media–based interventions for smoking reduction and cessation.

We operationalized two key intervention components to maximize the persuasive effects of social media platforms in promoting smoking reduction and cessation: (1) exposure to antismoking messages and (2) participant engagement and social support systems. Prior studies have demonstrated that exposure

XSL•FO
RenderX

to health campaign messages can enhance health behavior by changing one's beliefs about expected health outcomes [67] and by providing educational messages to increase necessary knowledge and skills [22]. Message exposure was successful as most of the posted materials were "seen by" almost everyone throughout the intervention weeks, although there was a slight decrease in message exposure during the last week. The diminished message exposure was not surprising but rather consistent with previous social media–based health interventions [62,68].

CDC's Best Practices Report, released in 2014, recommends the reuse of existing advertisements and campaign messages rather than producing new content in order to reduce the cost, time, and untested risks associated with developing new ones [1]. In this regard, we focused on systematically delivering evidence-based antismoking messages from existing campaigns and advertisements, which not only helped us save on time and cost but also enabled us to readily examine how message exposure and engagement with the intervention content led to smoking reduction.

At the time we developed this study, there was no standardized, evidence-based model or framework applicable to designing Facebook group–based interventions for smoking reduction and cessation. Thus, based on prior health communication and technology literature, we developed an intervention model using two main components: "persuasive message exposure" and "supportive engagement systems." We used a varied frequency of message exposure (three times per day or one time per day), as there was no empirical evidence on the optimal dose of message exposure for a social media–based intervention. We randomly juxtaposed these two components (high vs low message exposure frequency × high vs low engagement and social support) to develop our intervention model. This randomly juxtaposed combination led to four weekly designs, including high message exposure and low engagement and social support systems for week 1; low message exposure and high engagement and social support systems for week 2; high message exposure and high engagement and social support systems for week 3; and low message exposure and low engagement and social support systems for week 4. Our findings should be understood with caution. We did not examine which weeks resulted in the most successful intervention outcome (smoking reduction), but we tested the overall impact of the intervention as a whole (before and after the intervention) on smoking reduction. Thus, the risk of any possible confounding effect due to the varied frequency of message exposure is minimized because we did not test smoking reduction by individual week.

An interesting finding about user engagement is that although we manipulated the frequency of message exposure, there was no resulting effect on increasing user engagement, as shown in Figure 1. The targeted engagement and social support systems directly influenced user engagement, not the message exposure frequency. The distinctive weekly patterns shown in Figure 1 highlight the notion that in order to foster active user engagement, a significant predictor of smoking reduction, interventionists should directly target user interactivities. For example, the second week delivered a high level of user engagement by directly asking personally relevant questions and encouraging participants to share their progress with others in the Facebook group [69] (eg, "What motivates you to quit smoking?" and "Have you reduced the number of cigarettes today?"). Note that week 2 and week 4 had the same messaging frequency: antismoking message was posted once a day during these two weeks. When controlling for the messaging frequency as once per day, the number of user comments and wall posts (indicative of user engagement) in week 2 was higher than that of user comments and wall posts during week 4. This pattern indicates that interventionists should specifically target user engagement in addition to posting antismoking messages. We encourage researchers to adopt and apply effective persuasion tactics and principles to strategically target user engagement within social media–based health interventions.

Another strength of the study was 100% study retention throughout the MTurk-linked surveys at six different time points. Various technology features on MTurk, such as qualification assignment and the online payment system allowed us to conduct longitudinal surveys. We demonstrated that MTurk can be a platform for a wide range of research activities, ranging from recruitment of smokers living in the United States to multiple times of follow-up assessments with participants.

Limitations and Future Directions

Our findings successfully demonstrated the feasibility of social media technologies to offer smoking reduction and cessation interventions with the strategic delivery of engagement and social support systems. Our findings, however, should be understood within the limitations imposed by research budgets and the scope of the study. We did not objectively verify self-reported abstinence; thus, it is possible that the impact of the intervention may be inflated. The reported outcome on the reduced number of cigarettes per week does not correspond to the same level of reduction in toxicant exposure. In future technology-based interventions for smoking reduction and cessation, researchers should embrace practically feasible methods for measuring objective markers of nicotine toxicology [70].

Our sample size in this study was relatively small. Thus, rather than using a complex modeling approach such as latent growth curve modeling, we simplified our statistical model and directly examined the predictive value of Facebook-mediated engagement and social support in explaining smoking reduction outcomes. The dataset of 16 participants with no missing data still provided enough statistical power to detect the effect of the primary regression model outcomes.

Another limitation is that the gender and race of our sample were relatively homogeneous, mostly white women. In future research, we hope to replicate the interventions with bigger sample sizes and involve participants with characteristics that are more heterogeneous than the current sample to establish the generalizability and reproducibility of the findings. With an increased sample size, future studies should examine pathways of intervention processes with intermediate factors, such as enhanced self-efficacy and perceived social support, to reflect the dynamics of behavior change [71,72]. Also, Figure 1 shows different descriptive patterns by week. We did not examine a statistical difference using repeated measures by week on the

XSL•FO
RenderX

number of wall posts and comments (user engagement). In future research, we encourage to examine statistical difference on this engagement matrix, and perhaps to design multiple Facebook groups to prevent any potential confounding or spillover effects across weeks.

We examined the feasibility of the communication features in our Facebook group that were utilized to deliver theory-guided intervention components such as message exposure and engagement and social support systems among the optimal set of participants (self-motivated participants who wanted to quit smoking at baseline). In future work, another important question might be whether social media–based interventions can have a significant impact on enhancing these intermediate factors (eg, enhancing motivation to quit, and pro-quitting attitudes), even among those with low motivation to quit smoking at baseline. If social media–based interventions can successfully enhance those factors and smoking reduction and cessation among participants with low motivation, the expected significance of the interventions can be much greater than this study.

After our four-week intervention, followed by a two-week follow-up survey, we learned that participants continued to use our Facebook group and some participants expressed that they wanted the interventions for a longer period of time. Social media platforms provide novel opportunities to operationalize persuasive technologies for scalable interventions and to maintain active engagement and long-lasting intervention outcomes [18]. Future efforts in this line of research may examine and identify which communication and intervention strategies are most effective in sustaining active user engagement and maintaining long-lasting social ties for supportive networks. Additionally, there are several challenges and questions to consider for the future implementation of this work, including finding an optimal dose and information balance between support providers and support recipients and protecting the privacy of online intervention participants, especially for a large-scale intervention.

As reviewed, theory-driven and evidence-based interventions using Facebook for health promotions are promising. A growing line of research has shown positive effects of Facebook use on various health outcomes, from smoking cessation [62,73] and physical activities [74] to sexual health promotion [9,75]. The benefits of social media, however, go beyond its technological affordability, scalability, and accessibility. In fact, social media use provides psychological benefits that are essential to fundamental human needs. Researchers from various fields have examined psychological benefits and gratification from using Facebook [60,76-81], such as enhanced self-esteem and psychological well-being [78], increased social capital [79], and refinement of self-affirming values [77]. Furthermore, technology features of Facebook facilitate self-disclosure and reciprocal interactions with others, and these activities have been found to be intrinsically therapeutic and rewarding for humans [82]. Although targeting these psychological benefits were not within the scope of our study, and thus were not measured nor manipulated, we hope future work will consider how to actively facilitate these psychological benefits when using Facebook as an intervention tool for health promotions.

Conclusions

This study is the first Facebook-mediated intervention research that systemically promoted antismoking communication strategies and social support and engagement systems as mechanisms of behavior change within a Facebook group. We conceptualized Facebook “likes” and “wall postings and comments” as the manifestation of social support and user engagement. Our findings imply that receiving one Facebook “like” or posting on the Facebook group at least once predicts almost one less cigarette in the past 7 days. The study supports positive effects of Facebook-mediated communication, engagement and social support systems for smoking reduction and cessation, and highlights the public health potential of social media interventions for scaling-up tobacco control and prevention efforts. It also provides practical guidelines for designing communication strategies and persuasive, social media–based smoking reduction and cessation interventions that might be useful for future research.

Acknowledgments

This study was funded by the National Institutes of Health (NIH) NIDA Grant No. P30DA029926 Pilot Core. NIDA had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication. The authors would like to thank Gene Geter for his assistance in data collection, as well as Lamar D Moss for his sharp proofreading and valuable feedback.

Conflicts of Interest

None declared.

References

1. U.S. Department of Health and Human Services. Surgeon General. 2014. The health consequences of smoking—50 years of progress: a report of the Surgeon General URL: https://www.surgeongeneral.gov/library/reports/50-years-of-progress/full-report.pdf
2. Noar SM. A 10-year retrospective of research in health mass media campaigns: where do we go from here? J Health Commun 2006;11(1):21-42. [doi: 10.1080/10810730500461059] [Medline: 16546917]

XSL•FO
RenderX

3. Statista. 2016. Leading social networks worldwide as of January 2016, ranked by number of active users URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2017-01-02] [WebCite Cache ID 6nCrZGfxF]
4. Nabi RL, Prestin A, So J. Facebook friends with (health) benefits? exploring social network site use and perceptions of social support, stress, and well-being. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):721-727. [doi: 10.1089/cyber.2012.0521] [Medline: 23790356]
5. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
6. MacLean D, Gupta S, Lembke A, Manning C, Heer J. An analysis of an online health forum dedicated to addiction recovery. 2675146: ACM; 2015 Presented at: Proceedings of the 18th ACM Conference on Computer Supported Cooperative WorkSocial Computing; March 14–18, 2015; Vancouver, BC, Canada. [doi: 10.1145/2675133.2675146]
7. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
8. Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. J Med Internet Res 2011;13(4):e122 [FREE Full text] [doi: 10.2196/jmir.1696] [Medline: 22204869]
9. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
10. Phua J. Participating in health issue-specific social networking sites to quit smoking: how does online social interconnectedness influence smoking cessation self-efficacy? J Commun 2013 Sep 16;63(5):933-952. [doi: 10.1111/jcom.12054]
11. Bazarova N, Choi Y, Sosik V, Cosley D, Whitlock J. Social sharing of emotions on facebook: channel differences, satisfaction,replies. 2675297: ACM; 2015 Presented at: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; March 14-18, 2015; Vancouver, BC, Canada p. 154-164.
12. Zhao X, Cai X. The role of risk, efficacy, and anxiety in smokers' cancer information seeking. Health Commun 2009 Apr;24(3):259-269. [doi: 10.1080/10410230902805932] [Medline: 19415558]
13. Newman M, Lauterbach D, Munson S, Resnick P, Morris M. i don't have problems, i'm just not putting them on facebook: challenges and opportunities in using online social networks for health. 1958876: ACM; 2011 Presented at: It's not that Proceedings of the ACM 2011 conference on Computer Supported Cooperative Work; March 19-23, 2011; Hangzhou, China p. 341-350.
14. Lee K, Noh M, Koo D. Lonely people are no longer lonely on social networking sites: the mediating role of self-disclosure and social support. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):413-418. [doi: 10.1089/cyber.2012.0553] [Medline: 23621716]
15. Ellison N, Steinfield C, Lampe C. The benefits of facebook “Friends:” social capital and college students’ use of online social network sites. J Comput Mediat Commun 2007;12:1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
16. Davis MA, Anthony DL, Pauls SD. Seeking and receiving social support on facebook for surgery. Soc Sci Med 2015 Apr;131:40-47 [FREE Full text] [doi: 10.1016/j.socscimed.2015.02.038] [Medline: 25753284]
17. 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]
18. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]
19. Durkin S, Brennan E, Wakefield M. Mass media campaigns to promote smoking cessation among adults: an integrative review. Tob Control 2012 Feb 16;21(2):127-138. [doi: 10.1136/tobaccocontrol-2011-050345]
20. Hughes J, Carpenter M. The feasibility of smoking reduction: an update. Addiction (Abingdon, England) 2005;100(8):1074-1089.
21. Fishbein M, Cappella JN. The role of theory in developing effective health communications. J Commun 2006 Aug;56(s1):S1-S17. [doi: 10.1111/j.1460-2466.2006.00280.x]
22. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
23. Schooler C, Chaffee SH, Flora JA, Roser C. Health campaign channels tradeoffs among reach, specificity, and impact. Human Comm Res 1998 Mar;24(3):410-432. [doi: 10.1111/j.1468-2958.1998.tb00423.x]
24. Hornik R. The knowledge-behavior gap in public information campaigns: a development communication view. In: Salmon CT, editor. Information campaigns: Balancing social values and social change. Newbury Park, CA: Sage Publications; 1989.

XSL•FO
RenderX

25. Centers for Disease Control Prevention. Best practices for comprehensive tobacco control programs. In: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. 2014 Presented at: National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014; Atlanta, GA.
26. Gottlieb BH, Bergen AE. Social support concepts and measures. J Psychosom Res 2010 Nov;69(5):511-520. [doi: 10.1016/j.jpsychores.2009.10.001] [Medline: 20955871]
27. Namkoong K, Shah DV, Gustafson DH. Offline social relationships and online cancer communication: effects of social and family support on online social network building. Health Commun 2016 Nov 08:1-8. [doi: 10.1080/10410236.2016.1230808] [Medline: 27824257]
28. Havassy BE, Hall SM, Wasserman DA. Social support and relapse: commonalities among alcoholics, opiate users, and cigarette smokers. Addict Behav 1991;16(5):235-246. [Medline: 1663695]
29. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
30. Fox S. Pew Research Center. 2011. Peer-to-peer health care URL: http://www.pewinternet.org/2011/02/28/peer-to-peer-health-care-2/ [accessed 2017-05-02] [WebCite Cache ID 6qANEwqa7]
31. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting behavioral interventions for social media delivery. J Med Internet Res 2016;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
32. Warren JI, Stein JA, Grella CE. Role of social support and self-efficacy in treatment outcomes among clients with co-occurring disorders. Drug Alcohol Depend 2007 Jul 10;89(2-3):267-274 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.01.009] [Medline: 17329040]
33. Neuling SJ, Winefield HR. Social support and recovery after surgery for breast cancer: frequency and correlates of supportive behaviours by family, friends and surgeon. Soc Sci Med 1988;27(4):385-392. [Medline: 3175721]
34. Mason W, Suri S. Conducting behavioral research on Amazon's mechanical turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
35. Buhrmester M, Kwang T, Gosling SD. Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Feb 03;6(1):3-5. [doi: 10.1177/1745691610393980]
36. Kim SJ, Hancock JT. How advertorials deactivate advertising schema: MTurk-based experiments to examine persuasion tactics and outcomes in health advertisements. Comm Res 2016 Apr 25:- Epub ahead of print. [doi: 10.1177/0093650216644017]
37. Yu B, Willis M, Sun P, Wang J. Crowdsourcing participatory evaluation of medical pictograms using Amazon mechanical turk. J Med Internet Res 2013;15(6):e108 [FREE Full text] [doi: 10.2196/jmir.2513] [Medline: 23732572]
38. Snow R, O'Connor B, Jurafsky D, Ng A. Cheap and fast---but is it good? evaluating non-expert annotations for natural language tasks. 1613751: Association for Computational Linguistics; 2008 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing; October 25-27, 2008; Honolulu, HI p. 254-263.
39. Wilson DK, Wallston KA, King JE. Effects of contract framing, motivation to quit, and self-efficacy on smoking reduction. J Appl Social Pyschol 1990 Apr;20(7):531-547. [doi: 10.1111/j.1559-1816.1990.tb00426.x]
40. Sterling KL, Diamond PM, Mullen PD, Pallonen U, Ford KH, McAlister AL. Smoking-related self-efficacy, beliefs, and intention: assessing factorial validity and structural relationships in 9 th–12th grade current smokers in Texas. Addict Behav 2007;32(9):1863-1876. [doi: 10.1016/j.addbeh.2006.12.018]
41. Maryland Resource Center. The A3C clinical model URL: http://mdquit.org/fax-to-assist/module-2 [accessed 2017-01-11] [WebCite Cache ID 6nRaFo6X7]
42. Norman P, Conner M, Bell R. The theory of planned behavior and smoking cessation. Health Psychol 1999 Jan;18(1):89-94. [Medline: 9925050]
43. Douglas KA, Collins JL, Warren C, Kann L, Gold R, Clayton S, et al. Results from the 1995 National College Health Risk Behavior Survey. J Am Coll Health 1997 Sep;46(2):55-66. [doi: 10.1080/07448489709595589] [Medline: 9276349]
44. Kim SJ, Niederdeppe J. Emotional expressions in antismoking television advertisements: consequences of anger and sadness framing on pathways to persuasion. J Health Commun 2014;19(6):692-709. [doi: 10.1080/10810730.2013.837550] [Medline: 24417482]
45. Chung T, Maisto SA, Mihalo A, Martin CS, Cornelius JR, Clark DB. Brief assessment of readiness to change tobacco use in treated youth. J Subst Abuse Treat 2011 Sep;41(2):137-147 [FREE Full text] [doi: 10.1016/j.jsat.2011.02.010]
46. Wei R, Lo V, Lu H. Reconsidering the relationship between the third-person perception and optimistic bias. Comm Res 2007 Dec 01;34(6):665-684. [doi: 10.1177/0093650207307903]
47. Wright KB, Rosenberg J, Egbert N, Ploeger NA, Bernard DR, King S. Communication competence, social support, and depression among college students: a model of facebook and face-to-face support network influence. J Health Commun 2013;18(1):41-57. [doi: 10.1080/10810730.2012.688250] [Medline: 23030518]
48. Lund A. STC Usability SIG Newsletter. 2000. Measuring usability with the USE questionnaire URL: http://www.stcsig.org/usability/newsletter/index.html [accessed 2017-05-02] [WebCite Cache ID 6qAQFwDvc]
49. Lewis J. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]

XSL•FO
RenderX

50. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253. [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
51. R Core Team. R-project. Vienna, Austria; 2016. R: A language and environment for statistical computing URL: http://www.R-project.org [accessed 2017-05-02] [WebCite Cache ID 6qAQYB5kB]
52. IBM. 2013. IBM SPSS Statistics for Windows URL: https://www.ibm.com/us-en/ [accessed 2017-05-02] [WebCite Cache ID 6qAQimBxs]
53. 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]
54. Stead LF, Lancaster T. Interventions to reduce harm from continued tobacco use. Cochrane Database Syst Rev 2007 Jul 18(3):CD005231. [doi: 10.1002/14651858.CD005231.pub2] [Medline: 17636791]
55. Bjartveit K, Tverdal A. Health consequences of smoking 1-4 cigarettes per day. Tob Control 2005 Oct;14(5):315-320 [FREE Full text] [doi: 10.1136/tc.2005.011932] [Medline: 16183982]
56. Wennike P, Danielsson T, Landfeldt B, Westin A, Tønnesen P. Smoking reduction promotes smoking cessation: results from a double blind, randomized, placebo-controlled trial of nicotine gum with 2-year follow-up. Addiction 2003 Oct;98(10):1395-1402. [Medline: 14519176]
57. Vangeli E, Stapleton J, Smit ES, Borland R, West R. Predictors of attempts to stop smoking and their success in adult general population samples: a systematic review. Addiction 2011 Dec;106(12):2110-2121. [doi: 10.1111/j.1360-0443.2011.03565.x] [Medline: 21752135]
58. Godtfredsen NS, Prescott E, Osler M. Effect of smoking reduction on lung cancer risk. JAMA 2005 Sep 28;294(12):1505-1510. [doi: 10.1001/jama.294.12.1505] [Medline: 16189363]
59. Facebook. 2017. How is engagement rate defined? URL: https://www.facebook.com/help/178043462360087 [accessed 2017-05-02] [WebCite Cache ID 6qAR6gppN]
60. Bazarova NN, Choi YH. Self-disclosure in social media: extending the functional approach to disclosure motivations and characteristics on social network sites. J Commun 2014 Jun 19;64(4):635-657. [doi: 10.1111/jcom.12106]
61. Nunnally J. Psychometric theory. New York: McGraw-Hill; 1978.
62. Thrul J, Klein AB, Ramo DE. Smoking cessation intervention on facebook: which content generates the best engagement? J Med Internet Res 2015;17(11):e244 [FREE Full text] [doi: 10.2196/jmir.4575] [Medline: 26561529]
63. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008 Jun 30;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
64. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
65. Kim SJ, Marsch LA, Guarino H, Acosta MC, Aponte-Melendez Y. Predictors of outcome from computer-based treatment for substance use disorders: Results from a randomized clinical trial. Drug Alcohol Depend 2015;157:174-178. [doi: 10.1016/j.drugalcdep.2015.09.019]
66. Leventhal H, Leventhal EA, Contrada RJ. Self-regulation, health, and behavior: a perceptual-cognitive approach. Psychol Health 1998 Jul;13(4):717-733. [doi: 10.1080/08870449808407425]
67. Fishbein M, Yzer MC. Using theory to design effective health behavior interventions. Commun Theory 2003 May;13(2):164-183. [doi: 10.1111/j.1468-2885.2003.tb00287.x]
68. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013 Sep;7(3):355-368 [FREE Full text] [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]
69. Lee EJ, Kim YW. Effects of infographics on news elaboration, acquisition, and evaluation: prior knowledge and issue involvement as moderators. New Media & Soc 2015 Jan 14;18(8):1579-1598. [doi: 10.1177/1461444814567982]
70. Dallery J, Raiff BR, Kim SJ, Marsch LA, Stitzer M, Grabinski MJ. Nationwide access to an internet-based contingency management intervention to promote smoking cessation: a randomized controlled trial. Addiction 2016 Dec 06;112(5):875-883. [doi: 10.1111/add.13715] [Medline: 27923264]
71. Ajzen I, Fishbein M. Attitude-behavior relations: a theoretical analysis and review of empirical research. Psychol Bull 1977;84(5):888-918. [doi: 10.1037/0033-2909.84.5.888]
72. Fishbein M, Yzer MC. Using theory to design effective health behavior interventions. Commun Theory 2003 May;13(2):164-183. [doi: 10.1111/j.1468-2885.2003.tb00287.x]
73. Cobb NK, Jacobs MA, Wileyto P, Valente T, Graham AL. Diffusion of an evidence-based smoking cessation intervention through facebook: a randomized controlled trial. Am J Public Health 2016 Jun;106(6):1130-1135. [doi: 10.2105/AJPH.2016.303106] [Medline: 27077358]

XSL•FO
RenderX

74. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2014;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
75. Nguyen P, Gold J, Pedrana A, Chang S, Howard S, Ilic O, et al. Sexual health promotion on social networking sites: a process evaluation of The FaceSpace Project. J Adolesc Health 2013 Jul;53(1):98-104. [doi: 10.1016/j.jadohealth.2013.02.007] [Medline: 23583509]
76. Kim SJ, Hancock JT. Optimistic bias and Facebook use: self-other discrepancies about potential risks and benefits of facebook use. Cyberpsychol Behav Soc Netw 2015 Apr;18(4):214-220. [doi: 10.1089/cyber.2014.0656] [Medline: 25879379]
77. Toma CL, Hancock JT. Self-affirmation underlies Facebook use. Pers Soc Psychol Bull 2013 Mar;39(3):321-331. [doi: 10.1177/0146167212474694] [Medline: 23359086]
78. Gonzales AL, Hancock JT. Mirror, mirror on my facebook wall: effects of exposure to Facebook on self-esteem. Cyberpsychol Behav Soc Netw 2011;14(1-2):79-83. [doi: 10.1089/cyber.2009.0411] [Medline: 21329447]
79. Ellison NB, Lampe C, Steinfield C, Vitak J. With a little help from my friends: how social network sites affect social capital processes. In: Papacharissi Z, editor. A networked self: identity, community and culture on social network sites. New York: Routledge; 2011.
80. Burke M, Kraut RE. The relationship between facebook use and well-being depends on communication type and tie strength. J Comput Mediat Commun 2016 Jul 26;21(4):265-281. [doi: 10.1111/jcc4.12162]
81. Burke M, Marlow C, Lento T. Social network activitysocial well-being. 1753613: ACM; 2010 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA p. 1909-1912.
82. Tamir DI, Mitchell JP. Disclosing information about the self is intrinsically rewarding. Proc Natl Acad Sci U S A 2012 May 22;109(21):8038-8043 [FREE Full text] [doi: 10.1073/pnas.1202129109] [Medline: 22566617]

Abbreviations

Facebook ESSC Score: Facebook engagement and social support composite score

Edited by CL Parra-Calderón; submitted 22.09.16; peer-reviewed by J Mitchell, J Thrul; comments to author 15.10.16; revised version received 17.03.17; accepted 02.04.17; published 23.05.17.

Please cite as:
Kim SJ, Marsch LA, Brunette MF, Dallery J
Harnessing Facebook for Smoking Reduction and Cessation Interventions: Facebook User Engagement and Social Support Predict Smoking Reduction
J Med Internet Res 2017;19(5):e168
URL: http://www.jmir.org/2017/5/e168/
doi:10.2196/jmir.6681
PMID:28536096

©Sunny Jung Kim, Lisa A Marsch, Mary F Brunette, Jesse Dallery. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Social Media Performance of Hospitals in Spain: A Longitudinal and Comparative Study

Antonio Martinez-Millana[1*], MSc; Carlos Fernandez-Llatas[1,2*], PhD; Ignacio Basagoiti Bilbao[3*], MD; Manuel Traver Salcedo[1*], BSc (Comp Eng); Vicente Traver Salcedo[1,2*], PhD

[1]ITACA, Universitat Politècnica de València, Valencia, Spain

[2]Unidad Mixta de Reingeniería de Procesos Sociosanitarios (eRPSS), Instituto de Investigación Sanitaria, Hospital Universitario y Politecnico La Fe, Valencia, Spain

[3]Hospital de Manises, Valencia, Spain

[*]all authors contributed equally

Corresponding Author:
Antonio Martinez-Millana, MSc
ITACA
Universitat Politècnica de València
Buildign 8G Access B 1st Fl
Camino de Vera s/n. 46022
Valencia,
Spain
Phone: 34 963877606
Email: anmarmil@itaca.upv.es

Abstract

Background: Social media is changing the way in which citizens and health professionals communicate. Previous studies have assessed the use of Health 2.0 by hospitals, showing clear evidence of growth in recent years. In order to understand if this happens in Spain, it is necessary to assess the performance of health care institutions on the Internet social media using quantitative indicators.

Objectives: The study aimed to analyze how hospitals in Spain perform on the Internet and social media networks by determining quantitative indicators in 3 different dimensions: presence, use, and impact and assess these indicators on the 3 most commonly used social media - Facebook, Twitter, YouTube. Further, we aimed to find out if there was a difference between private and public hospitals in their use of the aforementioned social networks.

Methods: The evolution of presence, use, and impact metrics is studied over the period 2011- 2015. The population studied accounts for all the hospitals listed in the National Hospitals Catalog (NHC). The percentage of hospitals having Facebook, Twitter, and YouTube profiles has been used to show the presence and evolution of hospitals on social media during this time. Usage was assessed by analyzing the content published on each social network. Impact evaluation was measured by analyzing the trend of subscribers for each social network. Statistical analysis was performed using a lognormal transformation and also using a nonparametric distribution, with the aim of comparing *t* student and Wilcoxon independence tests for the observed variables.

Results: From the 787 hospitals identified, 69.9% (550/787) had an institutional webpage and 34.2% (269/787) had at least one profile in one of the social networks (Facebook, Twitter, and YouTube) in December 2015. Hospitals' Internet presence has increased by more than 450.0% (787/172) and social media presence has increased ten times since 2011. Twitter is the preferred social network for public hospitals, whereas private hospitals showed better performance on Facebook and YouTube. The two-sided Wilcoxon test and *t* student test at a CI of 95% show that the use of Twitter distribution is higher (P<.001) for private and public hospitals in Spain, whereas other variables show a nonsignificant different distribution.

Conclusions: The Internet presence of Spanish hospitals is high; however, their presence on the 3 main social networks is still not as high compared to that of hospitals in the United States and Western Europe. Public hospitals are found to be more active on Twitter, whereas private hospitals show better performance on Facebook and YouTube. This study suggests that hospitals, both public and private, should devote more effort to and be more aware of social media, with a clear strategy as to how they can foment new relationships with patients and citizens.

XSL•FO
RenderX

(J Med Internet Res 2017;19(5):e181) doi:10.2196/jmir.6763

KEYWORDS

public health; delivery of health care; Internet; social media; hospitals

Introduction

The Spanish health care system is one of the best ranked in the world as far as patient safety [1], efficiency [2], and satisfaction [3] is concerned. The health sector in Spain represents 9.5% (118.9/1.252 USD trillion) of the gross domestic product (2.5% for private purposes and 7% for public services). The Public health care copes the majority of delivery of health services, but in a recent study it was shown that private hospitals performed 32% (1.12/3.50 million) of surgeries, responded to 21% (9.91/47.2 million) of emergencies, and took up 15% of outpatients' referrals (11.83/78.9 million) [4]. Moreover, in a recent survey made on Spanish citizens looking for medical assistance, 60% said they preferred being assisted in a public hospital, whereas 28% preferred private hospitals [5]. This was because they considered public hospitals to have better technological installations, more capable physicians and nurses, and because public hospitals provided more and better information than private hospitals [5]. Therefore, information plays a big role in the way citizens make their choices about health management, and the need for information is increasing exponentially [6]. The need for accurate information is particularly critical when the information involves health, well-being, and disease, especially when Internet sources are winning the battle over traditional sources of information [7].

Internet health-related queries rose from 2010 [7-9], and they were mainly used to support a decision, such as looking for a second opinion or even purchasing drugs [10]. Even junior physicians consult information provided on the Internet to reinforce the diagnosis and treatment decisions they make on a daily basis [11]. One example of how people use Internet sources to assess their illness is Wikipedia [12,13], which hosts a large quantity of information about medical data [14], even comparable with commercial encyclopedias [15]. The easy access and the easy-to-understand development of health topics are turning Wikipedia into the first-choice Internet source to find brief and clear definition of a specific term, including health terminologies [14].

This scenario is defining a new paradigm in which health services' consumers and procurers (patients and health professionals) share a new framework for information exchange [16]. The unstoppable advance of social media in medicine is now a reality, and it is pushing health professionals and hospitals to learn, start, and increase their use of social media as a communication channel. Business concepts are currently being studied to improve marketing strategies for hospitals [17] and to amplify health values and principles [18]. These developments reflect the manner in which health professionals are applying their knowledge and experience in terms of interacting with patients [19].

Hospitals cannot control the information in social media [20]; on the other hand, patient communities have taken the lead in allowing the sharing of medical experiences on social media [21], and some social media sites have empowered patients to provide personal ratings on their health care experiences [22]. The relationship between hospital social media activity and quality key performance indicators are still quite unexplored; however, it has become increasingly critical to find effective ways of communicating with the community outside clinical environments as traditional communication channels such as Web 1.0, electronic mail, and media campaigns are being replaced by new communication channels [16]. Three of the key indicators used previously are (1) the presence, defined as the rate of health care entities with a profile or page on a social network; (2) the use (or usage), defined as the number of posts with content published in a time window; and (3) the impact, defined as the capability of an entity to gain subscribers [23,24]. In this paper we present a 5-year longitudinal study on the use of webpages and social media among public and private hospitals in Spain to evaluate the aforementioned indicators. Our hypothesis is that public and private hospitals perform differently, as regards to the final target of each type of entity. Presence, use, and impact of social media profiles have been analyzed to determine how metrics have evolved over time and which direction they will take in the future, by comparing the performance of public and private hospitals over 5 years on the 3 main social networks: Facebook, Twitter, and YouTube. Finally, results are compared with previous publications in the United States and Western Europe. The statistical analysis of the data allows us to confirm that there is a statistically significant difference in the use of Twitter between private and public hospitals. Spain is progressing well in the adoption and use of social media, but our findings reinforce the need to promote new forms of communication by public hospitals in the era of social communication, by using innovative channels to reach a bigger audience.

Methods

Study Design

A longitudinal review of hospitals that have presence on 3 of the most popular social networks—Facebook, Twitter, and YouTube—was conducted. For each hospital, data about the use and user acceptation of the generated content was collected, as well as general information about the hospital (eg, public or private ownership).

Data Collection

The studied cohort included all the hospitals listed in the "National Catalog of Public and Private Hospital Centers (NHC)" maintained by the Spanish Ministry of Health and Social Affairs [25]. The overall study cohort included 787 hospitals. Webpages (Web 1.0) and social media profiles were discovered using contact data, such as the name of the institution, address, and municipality on the Google search engine. The sites were validated by accessing the search

XSL•FO
RenderX

resulting pages manually and verifying that the content corresponded to the appropriate hospital. Only institutional profiles were included in the study population. Personal, department, service, or unofficial profiles were not included. Hospital ownership (public or private) was obtained from the NHC. Retrieved social media profiles for each hospital were classified as belonging to 1 of 3 social media networks—Facebook, Twitter, or YouTube. In order to avoid the effect of stationary events (eg, winter or summer campaigns), the temporal window to retrieve data was fixed from January 2011 to December 2015. As in previous studies, these 3 social networks were selected because of their popularity and the possibility of accessing performance metrics. Data included whether hospitals had accounts on social networks (presence), their activity on those accounts (use), and how those activities were received by the intended audience (impact).

Statistical Analysis

The study assessed 3 factors: presence, usage, and impact, similar to previous studies [23,26]. The percentage of hospitals having Facebook, Twitter, and YouTube profiles was used to indicate the presence of hospitals on social media. Usage was assessed by analyzing the content generated on each social network (eg, number of tweets and videos) over the period studied. Impact was measured by the number of subscribers for each social media account. These 3 factors are considered sufficient to evaluate the extent to which hospitals are present in social media, to assess their performance on social media in terms of communication (using number of posts), and whether users were consuming the disseminated content by subscribing to a particular account or channel. All these indicators were analyzed with respect to hospital ownership: public or private. For all the observed variables, we present the median, interquartile range, and kurtosis value. Goodness of fit to a normal distribution was evaluated by two techniques. First, according to [26], a lognormal transformation was used to approximate the skewed distribution of the variables to a normal distribution; a nonparametric distribution was used to obtain the raw probability density function. Both approximations were compared with raw data to assess reliability. Independence of public and private hospital results was assessed with the two-sided, *t* student test, and Wilcoxon test (CI of 95%) for each distribution. Statistical significance was considered for *P* values under .05. MATLAB statistics toolbox (version 2016R) was used to perform correlation, transformation, and independence tests [24].

Results

Presence Dimension

From a total of 787 hospitals identified in the NHC of the Ministry of Health and Social Affairs, 550 had an institutional webpage, and 269 of them had at least one profile in one of social networks considered in December 2015. Figure 1 shows the evolution of Internet presence among public and private hospitals from 2011 to 2015. Even though Internet presence increased by more than 450.0% (787/172) and social media presence increased ten times since 2011, there are still many hospitals without Web 1.0 and social media profiles. However, the correlation is strong between the evolution of hospitals with an institutional webpage (Web 1.0) and presence in social media, with a value coefficient of 0.949. In December 2015, from the total number of hospitals (787) only 69.9% (550) had an official Web 1.0 and only 34.2% (269) were present in any of the social networks considered. Taking into account only those hospitals with an Internet presence, up to 48% (264) of them had a profile page in at least one of the studied social networks.

Beyond new hospital openings and new profiles on social networks, variations in the number of hospitals (continuous line in Figure 1) are due to modifications and updates done in the NHC. Variations in the number of Web 1.0 pages and social media profiles are due to the corporative acquisitions of the owners of private hospitals and profile relabeling, resulting in a merger of social media profiles.

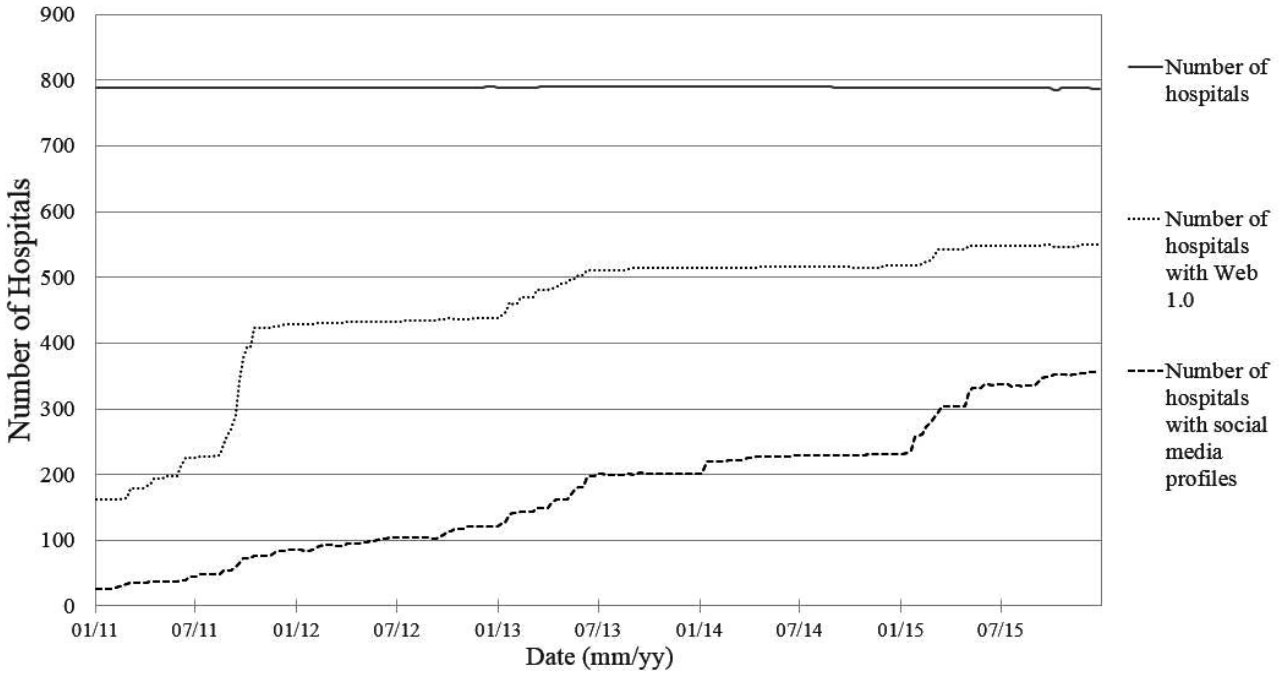
The aggregated and comparative distribution of the presence of public versus private hospitals shows that public hospitals have less Internet presence than private hospitals, not only regarding Web 1.0 pages where we find 24% (186) versus 29% (283) but also in the use of social media to disseminate information on their activities, which accounts for 33% (262) versus 60% (419). Focusing on the presence (profiles in social media), Figure 2 shows the trend from 2011 to 2015. It shows that the creation of profiles is very similar (correlation coefficient=.971), but that private hospitals present a viral growth in two specific periods (first semesters of 2013 and 2015—plain transition means no variation and a step transition means a high transition).

Despite the fact that social media has a lower presence overall in public hospitals, there are more public than private hospitals using Twitter (44% vs 36%), whereas with Facebook and YouTube, it is the opposite (31% vs 36% and 24% vs 27%). With respect to YouTube, the presence percentage is similar for private and public hospitals.

Data of the historical evolution of social media profiles disaggregated by social media type and public or private hospitals (Figure 2) confirms the previous result, and even though there exists a strong correlation between the creations of social media profiles, a slightly different volume is observed depending on the social network and entity; public hospitals have fewer profiles considering absolute numbers.

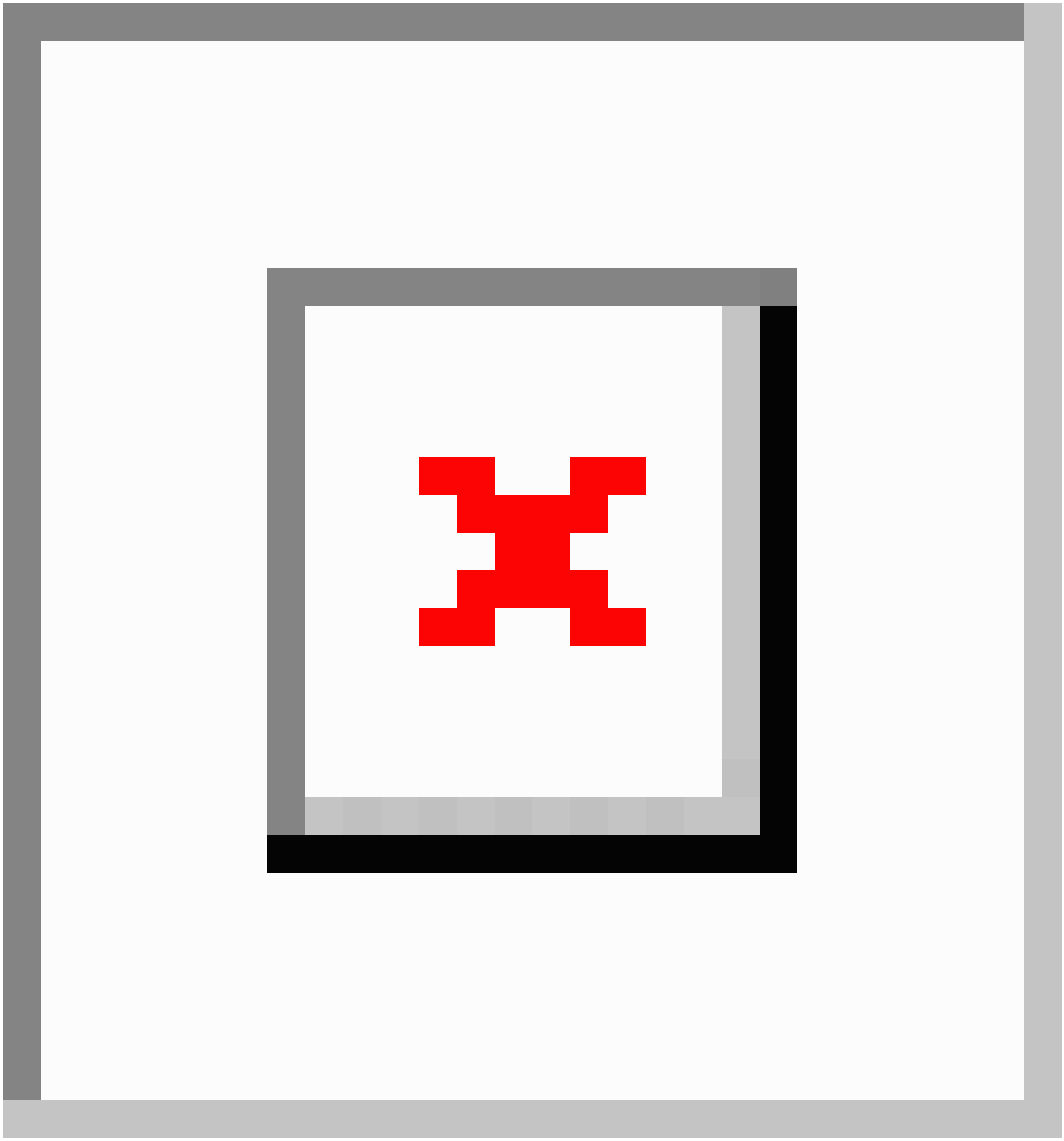
XSL•FO
RenderX

Figure 1. Evolution of Internet presence of Spanish hospitals in the period 2011-2015.



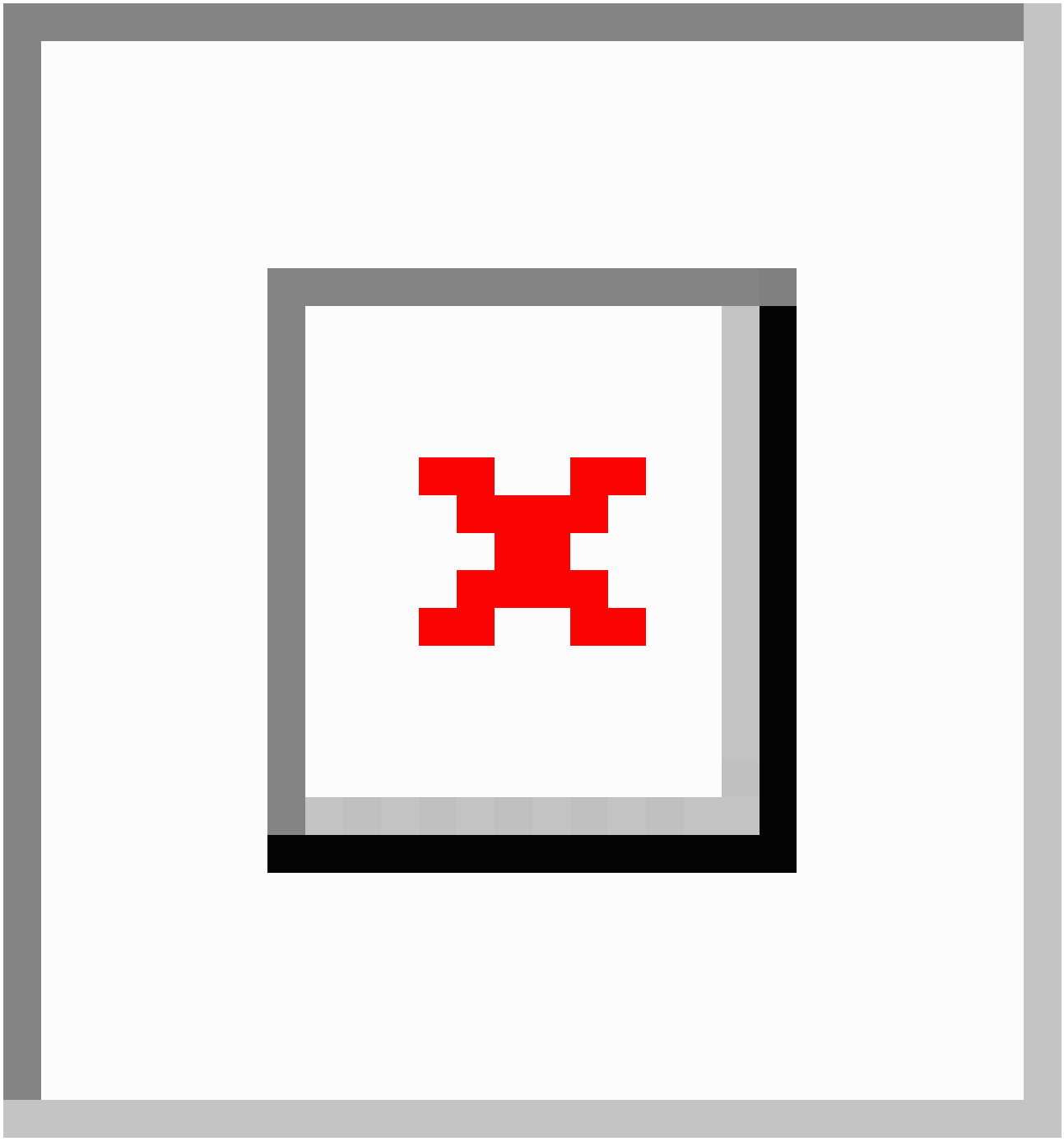
XSL•FO
RenderX

Figure 2. Evolution of social media profiles of public and private hospitals from 2011 to 2015. Facebook (FB), Twitter (TW), and YouTube (YT) and the type of hospital are represented using a private or public token (eg, FB Private stands for private hospital Facebook profiles).



XSL•FO
RenderX

Figure 3. Evolution of use of Twitter (left y-axis) and YouTube (right y-axis) among private and public hospitals in the period 2011-2015.



Use Dimension

Where analysis of content generation by hospitals in social media (Figure 3) is concerned, different behavior depending on social media and the type of entity is evident. Both public and private hospitals seem to have the same activity in Twitter, whereas private hospitals double the use of YouTube with respect to the public sector.

Impact Dimension

The evolution in the number of subscribers for the studied social networks for both private (dotted line) and public (continuous line) hospitals is exponentially increasing (Figure 4). All the profiles show continuous incremental growth, similar to an exponential function. The coupled analysis of the subscribers shows that Facebook is the most popular network, followed by Twitter. YouTube is, by far, the social network with the lowest subscriptions. Analysis of ownership shows that subscriptions to Facebook private hospitals profiles are significantly higher than those for public hospitals. With respect to Twitter subscribers, public hospitals have almost the same number as private and show a very similar trend of growth over the years. This growth is proportional to the number of tweets published by the hospitals (Figure 3). Although the number of Twitter accounts of private hospitals is greater than public ones, the number of people subscribed to private hospitals has been historically lower than those subscribed to public ones, until January 2015, when the trend reversed. Subscriptions to

XSL•FO
RenderX

YouTube are dramatically different between private to public hospitals.

As opposed to our findings regarding Twitter, analysis of the evolution of Facebook profiles (Figure 2) compared with the subscriptions (Figure 4, deep blue lines) demonstrates that the acceptance of private hospitals profiles is greater than public ones. From mid-2011, with a similar number of Facebook accounts, the number of subscriptions for private hospitals is dramatically higher than for public hospitals. Regarding videos (Figure 3), before 2012, public hospitals published more videos than private hospitals. However, since the end of 2011, the number of videos posted by private hospitals has increased more rapidly.

The position and magnitude attributes of observed variables in Table 1 for each social network separated show a skewed distribution. Kurtosis values (>3) confirm the leptokurtic distribution of all the variables. A similar study in the United States also reported skewed distribution of social media metrics [26]. In this case, statistical analysis was conducted by approximating the skewed distribution to a normal distribution with a lognormal transformation. Our approach is to be as realistic as possible without transforming raw data and using other kinds of statistical tests that can work with nonparametric distributions, such as the Wilcoxon signed-rank test [27]. In Figure 5, we show an example of the raw Twitter follower distribution for private hospitals (histogram) and superimposed the approximation to a normal distribution by applying a lognormal transformation (red line) and a nonparametric distribution (blue line). The goodness of fit or the nonparametric approach is more realistic than the lognormal transformation, and thus, closer to the actual real values of the distribution.

Table 1. Magnitude of the observed variables by means of median values, interquartile range, and kurtosis values for public and private hospitals with regard to December 2015.

Variables	Private hospitals			Public hospitals		
	Median	IQR[a]	K	median	IQR	K
Facebook friends (n=266,311)	1253	784.25-2306.75	4.329	740	341.25-2092.25	63.880
Tweets (n=250,040)	1244	641-2780	8.801	750	241-1863	15.565
Twitter followers (n=172,691)	895	502-1576	12.243	437.5	206-973	9.184
YouTube videos (n=4269)	19	5.25-72.25	7.760	19	15.4-62	15.982
YouTube subscribers (n=59,506)	25	4.25-113.5	5.493	12	4-56.75	58.464
YouTube video replays (n=20,488,992)	7386	1530-63,030	14.440	5908	1722.75-38,318.25	50.3166

[a]IQR: interquartile range.

Table 2 shows the results of the independent associations test between public and private hospitals for the observed raw magnitudes, by comparing the nonparametric distribution analysis (two-tailored Wilcoxon test , alpha=.05) and the lognormal approximation to a normal distribution (two-tailored *t* student test, alpha=.05).

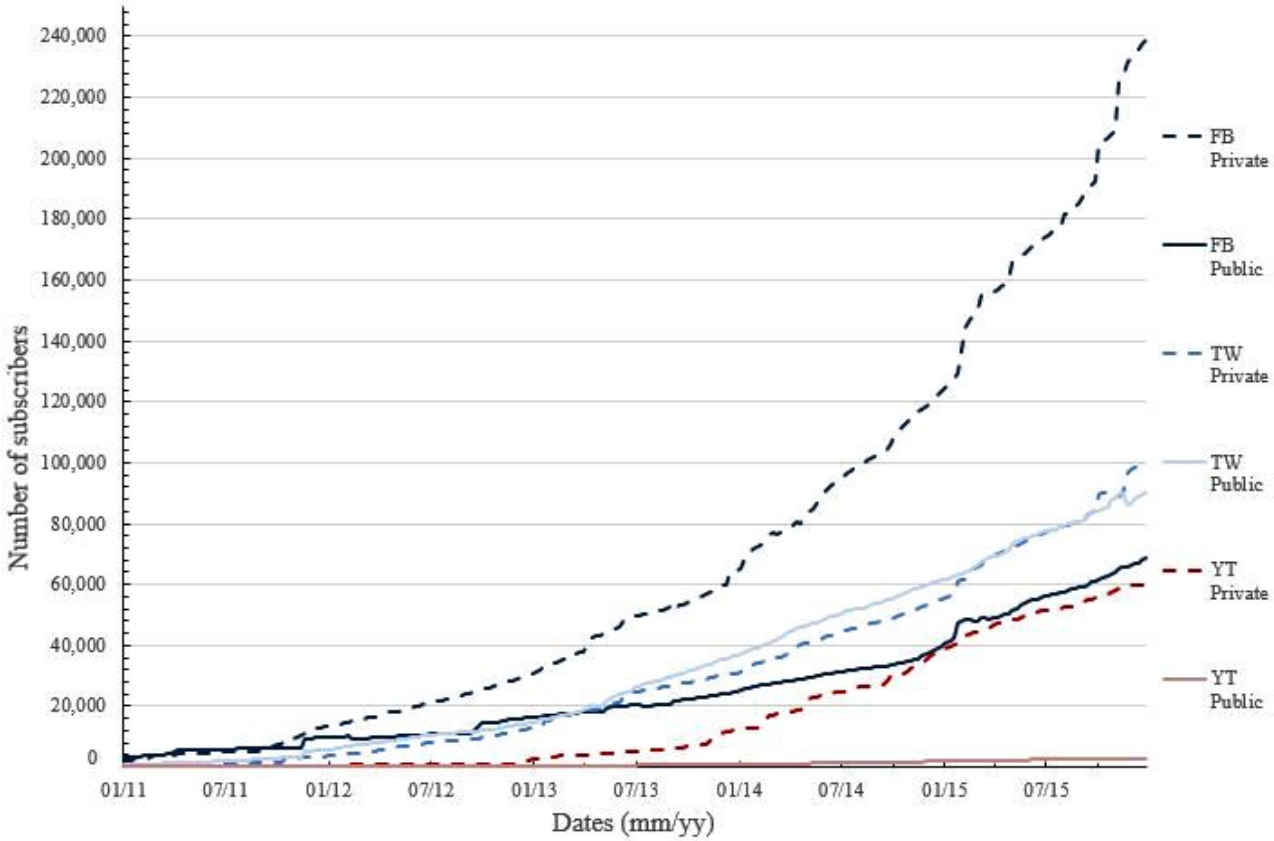
XSL•FO
RenderX

Table 2. Comparative table of the two-sided *t* student test and Wilcoxon test at a 95% CI for private and public hospital comparison on each of the observed variables.

Variables	Lognormal transformation		Nonparametric distribution	
	P	CI	*P*	*z*
Facebook friends (n=266,311)	*<.001* [a]	25.277-32.399	.07	1.804
Tweets (n=250,040)	*<.001*	5.728-9.295	*<.001*	2.542
Twitter followers (n=172,691)	*<.001*	4.090-5.220	*<.001*	3.503
YouTube videos (n=4269)	.20	.47	0.709	−0.232 to 1.084
YouTube subscribers (n=59,506)	*.03*	0.048-1.170	.51	0.647
YouTube video replays (n=20,488,992)	*<.001*	4.8474-6.5396	.87	−0.15849

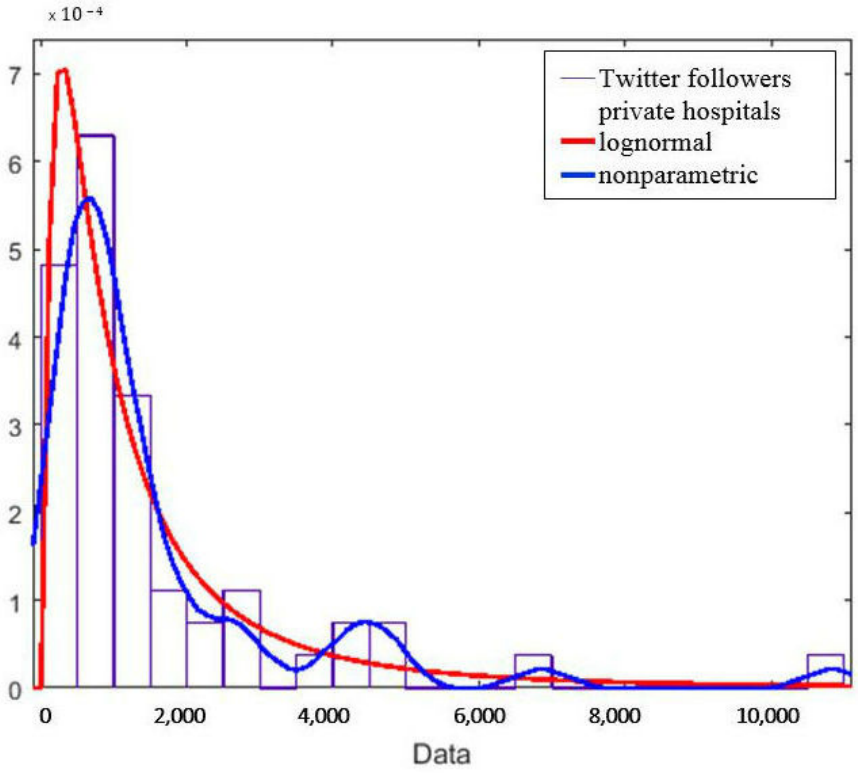
[a]Statistically significant values are given in italics.

Figure 4. Impact evolution of private and public hospitals in social media. Facebook (FB), Twitter (TW), and YouTube (YT) and the type of hospital are represented using a private or public token (eg, FB Private stands for private hospital Facebook profiles).



XSL•FO
RenderX

Figure 5. Twitter followers for private hospitals' distribution (rectangles) fitting comparison between a lognormal distribution (red line) and a nonparametric distribution (blue line).



Discussion

Principal Findings

The Internet presence of Spanish hospitals is high (550/787, 69.9%); nonetheless, the presence on the 3 main social networks is not as high (269/787, 34.2%) when compared with results of previous studies in the United States and Western Europe—which did not include Spain (Table 3).

Hospitals in Spain are adopting different types of social networks and using them to disseminate content. Considering all the hospitals with presence on social media, we find that Twitter had the most profiles, followed by Facebook and YouTube. The comparison of presence, use, and impact dimensions between public and private hospitals shows that private hospitals have a better performance and growth on Facebook and YouTube, whereas the growth on Twitter is similar for both public and private hospitals.

Table 3. Comparative table of presence among social media of the outcomes of the study and literature.

Social media presence	United States [26]		Western Europe [23]		Spain	
	2010	2014	2011	2014	2011	2014
	%	%	%	%	%	%
Facebook	18	90	67	-	4	10
Twitter	16	40	40	-	3	12
YouTube	-	-	19	-	3	7

The evolution of the presence, use, and impact of Spanish hospitals on social media is of increasing importance lately. Although the presence of Spanish hospitals on social media is low when compared with other countries, results of the analysis on presence, use, and impact of hospitals profiles on social media show that usage is constantly increasing. Since 2011, the presence of private hospitals on the Internet has grown exponentially. Private hospitals have deeper penetration on Facebook and YouTube than public hospitals, especially in the case of YouTube, where the number of subscribers is three times that of public hospitals. These results contrast with the fact that Spanish citizens prefer public hospitals rather than private [5].

Spanish citizens prefer public hospitals, but, could social media campaigns affect this perception depending of the type of social media? The answer to this question may rely on the nature of each social media and their main type of users. Twitter users look for an easy and quick way to gather information. It may be the case that Spanish social media users prefer tweets from public hospitals due to the high confidence of citizens in public health systems (they are more likely to post health content than advertisements) whereas Facebook and YouTube users look for interesting multimedia content. The results of this study show that despite Spanish citizens having an inferior impression of private hospitals, they achieve better metrics on social media

XSL•FO
RenderX

than public hospitals. This might mean that marketing campaigns of private hospitals and the higher investment devoted to generating more attractive multimedia material than public hospitals have a positive impact on users. In fact, *Son Espases* [28] and *Sant Joan de Deu* are two hospitals that have proved that marketing campaigns can be very effective in increasing the presence of hospitals on social media. The case of *Son Espases* hospital is very interesting. This hospital was inaugurated in 2011 with a massive marketing campaign carried out during its construction by a social media company. As a result, when this hospital was inaugurated and included on the NHC, the number of followers was significantly high, accounting for almost 32% (3521) of Twitter subscribers of public hospitals at that time. However, after the inauguration, *Son Espases '* activity on social media dropped, most probably because the social media company was no longer contracted. This abrupt change correlated with the large decrease in tweets produced at that time. From Figure 2, it can be seen that after the inauguration, Twitter activity of the public hospital maintained moderate growth, while private hospitals saw an increased slope. This suggests how a good media campaign may affect the visibility of a hospital, but also reveals that these activities need to be sustained for the long term.

Similarly, Facebook and YouTube are social platforms that allow hospitals and users to hold conversations (post and comments) in a very different way from Twitter. A previous study suggests that dialogue between hospitals and users (even patients) on these social networks may be a good source of information regarding service quality [29].

If we look at 2015, only 30.5% (96/314) of public and 36.6% (173/473) of private hospitals are involved in at least one of the observed social networks. This shows that hospital management and marketing teams are not aware of the opportunity that social media provide as an effective communication channel. The ability to respond in real time to users and give information on special situations or health campaigns through social media provides a new method of collecting data and assessing the quality of service, which is faster than traditional phone and onsite surveys.

Statistical analysis of the variables of private and public hospitals observed in the study, including Facebook friends, Twitter followers, number of tweets and YouTube videos, subscribers, and replays was performed in two stages. In the first stage, due to the leptokurtic distribution of all the variables (Table 1), a lognormal transformation to approximate a normal distribution was made, following the approach used by other authors [26]. The two-sided *t* student test with a CI of 95% showed that Facebook friends, Twitter followers, and tweets and YouTube video replays reveal a statistically significant different behavior between private and public hospitals ($P<.001$), whereas YouTube videos and subscribers showed a different, nonsignificant behavior. These findings confirm the results of Griffis and colleagues [26]. In the second stage, instead of transforming data, we used a nonparametric distribution to approximate the probability density function and a two-sided Wilcoxon test with a CI of 95%. This time, results show that only Twitter followers and tweets have a statistically significant different behavior ($P<.001$), whereas the other observed variables had a nonsignificant different behavior. The reason for this difference may be that for the second-stage analysis (nonparametric), raw data was used without performing any transformation, which may have led to bias in the data toward a certain direction. Nonetheless, the two-checked statistical significance of different behaviors in the use of Twitter for private and public hospitals confirms the hypothesis of our study.

This study focuses on analyzing the different trends and behaviors that public and private hospitals have in the use of social media, but the statistical tools used to pursue this analysis are a critical issue. This paper suggests using nonparametric techniques such as Wilcoxon test, used before in other scientific studies [27], which performs a better approximation to real data than the lognormal approximation, but the authors cannot confirm this extent.

The research question analyzed in this study is relevant in the era of social communication. Hospitals should have a strong presence on social media just as other entities and corporations do, as they provide an extraordinary and innovative instantaneous communication channel that reaches a wide audience. Unlike traditional communication campaigns, social media allow the release of information in new media formats (infographics, hashtags, audio, and video), adding formal and informal messages (eg, use of smileys) in very short time periods. Another singular characteristic is that hospitals can instantly assess the impact that the message or communication campaign has regarding new subscribers, retweets, or likes. Our study confirms that the performance on Twitter is different between private and public hospitals (Table 2), whereas at the same time, we observe a close evolution in the number of subscribers for both public and private hospitals (Figure 4). The cause may depend on the type of subscribers (age, profession, or interests) and in the type of broadcasted content (promotion, advertisement, or awareness).

When analyzing the performance of hospitals on social media, the risks as well as the ethical aspects of the use of social media in health care should also be considered. It is an important issue that is beyond the scope of our study. Scientific literature has addressed these issues in recent years [30], and several health care and health professional organizations have defined rules and policies on the use of social media. A good example is the American Medical Association (United States) policy, and more specifically, the Use and Style Guidelines on Social Networks published by the Regional Healthcare Agency of Andalucia (Spain) [31]. In spite of the multiple benefits of social media, there are significant risks that should be taken into account, mainly concerning security and privacy of the users. Broadcasters (hospitals) and consumers (patients and citizens) are not usually aware of the size of the audience they reach when a comment or an opinion is posted on any of the social media. The convenience of communicating with digital friends may lead users to publish harmful or inappropriate material that may affect their reputation and which may be very difficult (not to say impossible) to erase. Therefore, the use of social media by hospitals should adhere to a high level of compliance to published guidelines and rules from relevant organizations.

XSL•FO
RenderX

Limitations

Our findings are based on the entire population of public and private hospitals in Spain; nonetheless, even though the data collection method was based on previous publications, the authors cannot guarantee the location of all the webpages and social media profiles of the hospitals in the NHC. The fact that social media are constantly and rapidly changing affects the way data are collected and analyzed. Finally, some hospitals are outliers in the way that their performance shows a comparatively higher use and impact on social media than others, regardless of being public or private, and this could need further analysis, which is out of the scope of this study.

Future Work

Future work will tackle analysis of the subscribers' public profile and each profile on each social media, and among them, which type of subscribers are more likely to interact with posted content (likes, comments, shares) and which type of information has greater or lesser impact.

Conclusions

The presence of Spanish hospitals on social media is constantly evolving, showing an incremental growth year by year; however, it is very low compared with hospitals in the United States and Western Europe. Public hospitals are more active on Twitter, whereas private hospitals have a better performance on Facebook and YouTube. The Spanish health care system needs to maintain a high-ranking position, and to do so, this study suggests that hospitals, both public and private, should devote more effort to and be more aware of social media. The study conclusion is that private hospitals and public hospitals show statistically significant different behaviors in their use of Twitter (number of tweets and number of followers).

Acknowledgments

The authors wish to acknowledge the ITACA Institute (Universitat Politècnica de València) for making possible the publication of this paper through the Excellence Support program for the publication in high-impact international journals.

Conflicts of Interest

None declared.

References

1. Drösler SE, Romano PS, Tancredi DJ, Klazinga NS. International comparability of patient safety indicators in 15 OECD member countries: a methodological approach of adjustment by secondary diagnoses. Health Serv Res 2012 Feb;47(1 Pt 1):275-292 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01290.x] [Medline: 21762143]
2. Chatal O Health status determinants: lifestyle, environment, health care resources and efficiency, OECD economics department working papers, no. In: Health Care Systems: Efficiency and Policy Settings. Paris: OECD Publishing; 2008.
3. Deaton A. Income, health, and well-being around the world: evidence from the Gallup World Poll. J Econ Perspect 2008 Mar;22(2):53-72 [FREE Full text] [doi: 10.1257/jep.22.2.53] [Medline: 19436768]
4. Spanish Government Official Sociological Research Center (CIS) Health Barometer 2015. Gob. Madrid: Ministerio de Sanidad; 2016. Ministerio de Sanidad, Servicios Sociales e Igualdad, Secretaría General de Sanidad y Consumo URL: http://msssi.gob.es/eu/estadEstudios/estadisticas/BarometroSanitario/home_BS.htm [accessed 2017-05-06] [WebCite Cache ID 6qGipC5Oi]
5. Fundación IDIS. Primer Barómetro de la sanidad privada URL: http://www.fundacionidis.com/wp-content/uploads/2012/05/InformeIDIS_Barometro.pdf [accessed 2017-03-08] [WebCite Cache ID 6oowSChtL]
6. Murugesan S. Understanding web 2.0. IT Prof 2007 Jul;9(4):34-41. [doi: 10.1109/MITP.2007.78]
7. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012;14(3):e74 [FREE Full text] [doi: 10.2196/jmir.1943] [Medline: 22672889]
8. 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]
9. Kummervold PE, Chronaki CE, Lausen B, Prokosch H, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
10. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
11. van Uden-Kraan C F, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
12. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]

XSL•FO
RenderX

13. Callis KL, Christ LR, Resasco J, Armitage DW, Ash JD, Caughlin TT, et al. Improving wikipedia: educational opportunity and professional responsibility. Trends Ecol Evol 2009 Apr;24(4):177-179. [doi: 10.1016/j.tree.2009.01.003] [Medline: 19269059]
14. Friedlin J, McDonald CJ. An evaluation of medical knowledge contained in wikipedia and its use in the LOINC database. J Am Med Inform Assoc 2010 May;17(3):283-287 [FREE Full text] [doi: 10.1136/jamia.2009.001180] [Medline: 20442145]
15. Giles J. Internet encyclopaedias go head to head. Nature 2005 Dec 15;438(7070):900-901. [doi: 10.1038/438900a] [Medline: 16355180]
16. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
17. Thaker SI, Nowacki AS, Mehta NB, Edwards AR. How U.S. hospitals use social media. Ann Intern Med 2011 May 17;154(10):707-708. [doi: 10.7326/0003-4819-154-10-201105170-00021] [Medline: 21576547]
18. Ogilvy Public Relations. Using social media platforms to amplify public health messaging URL: http://socialchange.ogilvypr.com/using-social-media-platforms-to-amplify-public-health-messaging/ [accessed 2017-03-08] [WebCite Cache ID 6oowffMTu]
19. Wiener L, Crum C, Grady C, Merchant M. To friend or not to friend: the use of social media in clinical oncology. J Oncol Pract 2011 Dec 27;8(2):103-106. [doi: 10.1200/JOP.2011.000357]
20. Kamel BM, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
21. Clark C. Health Leaders Media. Tweet smears hospital's discharge incentive campaign URL: http://www.healthleadersmedia.com/quality/tweet-smears-hospitals-discharge-incentive-campaign [accessed 2017-03-08] [WebCite Cache ID 6oowthEhF]
22. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
23. Van de Belt TH, Berben SA, Samsom M, Engelen LJ, Schoonhoven L. Use of social media by Western European hospitals: longitudinal study. J Med Internet Res 2012;14(3):e61 [FREE Full text] [doi: 10.2196/jmir.1992] [Medline: 22549016]
24. Mathworks. MatlabTM URL: http://www.mathworks.com/ [accessed 2017-03-08] [WebCite Cache ID 6oox5uEQD]
25. Gob. 2015. National Hospitals Catalog URL: https://www.msssi.gob.es/ciudadanos/prestaciones/centrosServiciosSNS/hospitales/docs/CNH2016.pdf [accessed 2017-05-06] [WebCite Cache ID 6qGjGKBze]
26. Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. J Med Internet Res 2014;16(11):e264 [FREE Full text] [doi: 10.2196/jmir.3758] [Medline: 25431831]
27. Rosner B, Glynn RJ, Lee MT. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 2006 Mar;62(1):185-192. [doi: 10.1111/j.1541-0420.2005.00389.x] [Medline: 16542245]
28. Official Web of Son Espases Hospital. URL: http://www.hospitalsonespases.es/index.php?lang=es [accessed 2017-05-06] [WebCite Cache ID 6qGj1qNLN]
29. Timian A, Rupcic S, Kachnowski S, Luisi P. Do patients “like” good care? measuring hospital quality via facebook. Am J Med Qual 2013;28(5):374-382. [doi: 10.1177/1062860612474839] [Medline: 23378059]
30. Gholami-Kordkheili F, Wild V, Strech D. The impact of social media on medical professionalism: a systematic qualitative review of challenges and opportunities. J Med Internet Res 2013 Aug;15(8):e184 [FREE Full text] [doi: 10.2196/jmir.2708] [Medline: 23985172]
31. Redsaludandalucia. Guidelines on use and style in social networks from the Regional healtchcare system of Andalucia URL: http://www.redsaludandalucia.es/sites/default/files/guiav1.pdf [accessed 2017-03-08] [WebCite Cache ID 6ooxXtmrE]

Abbreviations

FB: Facebook
IQR: interquartile range
NHC: National Hospitals Catalog
TW: Twitter
YT: YouTube

XSL•FO
RenderX

Edited by G Eysenbach; submitted 07.10.16; peer-reviewed by G Giunti, S Kachnowski, I de la Torre, E Gabarron; comments to author 04.01.17; revised version received 26.01.17; accepted 27.03.17; published 23.05.17.

Please cite as:
Martinez-Millana A, Fernandez-Llatas C, Basagoiti Bilbao I, Traver Salcedo M, Traver Salcedo V
Evaluating the Social Media Performance of Hospitals in Spain: A Longitudinal and Comparative Study
J Med Internet Res 2017;19(5):e181
URL: http://www.jmir.org/2017/5/e181/
doi:10.2196/jmir.6763
PMID:28536091

©Antonio Martinez-Millana, Carlos Fernandez-Llatas, Ignacio Basagoiti Bilbao, Manuel Traver Salcedo, Vicente Traver Salcedo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Shifting Practices Toward Recovery-Oriented Care Through an E-Recovery Portal in Community Mental Health Care: A Mixed-Methods Exploratory Study

Deede Gammon[1,2], PhD; Monica Strand[1,3,4], MSc; Lillian Sofie Eng[1]; Elin Børøsund[1], PhD; Cecilie Varsi[1], PhD; Cornelia Ruland[1,4], PhD

[1]Center for Shared Decision-Making and Collaborative Care Research, Oslo University Hospital, Oslo, Norway

[2]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[3]Department of Psychiatry Blakstad, Division of Mental Health and Addiction, Vestre Viken Hospital Trust, Asker, Norway

[4]University of Oslo, Faculty of Medicine, Oslo, Norway

Corresponding Author:
Deede Gammon, PhD
Center for Shared Decision-Making and Collaborative Care Research
Oslo University Hospital
Klaus Thorgårdsvei 3, 3rd Floor
Oslo, N-0377
Norway
Phone: 47 90977963
Fax: 47 23075450
Email: deede.gammon@rr-research.no

Abstract

Background: Mental health care is shifting from a primary focus on symptom reduction toward personal recovery-oriented care, especially for persons with long-term mental health care needs. Web-based portals may facilitate this shift, but little is known about how such tools are used or the role they may play in personal recovery.

Objective: The aim was to illustrate uses and experiences with the secure e-recovery portal "ReConnect" as an adjunct to ongoing community mental health care and explore its potential role in shifting practices toward recovery.

Methods: ReConnect was introduced into two Norwegian mental health care communities and used for 6 months. The aim was to support personal recovery and collaboration between service users and health care providers. Among inclusion criteria for participation were long-term care needs and at least one provider willing to interact with service users through ReConnect. The portal was designed to support ongoing collaboration as each service user-provider dyad/team found appropriate and consisted of (1) a toolbox of resources for articulating and working with recovery processes, such as status/goals/activities relative to life domains (eg, employment, social network, health), medications, network map, and exercises (eg, sleep hygiene, mindfulness); (2) messaging with providers who had partial access to toolbox content; and (3) a peer support forum. Quantitative data (ie, system log, questionnaires) were analyzed using descriptive statistics. Qualitative data (eg, focus groups, forum postings) are presented relative to four recovery-oriented practice domains: personally defined recovery, promoting citizenship, working relationships, and organizational commitment.

Results: Fifty-six participants (29 service users and 27 providers) made up 29 service user-provider dyads. Service users reported having 11 different mental health diagnoses, with a median 2 (range 1-7) diagnoses each. The 27 providers represented nine different professional backgrounds. The forum was the most frequently used module with 1870 visits and 542 postings. Service users' control over toolbox resources (eg, defining and working toward personal goals), coupled with peer support, activated service users in their personal recovery processes and in community engagement. Some providers (30%, 8/27) did not interact with service users through ReConnect. Dyads that used the portal resources did so in highly diverse ways, and participants reported needing more than 6 months to discover and adapt optimal uses relative to their individual and collaborative needs.

Conclusions: Regardless of providers' portal use, service users' control over toolbox resources, coupled with peer support, offered an empowering common frame of reference that represented a shift toward recovery-oriented practices within communities. Although service users' autonomous use of the portal can eventually influence providers in the direction of recovery practices,

XSL•FO
RenderX

a fundamental shift is unlikely without broader organizational commitments aligned with recovery principles (eg, quantified goals for service user involvement in care plans).

(J Med Internet Res 2017;19(5):e145) doi:10.2196/jmir.7524

KEYWORDS

recovery; eHealth; mental health; psychiatry; user involvement; empowerment; working relationships; participatory research; secure email; e-recovery

Introduction

Mental health care policies for those with long-term care needs are shifting from a primary focus on symptom reduction toward partnership models and personal recovery-oriented care [1]. At the same time, public health policies are promoting broad-scale implementation of eHealth technologies to strengthen people-centered care and public health capacity [2]. Several developments in eHealth are relevant to recovery-oriented care in mental health, such as enabling people access to their own electronic health records [3], shared decision making [4], self-management [5,6], peer support [7,8], online patient-reported outcomes [9], and service user involvement in research [10,11].

A common denominator of these developments is a shift in "locus of control" from health care providers toward service users by increasing the transparency of care decisions, as well as facilitating the voice and resources of service users in their care. In contrast to biomedical approaches that focus mainly on reducing symptoms, recovery-oriented approaches support people in articulating and regaining control over progress toward personal well-being goals [1,12,13]. Conceptualizations of the holistic and multifaceted nature of recovery are evolving in interaction with related fields such as self-determination and strength-based approaches [14,15], and is sometimes referred to as paradigmatic in that it disrupts established practice norms, priorities, and professional skill sets [16,17]. Accompanying emerging frameworks and guidelines for recovery-oriented practices are efforts to identify meaningful outcome measures across cultures and contexts [18-20]. Considerable work still lies ahead in identifying active ingredients of recovery, for whom, and under what conditions [15,21].

It is within this evolving landscape that this study describes the use of a recovery-oriented eHealth ("e-recovery") portal "ReConnect" in two Norwegian community mental health sites during a 6-month period (2015-2016). ReConnect was designed using participatory methods. The rationales for portal design, including our path toward recovery as the guiding framework, are described elsewhere [22] (note that ReConnect was called "PsyConnect" in this previous publication). In this study, we sought insights into the question: what uses evolve when an e-recovery portal is made available in community mental health practices and what role does it play in terms of shifting practices toward recovery-oriented care?

Methods

Design

This study had a participatory design [23] and used mixed methods [24] to explore uses of an Internet-based intervention designed to support recovery-oriented practices in mental health care for people in need of long-term mental health care. The intervention was studied for a period of 6 months in two separate communities. Heeding calls for service user involvement in research [25,26], this study was conducted in collaboration with a service user consultant.

Setting

Norway has universal health care that is funded by the public as part of the through general and earmarked grants [27]. The municipalities are responsible for providing primary health care and social services, whereas the Regional Health Authorities provide specialist services (eg, acute wards, district psychiatric centers). "Communities" in this paper refer to both levels of care provided to residents of two municipalities in Norway. These communities differed in characteristics, thus providing an intended variation in context: a small rural community with approximately 5500 inhabitants within an area of 1493 km^2 versus a large community on the outskirts of the capital with approximately 52,000 inhabitants within an area of 100 km^2. Management in the two communities expressed commitments to policies promoting eHealth, user involvement, and collaborative practices. The largest community explicitly expressed commitments to recovery principles in policy and strategy documents [28].

Inclusion Criteria and Recruitment

The two communities became involved in the project through prior contacts with the principal investigator (DG). Multiple service entities at primary and specialist levels of care, as well as local service user organizations within the two communities, received written information and verbal presentations about the study. This information included the project's overall aim of gaining insights into user needs and how e-recovery might facilitate or undermine service user involvement in treatment and collaboration with providers. Service users interested in participating in the study needed to fulfill the following criteria: they had to be older than 18 years; had to have received mental health services for at least 6 months prior to inclusion; and needed expectations of requiring services at least 6 months forward, electronic ID (see subsequently), and at least one provider willing to interact with them through ReConnect. As an exploratory study, efforts were made to recruit a wide range

XSL•FO
RenderX

of participants in terms of gender, age, health issues, and types of ongoing support.

Ethics

The study was approved by the Regional Committees for Medical and Health Research Ethics in Norway and the Privacy Protection Committees at the participating sites. Participants signed an online consent form before inclusion in the study.

Organizational Anchoring

Local steering committees were established in both communities and consisted of primary and secondary health representatives (both clinicians and authorities), information technology (IT) management, and service user representatives. Their mandate was to ensure access to necessary resources (eg, clinician time, IT support, localities for training), and that the project harmonized with ongoing activities. Two hours of group training and/or individual training were held within both communities for service users and providers initially and when requested during the study period.

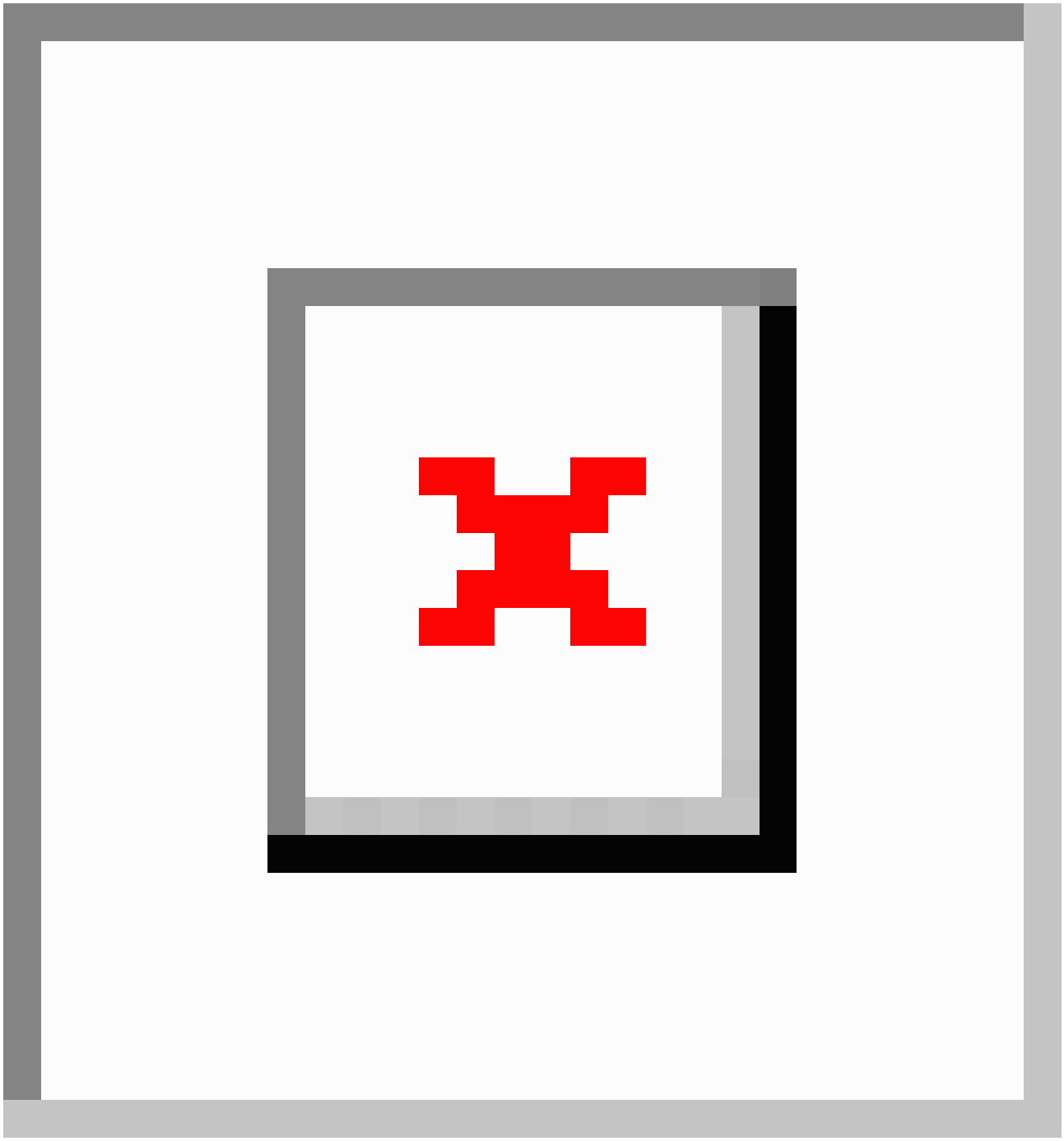
The E-Recovery Intervention

ReConnect was introduced to participants to support ongoing mental health care and treatment—whatever that treatment may be and as they saw fit. The stated objective of the portal was to support service user involvement in care, service user-provider collaboration, and personal recovery.

As depicted in Figure 1, the portal consists of a toolbox, anonymous peer support discussion forum, and messaging with providers. Users log on using their electronic ID (eg, BankID), which is approved by the Norwegian government to allow patients to share personal health information in electronic and mobile apps. This ID is the same whether users log on to public services or online banking and is thus familiar to most Norwegians.

XSL•FO
RenderX

Figure 1. ReConnect portal.



Toolbox

The toolbox can be likened to an interactive "workbook" and offers a wide range of resources that support service users in articulating and working with their personal recovery processes, including life domain status (eg, employment, social network, health), goals and activities, medications, network map, crisis plan, diary, and exercises (eg, sleep hygiene, mindfulness). Information related to patient rights and organizations were accessible either in the portal itself or through links. Simple help texts were available in all modules, as well as "good-to-know" texts (eg, how to formulate meaningful goals). Service users "owned" ReConnect in the sense that they determined how to use it to articulate aspects of "my life" and decide which providers had access to the user-generated content.

Forum

The anonymous peer-to-peer forum with service users from the two communities was moderated by LSE to ensure a safe and supportive environment.

Cafés

There were local real-life "ReConnect cafés" where service users could meet socially to discuss their uses of the portal and their own recovery processes.

XSL•FO
RenderX

Collaboration

Services users' interactions with providers through ReConnect took place by messaging, by providers remotely accessing and commenting on the content of service users' modules, or by sitting together and working with modules during consultations. The providers' user interface included an overview of all their clients who used ReConnect, and they could remotely access the service users' modules with some exceptions (eg, diary and forum). Thus, providers could follow the progression of service users' activities (eg, homework in between consultations) and provide feedback as they saw fit.

Service users consented to using ReConnect exclusively for nonemergency purposes, and that ordinary channels had to be used for acute needs. Other than that, collaborative uses of ReConnect were determined by each respective service user-provider dyad as described previously. These were encouraged to clarify mutual expectations and routines, such as response time for messages (eg, daily, or once a week), absences (eg, holidays), and types of content (eg, providers might acknowledge receipt of messages with brief responses, but reserve therapeutic responses for consultations).

Quantitative Data Collection and Analysis

After online registration and completion of the consent form, participants completed an online questionnaire containing questions about demographic characteristics, previous use of the Internet, and the following psychosocial measures:

1. Well-being was measured with the WHO-5 Well-being Index. The WHO-5 score can range between 0 to 100, in which 100 indicates the best possible well-being [29].
2. Anxiety and depression were measured with The Hopkins Symptom Checklist-25 (HSCL-25). The total HSCL-25 score can range between 1.0 and 4.0, and values greater than 1.75 indicate a need for help for the symptoms [30].
3. Patient activation was measured with the Patient Activation Measure (PAM). The score can range between 0 to100, and 100 indicates the best possible patient activation [31,32].
4. Satisfaction was measured with the Satisfaction Life Domains Questionnaire. The range of scores are 1 to 7, in which 7 indicates the best possible score [33].
5. Recovery was measured with the Recovery Assessment Scale (RAS). The range of scores are 1 to 5, and 5 indicates the highest recovery possible [34].

Data on portal use were extracted from the ReConnect user log on the server. All log-ins and uses per module were recorded in the system log.

Descriptive statistics were used to analyze questionnaire and system log data using SPSS version 21 (SPSS Inc, Chicago, IL, USA). Data are presented as medians and range for continuous variables due to skewed distributions and as proportion and percentages for categorical data. Demographics from the two communities are aggregated together to protect anonymity.

Qualitative Data Collection and Analysis

Six focus group interviews were held, three with service users and three with providers, separately at each community after approximately 3 months of use. The focus group interviews were conducted by MS and LSE together. Questions were semistructured addressing current practices and expectations, experiences related to use, recovery processes, collaboration, and desired changes in the portal. A total of 12 cafés (six in each community) were moderated by LSE who facilitated discussions about use of ReConnect and about recovery processes. Both the focus group interviews and café sessions lasted approximately 1.5 hours and were audiorecorded. All focus group interviews were transcribed, whereas cafés were transcribed only when LSE's notes indicated areas of interest to research analysis. Forum postings also served as data. Along with commenting on participant postings in the forum, LSE introduced topics for discussion relevant for recovery, such as hope, strengths, and citizenship. At times, authors DG, MS, and LSE discussed questions that LSE in turn posed to participants (eg, "Have any of you changed your way of collaborating with your provider through ReConnect and if so how?").

Publicly available documents (eg, minutes of meetings) and personal communications regarding stakeholders' actions following the study are referred to when reporting findings regarding the domain "organizational commitment" [35].

All transcripts from the focus groups, cafés, and forum were read through by authors MS, LSE, and DG, who added codes corresponding to the first two of Braun and Clarke's six stages thematic analysis approach [36]. For this paper, DG searched for codes and quotes that illustrated Le Boutillier et al's [35] four practice domains: (1) personally defined recovery, (2) promoting citizenship, (3) working relationships, and (4) organizational commitment. Because our aim was not to present themes based on an analysis of the total dataset (a focus of subsequent publications), but rather use the data to illustrate the practice domains, we present 3 to 5 illustrative quotes per practice domain. MS and LSE, who were most familiar with the informants and contexts from which quotes were extracted, reviewed the quotes in terms of credibility in illustrating the practice domains.

Results

Participants

Service User Characteristics

Of the 33 registered service users recruited, two withdrew right after recruitment. Twenty-nine answered the questionnaires and remained participants throughout the 6-month study period.

As presented in Table 1, participants had a median age of 44 (range 21-62) years, were predominantly female (86%, 25/29), single (69%, 20/29), and had an educational level of high school or less (69%, 20/29). In all, 31% (9/29) were employed either full-time or part-time, 28% (8/29) were on work assessment allowance, and 35% (10/29) were on disability benefits or retired. The service users reported a median of 2 (range 1-7) diagnoses (see Table 2 for elaboration on diagnosis).

XSL•FO
RenderX

Table 1. Demographic and illness characteristics among service users (N=29).

Characteristics	Participants
Age (years), median (range)	44 (21-62)
Gender, n (%)	
Female	25 (86)
Male	4 (14)
Marital status, n (%)	
Married/cohabitating	6 (21)
Divorced	3 (10)
Single	20 (69)
Education, n (%)	
Elementary/high school	20 (69)
University/college	9 (31)
Employment status, n (%)	
Full-time/part-time work	9 (31)
Work assessment allowance	8 (28)
Disability leave/retired	10 (35)
Sick leave	1 (3)
Student	1 (3)
Site, n (%)	
Small community (5500 inhabitants)	14 (48)
Large community (52,000 inhabitants)	15 (52)
Number of psychiatric diagnosis, median (range)[a]	2 (1-7)
Psychosocial factors, median (range)	
Well-being	44 (0-80)
Anxiety/depression	2.08 (1.24-3.68)
Patient activation	56.40 (32.20-100)
Satisfaction with life domains	4.11 (1.44-5.44)
Recovery, total score	3.67 (2.33-4.50)

[a] See Table 2 for list of diagnoses.

XSL•FO
RenderX

Table 2. Diagnoses reported by service users (N=29).

Reported diagnosis	n (%)
Depression	17 (59)
Panic anxiety	8 (28)
Generalized anxiety	8 (28)
Posttraumatic stress disorder	8 (28)
Phobic anxiety	5 (17)
Drug/alcohol addiction	4 (14)
Bipolar illness	4 (14)
Personality disorder	3 (10)
Schizophrenia	2 (7)
Obsessive-compulsive disorder	2 (7)
Schizoaffective illness	1 (3)
Mania	1 (3)
Other	4 (14)

The participants reported a median score of 44 (range 0-80) on the WHO-5 Well-being Index and a median score of 2.08 (range 1.24-3.68) on the HSCL-25 (anxiety and depression) indicating low well-being and an overall need for help with anxiety and depression symptoms (HSCL-25 cut off: 1.75). Their scores on patient activation, satisfaction, and recovery measures were in the middle of these scales, indicating room for improvements.

A total of 90% (26/29) used email daily or weekly, and 76% (22/29) used social media daily or weekly (see Multimedia Appendix 1 for more details of media use).

Health Care Provider Characteristics

Of the 27 participating health care providers, 14 worked in the municipalities and 13 worked in secondary level (DPC). They were predominantly women (89%, 24/27), 40 years or older (85%, 23/27), and most were nurses (11/27, 41%), social workers (5/27, 19%), and physicians (3/27, 11%). The remaining eight (28%) had different professions such as occupational therapist, psychologist, priest, interdisciplinary specialists, bachelor of psychology, or home care worker. There was a median of 19 (range 1-45) years since graduating from health professional education, and they had been working a median 10 (range 1-38) years within the field of mental health (Multimedia Appendix 2). All 27 providers used email and the majority used it daily (25/27, 93%) (see Multimedia Appendix 1 for more details on media use).

The dyads were highly diverse in terms of the diagnoses that service users reported and the professions reported by health care providers. Two service users had more than one participating provider.

Types and Frequencies of Use

The median number of log-ins was 17 (range 1-151) (Table 3). Median number of messages sent was 2 (range 0-43). Modules not used by most participants could be frequently used and valued by one or two participants. This was particularly the case for crisis plan, network map, the medication list, and the diary. Some reported that having the options was valued, even though they had not used them yet.

XSL•FO
RenderX

Table 3. Usage of different components and activities in ReConnect during 6 months of access among service users (N=29).

Components and activities	Median (range)
Number of log-ins	17 (1-151)
“Good-to-know” article visits	3 (0-12)
Read article	0 (0-9)
Number messages views	14 (0-93)
Messages sent	2 (0-43)
Messages received	3 (0-40)
Crisis plan created	0 (0-1)
Diary post entries	0 (0-51)
Exercise visits	0 (0-9)
Forum Visits	21 (0-364)
Forum posts	3 (0-149)
Forum treads views	33 (0-508)
Medicine visits	1 (0-5)
Network map visits	2 (0-10)
Plan visits	10 (0-45)
Activity plan creation	0 (0-18)
Goal plan creation	0 (0-15)
Sub goal plan creation	0 (0-14)
Registration visits	6 (0-109)
Registration create	2 (0-128)
Update life domains assessment	1 (0-22)

Collaborative Use

Of the 27 health care providers, 19 (70%) answered secure messages from the service users. They answered a median of 6 (range 1-27) messages. (The system log failed to register types and frequencies of provider’s access to their service user’s module; therefore, we are unable to report this.) Both service users and providers reported that 6 months was too short of time to learn and optimally adapt their use of the various toolbox resources to their individual and collaborative needs. Examples mentioned included discovering relevant exercises after learning from peers in the forum, and that optimal use of the portal could differ when health was in a good versus a bad phase.

Forum

The forum was visited a median 21 (range 0-364) times per service user during their respective 6-month participation periods. During these 6 months, the service users posted 542 postings and viewed forum posts 1870 times in total (data not shown). The peer-moderator (LSE) initiated 167 of 542 postings (30.8%). Ten service users were active posters (>10 posts). No postings had to be removed due to inappropriate content. One service user reported obsessive use of the forum and together with his/her provider found ways to control use.

Cafés

In the 12 face-to-face gatherings, a total of 17 service users participated (range 3-9 per meeting). Several reported that becoming secure in the forum had been a prerequisite for mustering the courage to participate in the face-to-face cafés.

Recovery-Oriented Practices

In the following, the experiences reported in focus groups, forum postings, and cafés by service users and providers are presented relative to Le Boutillier et al’s [35] four practice domains that were derived from a qualitative analysis of 30 international recovery-oriented practice guidance documents. These domains are summarized in Textbox 1. Although “practices” typically refers to actions taken by providers, we included the actions taken autonomously by service users through their use of ReConnect.

XSL•FO
RenderX

Textbox 1. Practice domains.

Supporting personally defined recovery

- Practitioners focus on supporting personally defined recovery heart of practice and not as an additional task. Individuals are supported to define their own needs, goals, dreams, and plans for the future to shape the content of care. Individuality, informed choice, peer support, strengths focus, and holistic approach are contained in this practice domain.

Promoting citizenship

- The core aim of services is to support people who live with mental illness to reintegrate into society and to live as equal citizens. Citizenship is central to supporting recovery, in which the right to a meaningful life for people living with severe and enduring mental illness is advocated. Seeing beyond "service user," service user rights, social inclusion, and meaningful occupation are grouped in this practice domain.

Working relationships

- Practitioner interactions demonstrate a genuine desire to support individuals and their families to fulfill their potential and to shape their own future. A therapeutic relationship, characterised as a partnership, is essential to supporting recovery in which hope is promoted.

Organizational commitment

- Organizations that support recovery orientation demonstrate a commitment to ensure that the work environment and service structure are conducive to promoting recovery-oriented practice. The organizational culture gives primacy to recovery and focuses on and adapts to the needs of people rather than those of services. Recovery vision, workplace support structures, quality improvement, care pathway, and workforce planning are included in this practice domain.

Supporting Personally Defined Recovery

Types of ReConnect uses that were particularly reflective of this practice domain were life domains, goal/activities, peer support, and the process of writing. Service users reported being helped in gaining an overview of their lives and becoming more conscientious of where they were headed and what kind of help they needed. For example, in a café discussion, one service user offered advice to another participant who was not getting the help he/she needed:

> *What I experienced as unbelievably positive for me was to sit down and divide up my life into the different life areas. It really increased my awareness. It became clearer for me where I stood, and where I wanted to head. I can really recommend it. You create for yourself a direction in life. At least that was my experience...[a lot of talking erupted in the group]...Maybe it would be easier for your helper to follow you up if she had something more concrete to work on...Maybe if you write it down it is easier for her to get a grip on what you need?*

Some providers shared this assessment of the same modules. As one provider stated in a focus group:

> *The goal module has really helped. When he/she says "I wish I'd do more of this," then I can put pressure on. When it's written down in there as a concrete goal, then it kind of lights up a fire of sorts.*

Another provider highlighted the value of service users' creating descriptions of life domains and goals/activities in their own words:

> *It's become a good way of structuring our work together. In a way, it's clearer. What is his/her assignment or expectation of me?...It ensures that it is in fact his/her goal and not something I've written. One might think it's the same, but the nuances in language can make a decisive difference in the actions we take.*

The peer-to-peer forum and café gatherings were used to share experiences with the exercises (eg, mindfulness, strengths, self-created exercises) in support of defining one's own direction. Among the many illustrative quotes in the forum:

> *I've just logged and made an exercise, I use all of them [exercises] except the ones for drug abuse. It's a nice support for me when I'm working with myself. And it HAS helped me. From being isolated and very depressed to now getting out more. It's helped to the point that I'm now working in a job 50%.*

Promoting Citizenship

The peer support activities in the forum and cafés can be viewed as promoting community involvement (citizenship) in and of itself. Initially, peer support was established and maintained through the forum and subsequently expanded on and enriched through both the café gatherings and the focus group interviews. Friendships developed and plans were made for getting involved (eg, volunteering) in local activities. This included reflections on the role that community involvement can play in promoting health, and that providers need to support service users in this process. An example from a café discussion, that also illustrates the next domain (working relationships), is the following:

> *My mental health gets better when I help others. Be useful, do something meaningful, contribute to community. Those are things that helps your health and recovery. How can we get our helpers to support us in that kind of thing?*

The issue that ignited the liveliest "community engagement" was at the end of the study when it was unclear whether ReConnect would be continued as a service within the two communities. This was evident in extensive forum discussions about how to influence community decision makers:

XSL•FO
RenderX

> *In any case, we've got to behave in the right way and talk to the people who are affected first, before we go to the newspaper, so we don't step on the wrong people's toes? But we can do this, right you guys? I hope you guys in [large community] are as enthusiastic as we [small community] are because we're pretty fired up about keeping this service (six smilies)*

A service user from the larger community responded:

> *If we're going to the newspapers, we need to have a positive angle—not that we're angry, or going to the barricades to fight, if we lose ReConnect. The smartest might be to go to the membership paper of the Norwegian Mental Health Association.*

Eventually, two service users from the largest community contacted one of the project's funding agencies who interviewed and photographed them for an article on their website.

Working Relationship

Dyad collaboration through ReConnect ranged from not at all to almost daily. This domain overlapped particularly with the first domain (supporting personally defined recovery) in that providers who supported service users in working with life domains, goals, and activities also reported having good working relationships. Collaborative uses included messaging, providers commenting the content of service users' modules, and/or by sitting together and working with modules during consultations.

The life domains and goal/activity modules were frequently referred to by service users as helping collaboration with providers become more focused on their needs. The types of goals reported were typically short term (eg, per week) and very concrete. As one service user reported in a focus group:

> *Earlier it's always been that [provider] asked me if I'd taken my medications, and then what openings there were in our calendars for my next consultation. Those two issues were what [provider] seemed mainly preoccupied with. Now with ReConnect we work more on my resources and goals—it can be as simple as managing to get through Christmas. How do I do it? Subgoals and activities can be buy the steak, avoid stress, get everything in the house, that type of thing—it was actually very useful to get ideas from another perspective—how to break down the problem...It really helps to break down the problem into smaller pieces.*

Some service users expressed frustration that providers repeatedly told them how busy they were as an excuse for why they had not worked with them through ReConnect. For example, in a café discussion:

> *Why did [provider] agree to work with me through this tool if she never expected to do it? She should have just said no. You get so disappointed. That's why it's good to have each other [forum participants]—to call you my helpers. So we can share things.*

This started a series of discussions about taking care not to overwhelm providers with messages or tasks, which caused one participant to react:

> *It's completely understandable that constantly hearing how busy your helper is—I mean you don't want to make life miserable for them. You don't. But it's just not right that us service users have to go around protecting our helpers.*

These types of discussions in forum, focus groups, and cafés were typically accompanied by constructive suggestions for how to positively engage providers. One such exchange took place in a café discussion:

> *You're right, it's important for them [providers] to see that they're useful to us—productive. The more specific we can be about what we need, the greater the chances that they'll respond to us and our needs.*
>
> *I think it was some smart advice from [another participant]. She gave her helper a clear assignment as to how to follow her up. I think several of us should do that. That's how we create communication.*

Some service users appreciated the flexibility that ReConnect introduced relative to in-person consultations that were sometimes described as unnecessary or unproductive. One service user, who received regular home visits, argued that flexibility could also benefit providers. As said in a focus group:

> *Maybe they don't have to come so often if we can contact them [through the portal] when we're working on something and need follow-up. Follow-up is what we need.*

Providers, on the other hand, expressed concerned about pressuring service users to use ReConnect in ways that could be an added burden on them. For example, one reported in a focus group interview:

> *I've heard my service user say, "Unfortunately I've not answered, or done it" ...sort of like they have to apologize for not doing it [used ReConnect]. That's why I'm kind of afraid of...it can be an extra burden on them...just following up things...Many are really vulnerable for stress.*

This coincided with several providers who reported not wanting to put pressure on service users to use ReConnect, but that they were available if service users took the initiative.

Organizational Commitment

Most providers told of being committed to user involvement in care (a key recovery principle), whereas several reported barriers to committing to use of ReConnect as an ordinary service. Technical infrastructure-related barriers included inconveniences of having to log in with their private electronic ID (due to lack of integration with secure log-in system used by health care), multiple overlapping systems, and lack of integration with electronic health records.

Leadership in the large municipality initiated processes to address infrastructure barriers with the intention to implement ReConnect as a permanent service (minutes of meetings). Both

XSL•FO
RenderX

the political and administrative leadership had committed to personal recovery—principles in all major policy and strategy documents [28]. This included a commitment to quantifying the extent of user involvement in individual care plans along with ambitious goals for an increase. ReConnect was viewed by leadership as enabling more effective progress toward policy goals (minutes of meetings). The smaller municipality also had user involvement as a goal, but without a specific approach or quantified goals. Here, the technical and financial commitments required to implement ReConnect were considered too great at the time.

Providers reported other barriers to committing to ReConnect. These included blurring lines between work and private life, lack of time allotted to answering messages, and concerns about the frequency and volume of written responses that might be expected by service users. Providers who appeared most positive toward ReConnect also reported being explicit about what service users could expect from them. One focus group participant, who described the portal as an asset to her work and benefit to service users, reported giving service users' clear expectations:

> *I've told my clients that I answer messages Monday and Thursday mornings. That's when they can expect answers from me. I need to have structure.*

Another satisfied provider reported making agreements with service users that they would only respond to service users' messages with brief responses to acknowledge receipt or clarify practical issues. More in-depth issues presented by service users would be acknowledged, then dealt with in their next consultation. Service users responded positively to these clarifications. Other providers valued saving time now that a service user had produced texts that could be taken directly into the statutory action plans. The fact that the service user also benefited from formulating and "owning" their own plans was referred to as "killing two birds with one stone."

Discussion

Principal Findings

This descriptive and exploratory study sought to illuminate the question: how is an e-recovery tool used as an adjunct to ongoing community mental health practices and what role can it play in shifting practices toward recovery-oriented care?

The service users who used the portal became more involved in activities reflecting the first two of Le Boutillier et al's [35] practice domains—personal recovery processes and citizenship—regardless of the practices of their provider. This was observable for the approximately 10 active forum posters and 17 café participants who also reported benefits similar to those reported in studies of online [37] and offline [38] peer support. Combining online and offline peer support with toolbox resources was an empowering common frame of reference for service users. Service users valued working more concretely on their personal life domains and goals, and in having a common vocabulary in discussing their experiences with peers. The opportunity to do so represented itself a shift toward recovery-oriented practices. Not only were service users offered a choice in terms of how they received mental health services, they could also choose to participate in defining their personal recovery processes and participate in community-promoting arenas. The service user who reported obsessive use of the forum, which was resolved together with his/her provider, was the only negative health-related experience reported among services users.

The positive role that the e-recovery portal played as a service separate from traditional services was highly dependent on the role played by the service user consultant (LSE) who moderated the forum and cafés. Although knowledge of optimal models for peer-run interventions is still evolving [12], communities who seek to promote recovery through similar portals will need to invest in similar types of expertise and role models for hope. Our experience suggests that the success of this role is closely linked with the acknowledgment of experiential knowledge as an asset within the community, in-depth familiarity of the principles of recovery, and the availability of discussion partners in health care when difficulties or dilemmas arise (LSE's experiences will be elaborated on elsewhere). When sufficiently supported, such consultants with "lived experience" can contribute to mobilizing resources among service users and communities in ways that also can be valuable for improving the quality of health care services [12].

Use of the portal to augment treatment and its role relative to working relationships (the third practice domain) was less obvious. Dyad diversity, along with the nondirective way in which ReConnect was introduced to dyads ("use it as you see fit"), was reflected in highly diverse uses of the various portal resources. A total of 30% of providers never initiated or responded to messages, a source of frustration for service users. After 6 months, both service users and providers reported they were still discovering resources in the portal and adapting uses to their needs and preferences. This may partly be due to the shift in locus of control in that service users' could now control the content of their own story and had a lowered threshold for linking documentation (eg, personal goals) to requests for follow-up. Both parties in working relationships can experience transitions of control as challenging [17,39], which likely adds to the time it takes to adjust.

Even if some dyads did not use the available resources in the portal to engage service users, the mere existence of the portal, and the dyads' agreements to use it, inserted the topic of control into service users' reports of their experiences in working with their provider. Some service users reported becoming empowered to make or request changes in the treatment they received, and that providers responded positively to these requests. However, such examples probably reflect good working relationships prior to use of ReConnect. Poor working relationships did not appear to improve through use of ReConnect, but rather were more clearly exposed as such. To explore how ReConnect can more systematically support working relationships in future studies, we have incorporated a short feedback-informed treatment measure to help dyads attend to the quality of their working relationship [40].

The largest community whose leadership had committed to recovery principles (ie, fourth practice domain) was also

XSL•FO
RenderX

prepared to address the infrastructure barriers to implementing ReConnect as a permanent service. This may reflect greater financial and political resources compared to the smaller community, who did not make implementation of ReConnect a priority. More importantly, however, the largest community viewed ReConnect as a means for more effectively reaching quantified policy goals for user involvement in individual action plans [28]. This type of match between organizational values and the characteristics of the eHealth tool is an important success criterion for eHealth implementations [41]. Once an organizational commitment is in place, portals such as ReConnect can facilitate more rapid shifts in practices toward recovery, in addition to more rapid dissemination of new knowledge within communities.

Limitations

We are not able to offer plausible explanations for the lack of men despite considerable efforts to recruit them, an issue which future service design studies need to address. Our opportunistic selection of quotes from superficially coded data to illustrate the four practice domains is not a balanced reflection of the experiences of participants. Thorough inductive analyses of participant experiences relative to collaboration and personal recovery are forthcoming. Nevertheless, we argue that the approach in this paper is justified in light of our aim of exploring the role such portals might play in shifting practices.

Comparison to Prior Work

This study complements reviews of technically supported self-management interventions in general [6,42-44], as well as more specific recovery-oriented self-management interventions [45,46]. Our own scoping review of e-recovery found 20 studies of six recovery-oriented portals in five countries [47]. These studies have promising, but as yet no definitive findings related to enhanced shared decision making [48], strengths and resilient self-care strategies [49], social connectedness and empowerment [7,50], and patient-centered care [51,52], to mention a few. This study is one of the few to use a participatory approach with an exploratory design using mixed methods and, to our knowledge, the first to discuss e-recovery findings more systematically in light of a recovery framework. Several of the components in our portal are similar to other solutions with promising findings (ie, access to health records [44], shared decision making [42], and peer support [45,46]). Combining multicomponents into a single portal, as we have done, increases the challenge of sorting out active ingredients. At the same time, our study of how such a multicomponent intervention is used and influences ongoing practices helps pave the way for implementation of subsequent, more evidence-supported interventions in communities.

Based on this exploratory study, the following hypotheses can be proposed for future studies:

Personal recovery: people who have Internet-based tools that help them articulate what is important to them, coupled with providers who help operationalize “what is important” into concrete goals, are more likely to become actively engaged in their recovery processes than those without such tools and support.

Citzenship: e-recovery portals that combine Internet-based peer support with local in-real-life peer support are more likely to lead to community engagement than those who have access exclusively to one or the other.

Working relationships: working relationships via e-recovery are more likely to be effective if coupled with low-threshold feedback mechanisms that monitor the quality of such relationships than those without.

Organizational commitment: organizations with commitments to recovery principles are more likely to invest in and benefit from e-recovery portals than those without such commitments.

Although policy-pushes toward recovery and eHealth are so far largely based on values and resource constraints, e-recovery is unlikely to survive without evidence of its efficacy in helping people live fulfilling lives. Progress toward efficacy trials will need to build on more in-depth understandings of how digital resources interplay with recovery processes and for which service users, dyads/teams, and community contexts. Future research would benefit from recovery researchers joining forces with computer scientists in sorting out key recovery-oriented factors that can be co-created, boosted, tested in larger controlled trials, and implemented through digital innovations.

Conclusions

The 24/7 availability of peer support and support for articulating personal goals in recovery processes represented itself a shift toward recovery-oriented practices within the participating communities. It is nevertheless the two practice domains, working relationships, and organizational commitment that are key to the more fundamental role that e-recovery portals can play in shifting practices toward recovery. Given organizational goals of monitoring service user involvement in care and the quality of working relationships, e-recovery portals can play a role in helping practices become more responsive to needs and aspirations as defined by service users.

Acknowledgments

This project was made possible by the Norwegian Research Council, the Norwegian ExtraFoundation for Health and Rehabilitation, the Norwegian Council for Mental Health, the Northern Norway Regional Health Authority, Vestre Viken Hospital Trust, and University Hospital in North-Norway. We are immeasurably grateful to all community participants whose experiences and insights gave life to this work.

XSL•FO
RenderX

Conflicts of Interest

DG and CMR are the developers of ReConnect, but have no financial or ownership rights to the portal. All authors are affiliated with the research center, but have no financial or nonfinancial interest to declare in relation to this study.

Multimedia Appendix 1

Service users' use of Internet and e-mail (n=29).

[PDF File (Adobe PDF File), 23KB - jmir_v19i5e145_app1.pdf]

Multimedia Appendix 2

Health care provider characteristics and experience with use of Internet and e-mail of (n=27).

[PDF File (Adobe PDF File), 34KB - jmir_v19i5e145_app2.pdf]

References

1. Mental Health Commission of Canada. Toward Recovery & Well-Being: A Framework for a Mental Health Strategy for Canada. Ottawa, ON: National Library of Canada; 2009.
2. World Health Organization. From Innovation to Implementation: eHealth in the WHO European Region 2016. Copenhagen: WHO Regional Office for Europe; 2016.
3. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A Web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
4. Stein BD, Kogan JN, Mihalyo MJ, Schuster J, Deegan PE, Sorbero MJ, et al. Use of a computerized medication shared decision making tool in community mental health settings: impact on psychotropic medication adherence. Community Ment Health J 2013 Apr;49(2):185-192. [doi: 10.1007/s10597-012-9528-8] [Medline: 22837104]
5. Beentjes TA, van Gaal BG, Goossens PJ, Schoonhoven L. Development of an e-supported illness management and recovery programme for consumers with severe mental illness using intervention mapping, and design of an early cluster randomized controlled trial. BMC Health Serv Res 2016 Jan 19;16:20 [FREE Full text] [doi: 10.1186/s12913-016-1267-z] [Medline: 26782620]
6. Karasouli E, Adams A. Assessing the evidence for e-resources for mental health self-management: a systematic literature review. JMIR Ment Health 2014;1(1):e3 [FREE Full text] [doi: 10.2196/mental.3708] [Medline: 26543903]
7. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
8. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
9. Krägeloh CU, Czuba KJ, Billington DR, Kersten P, Siegert RJ. Using feedback from patient-reported outcome measures in mental health services: a scoping study and typology. Psychiatr Serv 2015 Mar 01;66(3):224-241. [doi: 10.1176/appi.ps.201400141] [Medline: 25727110]
10. Markham S. Development of an online clinical trial recruitment portal for the NIHR mental health BRC. Res Involv Engagem 2016 Apr 1;2(1):1. [doi: 10.1186/s40900-016-0024-0]
11. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 2012;14(2):e46 [FREE Full text] [doi: 10.2196/jmir.1988] [Medline: 22397809]
12. Davidson L, Bellamy C, Guy K, Miller R. Peer support among persons with severe mental illnesses: a review of evidence and experience. World Psychiatry 2012 Jun;11(2):123-128 [FREE Full text] [Medline: 22654945]
13. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. Br J Psychiatry 2011 Dec;199(6):445-452 [FREE Full text] [doi: 10.1192/bjp.bp.110.083733] [Medline: 22130746]
14. Shogren KA, Wehmeyer ML, Lane KL. Embedding interventions to promote self-determination within multitiered systems of supports. Exceptionality 2016 Mar 23;24(4):213-224. [doi: 10.1080/09362835.2015.1064421]
15. Tse S, Tsoi EW, Hamilton B, O'Hagan M, Shepherd G, Slade M, et al. Uses of strength-based interventions for people with serious mental illness: A critical review. Int J Soc Psychiatry 2016 May;62(3):281-291. [doi: 10.1177/0020764015623970] [Medline: 26831826]
16. Le Boutillier C, Slade M, Lawrence V, Bird VJ, Chandler R, Farkas M, et al. Competing priorities: staff perspectives on supporting recovery. Adm Policy Ment Health 2015 Jul;42(4):429-438. [doi: 10.1007/s10488-014-0585-x] [Medline: 25134949]
17. Roberts G, Boardman J. Becoming a recovery-oriented practitioner. Adv Psychiat Treat 2014 Jan 02;20(1):37-47. [doi: 10.1192/apt.bp.112.010652]

XSL•FO
RenderX

18. Burgess P, Pirkis J, Coombs T, Rosen A. Review of Recovery Measures. Parramatta, Australia: Australian Mental Health Outcomes and Classification Network; 2010:1-78.
19. Shanks V, Williams J, Leamy M, Bird VJ, Le Boutillier C, Slade M. Measures of personal recovery: a systematic review. Psychiatr Serv 2013 Oct;64(10):974-980. [doi: 10.1176/appi.ps.005012012] [Medline: 23820592]
20. Sklar M, Groessl EJ, O'Connell M, Davidson L, Aarons GA. Instruments for measuring mental health recovery: a systematic review. Clin Psychol Rev 2013 Dec;33(8):1082-1095. [doi: 10.1016/j.cpr.2013.08.002] [Medline: 24080285]
21. Slade M. Measuring recovery in mental health services. Isr J Psychiatry Relat Sci 2010;47(3):206-212 [FREE Full text] [Medline: 21149985]
22. Gammon D, Strand M, Eng LS. Service users' perspectives in the design of an online tool for assisted self-help in mental health: a case study of implications. Int J Ment Health Syst 2014;8(1):2 [FREE Full text] [doi: 10.1186/1752-4458-8-2] [Medline: 24405765]
23. Baum F, MacDougall C, Smith D. Participatory action research. J Epidemiol Community Health 2006 Oct;60(10):854-857 [FREE Full text] [doi: 10.1136/jech.2004.028662] [Medline: 16973531]
24. Tariq S, Woodman J. Using mixed methods in health research. JRSM Short Rep 2013 Jun;4(6):1 [FREE Full text] [doi: 10.1177/2042533313479197] [Medline: 23885291]
25. Fulford B, Wallcraft J. Values-based practice: service user involvement in mental health research. In: Wallcraft J, Schrank B, Amering M, editors. Handbook of Service User Involvement in Mental Health Research. West Sussex: John Wiley & Sons; 2009:37-60.
26. Ministry of Health and Care Services. High Quality - Safe Services: Quality and Patient safety in the Health and Care Services. In: St Meld 10 (2012-2013). Oslo: Fagbokforlaget; 2012.
27. Royal Norwegian Ministry of Finance. National Budget 2017. Oslo: The Ministry of Finance; 2017.
28. Asker Municipality. 2015. Establishing mental health and substance abuse services: mandate description URL: https://www.asker.kommune.no/helse-og-omsorg/psykisk-helse/ [accessed 2017-04-28] [WebCite Cache ID 6q49RflTO]
29. World Health Organization. Well-Being Measures in Primary Health Care: The DepCare Project. Copenhagen: WHO Regional Office for Europe; 1998. URL: http://www.euro.who.int/__data/assets/pdf_file/0016/130750/E60246.pdf [accessed 2017-04-28] [WebCite Cache ID 6q4C96OUi]
30. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav Sci 1974 Jan;19(1):1-15. [Medline: 4808738]
31. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
32. Steinsbekk A. [Patient Activation Measure]. Tidsskr Nor Laegeforen 2008 Oct 23;128(20):2316-2318 [FREE Full text] [Medline: 19096487]
33. Andrews F, Whithey S. Social Indicators of Well-Being. New York: Plenum Press; 1976.
34. Corrigan PW, Salzer M, Ralph RO, Sangster Y, Keck L. Examining the factor structure of the recovery assessment scale. Schizophr Bull 2004;30(4):1035-1041. [Medline: 15957202]
35. Le Boutillier C, Leamy M, Bird VJ, Davidson L, Williams J, Slade M. What does recovery mean in practice? A qualitative analysis of international recovery-oriented practice guidance. Psychiatr Serv 2011 Dec;62(12):1470-1476. [doi: 10.1176/appi.ps.001312011] [Medline: 22193795]
36. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
37. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
38. Chinman M, George P, Dougherty RH, Daniels AS, Ghose SS, Swift A, et al. Peer support services for individuals with serious mental illnesses: assessing the evidence. Psychiatr Serv 2014 Apr 01;65(4):429-441. [doi: 10.1176/appi.ps.201300244] [Medline: 24549400]
39. Bjerkan J, Vatne S, Hollingen A. Web-based collaboration in individual care planning challenges the user and the provider roles - toward a power transition in caring relationships. J Multidiscip Healthc 2014;7:561-572 [FREE Full text] [doi: 10.2147/JMDH.S70470] [Medline: 25525367]
40. Duncan B, Miller S, Sparks J, Claud D. The Session Rating Scale: preliminary psychometric properties of a working alliance measure. J Brief Ther 2003;3(1):3-12.
41. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]
42. Barlow JH, Ellard DR, Hainsworth JM, Jones FR, Fisher A. A review of self-management interventions for panic disorders, phobias and obsessive-compulsive disorders. Acta Psychiatr Scand 2005 Apr;111(4):272-285. [doi: 10.1111/j.1600-0447.2005.00499.x] [Medline: 15740463]

XSL•FO
RenderX

43. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
44. van der Krieke L, Wunderink L, Emerencia AC, de Jonge P, Sytema S. E-mental health self-management for psychotic disorders: state of the art and future perspectives. Psychiatr Serv 2014 Jan 1;65(1):33-49. [doi: 10.1176/appi.ps.201300050] [Medline: 24129842]
45. Leitan ND, Michalak EE, Berk L, Berk M, Murray G. Optimizing delivery of recovery-oriented online self-management strategies for bipolar disorder: a review. Bipolar Disord 2015 Mar;17(2):115-127. [doi: 10.1111/bdi.12258] [Medline: 25238632]
46. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015 Aug 19;24(5):1-12. [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
47. Strand M, Gammon D, Ruland CM. Transitions from biomedical to recovery-oriented practices in mental health: a scoping review to explore the role of Internet-based interventions. BMC Health Serv Res 2017 Apr 07;17(1):257 [FREE Full text] [doi: 10.1186/s12913-017-2176-5] [Medline: 28388907]
48. Deegan PE. A web application to support recovery and shared decision making in psychiatric medication clinics. Psychiatr Rehabil J 2010;34(1):23-28. [doi: 10.2975/34.1.2010.23.28] [Medline: 20615841]
49. MacDonald-Wilson KL, Deegan PE, Hutchison SL, Parrotta N, Schuster JM. Integrating personal medicine into service delivery: empowering people in recovery. Psychiatr Rehabil J 2013 Dec;36(4):258-263. [doi: 10.1037/prj0000027] [Medline: 24320834]
50. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy. ACM Trans Comput-Hum Interact 2014 Feb 01;21(1):1-26. [doi: 10.1145/2513179]
51. Anttila M, Koivunen M, Välimäki M. Information technology-based standardized patient education in psychiatric inpatient care. J Adv Nurs 2008 Oct;64(2):147-156. [doi: 10.1111/j.1365-2648.2008.04770.x] [Medline: 18990095]
52. Emerencia A, van der Krieke L, Sytema S, Petkov N, Aiello M. Generating personalized advice for schizophrenia patients. Artif Intell Med 2013 May;58(1):23-36. [doi: 10.1016/j.artmed.2013.01.002] [Medline: 23419698]

Abbreviations

IT: information technology
PAM: Patient Activation Measure
RAS: Recovery Assessment Scale

Edited by G Eysenbach; submitted 20.02.17; peer-reviewed by M Stuckey; comments to author 21.03.17; revised version received 06.04.17; accepted 17.04.17; published 02.05.17.

Please cite as:
Gammon D, Strand M, Eng LS, Børøsund E, Varsi C, Ruland C
Shifting Practices Toward Recovery-Oriented Care Through an E-Recovery Portal in Community Mental Health Care: A Mixed-Methods Exploratory Study
J Med Internet Res 2017;19(5):e145
URL: http://www.jmir.org/2017/5/e145/
doi:10.2196/jmir.7524
PMID:28465277

©Deede Gammon, Monica Strand, Lillian Sofie Eng, Elin Børøsund, Cecilie Varsi, Cornelia Ruland. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 02.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates

Koustuv Saha[1], BTech (Hons); Ingmar Weber[2], PhD; Michael L Birnbaum[3,4,5], MD; Munmun De Choudhury[1], PhD

[1]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

[2]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

[3]The Zucker Hillside Hospital, Psychiatry Research, Northwell Health, Glean Olks, NY, United States

[4]Hofstra Northwell School of Medicine, Hempstead, NY, United States

[5]The Feinstein Institute for Medical Research, Manhasset, NY, United States

Corresponding Author:
Koustuv Saha, BTech (Hons)
School of Interactive Computing
Georgia Institute of Technology
Technology Square Research Building
85 Fifth Street NW
Atlanta, GA, 30308
United States
Phone: 1 4046929496
Fax: 1 4043856450
Email: koustuv.saha@gatech.edu

Abstract

Background: Schizophrenia is a rare but devastating condition, affecting about 1% of the world's population and resulting in about 2% of the US health care expenditure. Major impediments to appropriate and timely care include misconceptions, high levels of stigma, and lack of public awareness. Facebook offers novel opportunities to understand public awareness and information access related to schizophrenia, and thus can complement survey-based approaches to assessing awareness that are limited in scale, robustness, and temporal and demographic granularity.

Objective: The aims of this study were to (1) construct an index that measured the awareness of different demographic groups around schizophrenia-related information on Facebook; (2) study how this index differed across demographic groups and how it correlated with complementary Web-based (Google Trends) and non–Web-based variables about population well-being (mental health indicators and infrastructure), and (3) examine the relationship of Facebook derived schizophrenia index with other types of online activity as well as offline health and mental health outcomes and indicators.

Methods: Data from Facebook's advertising platform was programmatically collected to compute the proportion of users in a target demographic group with an interest related to schizophrenia. On consultation with a clinical expert, several topics were combined to obtain a single index measuring schizophrenia awareness. This index was then analyzed for differences across US states, gender, age, ethnic affinity, and education level. A statistical approach was developed to model a group's awareness index based on the group's characteristics.

Results: Overall, 1.03% of Facebook users in the United States have a schizophrenia-related interest. The schizophrenia awareness index (SAI) is higher for females than for males (1.06 vs 0.97, P<.001), and it is highest for the people who are aged 25-44 years (1.35 vs 1.03 for all ages, P<.001). The awareness index drops for higher education levels (0.68 for MA or PhD vs 1.92 for no high school degree, P<.001), and Hispanics have the highest level of interest (1.57 vs 1.03 for all ethnic affinities, P<.001). A regression model fit to predict a group's interest level achieves an adjusted R^2=0.55. We also observe a positive association between our SAI and mental health services (or institutions) per 100,000 residents in a US state (Pearson r=.238, P<.001), but a negative association with the state-level human development index (HDI) in United States (Pearson r=−.145, P<.001) and state-level volume of mental health issues in United States (Pearson r=−.145, P<.001).

Conclusions: Facebook's advertising platform can be used to construct a plausible index of population-scale schizophrenia awareness. However, only estimates of awareness can be obtained, and the index provides no information on the quality of the information users receive online.

XSL•FO
RenderX

(J Med Internet Res 2017;19(5):e156) doi:10.2196/jmir.6815

KEYWORDS

schizophrenia; psychotic disorders; online social networks; health awareness; mental health; public health; Facebook

Introduction

Background and Prior Work

Schizophrenia, although a relatively rare condition that occurs in approximately 1% of the world's population, is associated with devastating individual, familial, and societal loss [1]. Psychotic symptoms, including hallucinations, delusions, and disorganized thinking and behavior, typically emerge during precious years of adolescent and young adult development interfering with the establishment of healthy social, educational, and vocational foundations [2]. The substantial burden imposed by schizophrenia has been linked to its early onset as well its incurable nature often associated with persistent psychotic symptoms. Schizophrenia is ranked among the top 25 leading causes of disability worldwide [1]. The World Health Organization has estimated that direct costs associated with schizophrenia in the United States range from 1.6% to 2.6% of total health care expenditures [3]. The economic burden of schizophrenia is found to be more than US $60 billion per year [4].

Despite effective treatment options, major impediments to receiving early and appropriate care include lack of public awareness, misconceptions and misunderstanding, and high levels of stigma associated with schizophrenia and other related behavioral health disorders [5,6]. Professional help is often only sought after very long delays or a crisis, contributing to negative outcomes and poorer response to available treatments [7]. Furthermore, once in treatment, stigma and poor public awareness and understanding of schizophrenia and other psychotic disorders leads to worse outcomes including lowered self-esteem, social withdrawal, poor self-care, and drug and alcohol misuse [8].

The importance of public awareness and understanding of schizophrenia has been repeatedly demonstrated in the scientific literature [9]. Mental health literacy is defined as the knowledge and beliefs about mental disorders which aid their recognition, management or prevention and includes the ability to recognize specific behavioral health symptoms and disorders as well as available treatment options [10]. Existing mechanisms to assess (and ultimately improve) population awareness of schizophrenia, however, are challenged by the difficulty in gathering reliable, near real-time, and fine-grained population data [11]. A number of mental health programs and campaigns tend to employ survey- and questionnaire-based approaches to assess awareness, but these approaches are often expensive, time consuming, and unable to reach a large nation-wide audience. Moreover, these approaches thus far have provided limited data on schizophrenia awareness in different demographic groups, such as sex, race, age, or geography. Finally, most awareness assessment surveys are administered with large temporal gaps. This practice poses challenges in the ability to act on awareness-related information for deploying appropriate intervention programs, improving mental health service facilities, or for conducting mental health awareness campaigns. In this paper, we aim to tackle this challenge by utilizing unobtrusively gathered social media (Facebook) data in the design and development of a population index of schizophrenia awareness.

There is now convincing evidence that social media activity can be used to reliably monitor health-related thoughts and behaviors [12-14]; this forms the basis of our work. From a public health perspective, social media data has been used to infer information ranging from the spread of the influenza virus across the United States to rates of seasonal allergies, human immunodeficiency virus (HIV) infection, smoking, depression, obesity, and a variety of other county-level health statistics [15-20]. Chunara et al [21] argued that crowdsourced data, gathered via new technologies including the Internet and mobile phones offer the opportunity to fill gaps in and augment current epidemiological models. Early work by Ginsberg et al [22] demonstrated this promise by employing Google's search engine query data to monitor and track influenza trends in the United States.

There has been some work employing measurements of online activity, specifically Facebook "likes" as a mechanism for public health surveillance, and our work is motivated by these approaches. Gittelman et al [23] examined the predictive qualities of Facebook "likes" with regard to mortality, diseases, and lifestyle behaviors in 214 counties across the United States and 61 of 67 counties in Florida. Also employing Facebook "likes" as measures of topical interest, Chunara et al [24] found that activity-related interests across the United States and sedentary-related interests across New York City, inferred via Facebook's advertisement estimate algorithm, were significantly associated with obesity prevalence. Other research has also been successful in accurately assessing personality traits, intelligence, happiness, substance use, sexual orientation, religious and political views from individuals' "like" data on Facebook [25].

Prior work on the utility of social media activity data as a sensor of population-scale mental health awareness is limited. Korda and Itani [26] explored the role of social media as a platform for health promotion and behavior change, whereas Chou et al [27] explored this for health communication (also see [28]). Focusing on schizophrenia in particular, Birnbaum et al [29] studied how youth appropriated social media for seeking help around psychosis-related issues. Focusing on assessing the efficacy of awareness campaigns, Ayers et al [30] used the "Great American Smokeout" as a case study to observe cessation-related news reports and Twitter postings, and cessation-related help seeking via Google, Wikipedia, and government-sponsored quitlines. Other work conducted a randomized control trial to investigate the efficacy of an online depression awareness campaign that employed a Facebook-based recruitment strategy [31].

XSL•FO
RenderX

Study Objectives

As demonstrated in prior work, Facebook "likes" (henceforth referred to as Likes) are a powerful means by which users can express their own thoughts and interests, including health concerns. We note that this data may also reveal important details about the type of information and health education an individual is receiving. For example, individuals who like a specific Facebook page, such as the public page titled "Schizophrenia Awareness" [32] will be provided with information from that source directed back to the user, in the case of this page the stated mission being "By spreading awareness among a large and popular Facebook community, we can help rid the world of the misunderstandings and misconceptions surrounding this condition." Thus "Like" information can be particularly relevant for measuring schizophrenia awareness as it bears the potential to provide insight into the number of individuals who are expressing interest in or engaging with the topic of schizophrenia, as well as obtaining information related to schizophrenia through social media. Facebook additionally collects data on the interests of their users by passively monitoring any websites they visit, as long as those websites have Facebook Like or share functionality [33]. According to estimates, this kind of tracking happens on 75% of the 1000 most popular websites [34]. This information may further be useful to understand awareness around specific health topics such as schizophrenia. Together, in this paper, we employ these sources of information made available via Facebook to quantify population level awareness measures of schizophrenia among Facebook users in the United States.

Specifically, the three main objectives of our study are to

1. construct and examine an index that measures the awareness of different demographic groups around schizophrenia-related information on Facebook;
2. explore how this index differs across demographic groups and how it correlates with offline variables about population well-being; and
3. examine the relationship of Facebook derived schizophrenia index with other types of online activity as well as offline health and mental health outcomes and indicators.

Our specific source of data involves advertisement estimates of topics of interest made available by Facebook. Twitter, Facebook, and all other major social media platforms rely on targeted advertising for their revenue. To maximize the advertisement revenue, it is thus in the platform's interest to learn as much as possible about their users. This way Facebook and others can provide advertisers with access to highly targeted custom audiences that meet certain criteria, including demographics as well as certain topical "interests." According to Facebook's Business page [29], "Interests may include things people share on their Timelines, apps they use, advertisements they click, Pages they like, and other activities on and off of Facebook and Instagram. Interests may also factor in demographics such as age, gender, and location." Before an advertiser launches an advertising campaign, they are provided with an estimate of the audience size of the number of monthly active users on Facebook, so as to enable assessing the campaign's cost. We note that this mechanism of gathering audience estimates essentially acts as an on-demand digital census, answering questions of the form "how many female Facebook users in the state of New York aged 25-44, have at least a bachelor's degree and are interested in 'schizophrenia?'" In this way, with data gathered through such queries, we can assess the level of interest or engagement of a particular demographic group in schizophrenia, psychosis, and related topics. We refer to this topical interest or engagement as "schizophrenia awareness" within the context of the objectives proposed in this paper.

Methods

Data Sources

Facebook Advertisement Estimates

The main source of data for our study (objectives 1-3) comes from Facebook's Marketing application program interface (API) [35,36]. This API allows advertisers to limit their advertisement's reach to a highly targeted, custom audience. As noted above, this custom audience can be tailored in terms of user demographics and topical interests [37]. Some of these variables, such as the age and gender, are self-declared by Facebook users, whereas in particular the topical interests are inferred automatically based on interaction both on and off Facebook [38]. As an example, the reader can check the list of inferred interests for their own Facebook profile online [39].

To obtain a list of "marker interests" that indicate a user's interest in the schizophrenia topic, we made use of the (i) autocomplete and (ii) interest suggestion functionalities in Facebook's Adverts Manager [40]. Through the autocomplete functionality, we obtained a full list of topics matching "schizophren*" and "psychos*." This list was then pruned for relevance, removing topics such as "cosmic psychos." We then selected the topic "psychosis" and used the suggested related interests to expand our list. Interests that matched too few users were dropped from consideration. Table 1 shows the final list of the five selected marker interests. To arrive at this final list, we used an iterative method of adding and pruning marker interests, closely engaging with our psychiatrist coauthor, who is an expert in schizophrenia.

XSL•FO
RenderX

Table 1. List of the schizophrenia-related "marker interests" used in our study.

Interest	Global user count	US user count
Psychosis	5.8M	1.5M
Schizophrenia awareness	4.0M	480K
Hallucination	1.7M	120K
Schizoaffective disorder	57K	9.3K
Paranoid schizophrenia	46K	7.5K

With the Marketing API, we obtained various audience estimates for a range of combinations of demographic variables in combination with the marker interests. Concretely, we "or"-ed the five marker interests as their distinction is not clear cut, especially for typical Facebook users. Note that, to protect against targeting individuals, Facebook's Marketing API never returns values below 20, making 20 indistinguishable from 0. For our ensuing data analysis, tuples with an advertisement audience estimate of 20 were dropped from consideration.

Concerning demographic variables, we obtained separate audience estimates for (i) each of the 50 US states, (ii) both genders, (iii) five different age groups (13-17, 18-24, 25-44, 45-64, 65+ years), (iv) four different education levels (no high school degree; high school degree but no bachelor's degree; bachelor's degree but no master's or PhD degree; master's or PhD degree), and (v) four different ethnic affinities (Hispanic, African-American, Asian-American, none of the previous). For each of these dimensions, we also added a value of "all" which would define an audience across all possible values for that dimension. Note that Facebook carefully does not refer to "race" and that "ethnic affinity" is automatically inferred. So, a non-Hispanic user with a strong affinity toward Hispanic content and culture online might get labeled as Hispanic. Furthermore, no "White" ethnic affinity is offered. As of November 2016, Facebook has disabled the use of ethnic affinity for advertising related to housing, jobs, or credit, which have particular status concerning the protection against racial discrimination [41].

Google Trends

To get a comparative understanding with other complementary forms of online activity (objective 3), we obtained the US state-wise distribution of Google Trends [42] for the keywords present in our set of marker interests. Google Trends is a public Web facility of Google, based on their search engine, that shows how often a particular search-term (eg, "schizophrenia awareness") is entered relative to the total search-volume (in a US state).

Offline Population Data

In addition to the above online data, we obtained different types of offline data, as a way to examine the relationship and validity of schizophrenia awareness obtained from Facebook with related health and mental health outcomes and indicators (objective 3). First, we obtained the number of mental health institutions in each US state from the substance abuse and mental health services administration (SAMHSA) database [43]. For the purposes of our ensuing analysis, we normalized these numbers per state by dividing them by the corresponding state's population estimates (obtained from the United States Census Bureau [44]). Next, we intended to examine if factors like health, economy, and overall development of a state are associated with schizophrenia awareness. Therefore, we used the American human development index (HDI) of each US state, published by Measure of America, which is an initiative of the Social Science Research Council [45]. HDI is a composite index of life expectancy, education, and per capita income indicators. The American HDI is a modified version of the global HDI, using different indicators to better reflect the US context and to maximize the use of available data [46,47]. Intuitively, the HDI quantitatively summarizes the three dimensions of (i) health and well-being, (ii) knowledge, and (iii) standard of living. Third, we obtained the state-wise number of mental health issues among adults from The Henry J. Kaiser Family Foundation database [48].

Assessing Validity of the Acquired Facebook Data

To what extent is the population-level data given by Facebook's advertisement estimates representative of the US population? We note that it is important to assess the validity of the Facebook data we collected above before employing it in building a population index of schizophrenia awareness. For the purpose, we employed the same Marketing API to obtain the state-wise penetration of Facebook in United States. We obtained the number of users from the API for each state and computed their percentage on the corresponding population given by the US Census Bureau [44]. Based on this calculation, the national penetration of Facebook in United States was 63.16% and ranged from 56.35% (New Mexico) to 67.85% (Alaska), with a mean of 62.31% and standard deviation of 2.53%. This number is close to the Pew reported statistics of Facebook use in the US adult population (68%) [49].

Next, recall that one of our research objectives is to examine how awareness of schizophrenia varies across different demographic groups. However, one could anticipate that there are confounding factors between different demographic groups that may mediate this relationship; for example, young adults are perhaps more likely to be active Facebook users, leading to greater engagement manifested through Likes. To examine if these demographic group differences could be potential confounds, we again employed the Marketing API to evaluate the extent of activity of each demographic group on Facebook. Specifically, we obtained the count of audiences interested in either of entertainment, technology, music, and reading as marker interests; these topics were chosen because of the broad and popular scope. We then defined the percentage of active population as the ratio of the number of Facebook users with demographic attribute *d*, with at least 1 of 4 generic interests

XSL•FO
RenderX

(entertainment, technology, music, and reading), to the number of Facebook users with demographic attribute *d*. We found that for each of the demographic attributes: state, ethnic affinity, education, and age, the standard deviation of the percentages of active population lied within 1.5-8%, indicating that there is little difference in activity across groups, or in other words, Facebook activity is likely not a confounding factor in our study.

Index Construction

Our first research objective is to build an index which, for a given demographic group, measures the relative interest of that group in the topic of schizophrenia. Concretely, our SAI computes the proportion of the target demographic group, which has at least one of the marker interests:

$$SAI=(n_s[d]/\ n\ [d])\times 100$$

where $n_s(d)$ is the number of Facebook users with demographic attributes *d* and with at least one of the five schizophrenia marker interests, whereas *n* (*d*) is the total number of Facebook users with the same demographic attributes *d*. As an example, of the 19,000,000 women aged 18-24 years in the United States who use Facebook (n [d]), 230,000 have an interest in one of our schizophrenia interests (n_s[d]). So the SAI for women, in the 18-24 age group is 1.21. Similarly, for any given subgroup defined in terms of gender, age, ethnic affinity, and education level, their count of users with a schizophrenia-related interest is divided by the number of Facebook users for the same subgroup.

Statistical Models

First, to understand the interplay between different demographic variables and the SAI (objective 2), we computed and examined the trend of the index across changes in a single feature, such as across education levels or across ethnic affinities. Concretely, we used kruskal.test(...) function in R for Kruskal-Wallis rank sum test to compute *P* values and examine the statistical significance of data across demographic groups.

Next, to better understand the combined effects of several features (demographic attributes), we fitted linear regression models to predict the SAI for a given group. We experimented with two types of models, one including a binary indicator variable for each US state, and one where we replaced each state with their corresponding HDI, number of mental health institutions per 100,000 populations (MHP) and the percentage prevalence of mental health issues (MHI). The other independent variables for each of these models were the categorical features of age group, gender, ethnic affinity, and education. To ensure the stability of the model and the absence of statistical multicollinearity in the dataset, we computed the variance inflation factor (VIF). All computations were performed in R, using stats, reshape2, car, maps, maptools, and sp packages. We used coefficient of determination (R^2) to understand the goodness-of-fit of the linear regression model, and *P* statistic to understand the significance of the variables.

Finally, toward our third research objective, we employed the use of correlation tests (Pearson product-moment correlation coefficient), which revealed the association between state-wise SAI and state-wise measures of searches given by Google Trends for all of the marker interests, the state-wise numbers of mental health institutions, state-wise HDI, and state-wise MHI.

Results

Measurements Enabled by the Schizophrenia Awareness Index

Toward our first research objective, we computed the SAI separately for each gender (male, female), for each ethnic affinity (African-American, Asian-American, Hispanic, none of the above), for each age group, in years (13-17, 18-24, 25-44, 45-64, 65+ years), and for each education level (no high school degree; high school degree but no bachelor's degree; bachelor's degree but no master's or PhD degree; master's or PhD degree).

We found the mean SAI across states to be 1.11, with a standard deviation of 0.24 (Figure 1 a). Among the states, New Mexico (2.00) and West Virginia (1.73) have the highest awareness index, whereas Maryland (0.76) and New Jersey (0.75) have the lowest awareness index. Concerning gender, we observed that females (1.06) have greater SAI than males (0.97, $P<.001$). The awareness index across age groups ranges from 1.35 (age 25-44 years) to 0.27 (age 65+ years), with a $P<.001$ (Figure 1 b). The index decreases with an increase in education level, from its highest of 1.92 for Facebook users with no high school degree, to its lowest of 0.68 for Facebook users having a master's or PhD degree ($P<.001$) (Figure 1 c). Among ethnic affinities, the awareness index ranges from 1.57 (Hispanics) to 0.38 (Asian-Americans), with $P<.001$ (Figure 1 d). Each of these *P* values was computed using Kruskal-Wallis rank sum test.

Analysis of Schizophrenia Awareness Index Over Demographics

In the previous section, we examined the interplay between our SAI and individual demographic variables. As a next step and per objective 2, we examine how particular *combinations* of demographic variables relate to our index. For this, we start by looking at all combinations of gender, state, age group, ethnic affinity, and education level for which there were at least 10,000 Facebook users. Table 2 lists the top 10 combinations in terms of their SAI.

We observe that the groups with highest SAI in Table 2 are dominated by demographic combinations of individual groups with a positive association with the SAI. For example, all the top 10 combinations are for women with Hispanic ethnic affinity and from one of the two age groups, 18-24 and 25-44 years, which also individually have the highest SAI scores. This indicates that the effect of the SAI is potentially "additive" allowing more in-depth analysis via a linear regression model.

Next, in order to examine the variability in SAI across demographic groups, we fit a linear regression model (described in the section "Methods"), with the SAI of a demographic group as a dependent variable, and using gender, age group, ethnic affinity, and education level as independent variables. Results of this model are given in Table 3. Especially for states with small populations, such as Wyoming or Vermont, using counts

XSL•FO
RenderX

for, say, Hispanic women on Facebook aged 65+ years and holding a MA or PhD degree resulted in very low advertisement audience estimates. For the task of predicting the SAI for the remaining tuples, we obtained an adjusted R^2=0.55 (P<.001), indicating a significant relationship between the independent and dependent variables. Table 3 lists a few of the significant variables (based on their P values). To ensure the absence of multicollinearity between variables, we computed the generalized variance inflation factors (GVIFs). The GVIF for each of the variables, ranged from 1.01 to 1.30 and the GVIF^(1/[2df]) (where df is the degree of freedom, or number of coefficients in the subset), ranged from 1.01 to 1.04. As these values are very close to 1.0, this indicates that there was no large multicollinearity in the linear regression model.

Figure 1. (a) Choropleth map showing the levels of schizophrenia awareness index (SAI) in different states of the mainland US; (b) distribution of SAI over age and gender; (c) distribution of SAI over education level and gender; and (d) distribution of SAI over ethnic affinity and gender.

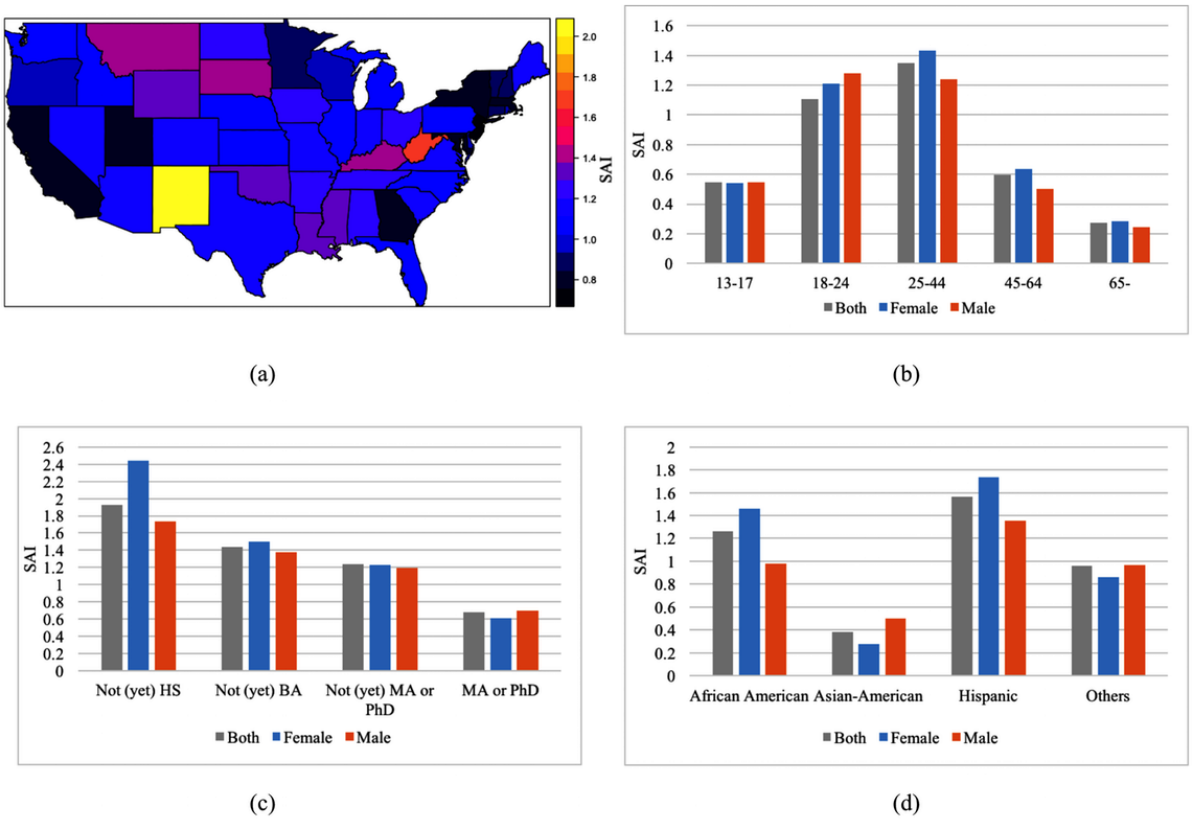


Table 2. Top 10 groups with highest schizophrenia awareness index (SAI) (at least 10,000 users).

Gender	State	Age (years)	Ethnic affinity	Education	SAI
Female	Ohio	18-24	Hispanic	Not (yet) BA	6.42
Female	Ohio	25-44	Hispanic	Not (yet) BA	6.36
Female	Colorado	18-24	Hispanic	Not (yet) HS	6.27
Female	Ohio	18-24	Hispanic	Not (yet) MA or PhD	5.91
Female	Kentucky	25-44	Hispanic	Not (yet) MA or PhD	5.88
Female	New Mexico	25-44	Hispanic	Not (yet) BA	5.52
Female	New Mexico	25-44	Hispanic	Not (yet) MA or PhD	5.47
Female	Ohio	25-44	Hispanic	Not (yet) MA or PhD	5.43
Female	Missouri	18-24	Hispanic	Not (yet) MA or PhD	5.38
Female	Michigan	18-24	Hispanic	Not (yet) BA	5.25

XSL•FO
RenderX

Table 3. Significant variables of Linear Regression Model 1 (dependent variable: schizophrenia awareness index [SAI]; independent variables: gender, state, age, education, and ethnic affinity).

Variable	Estimate (beta)	Std error	*t* statistic	*P* (>| *t* |)
(Intercept)	.00930	0.00170	5.5	<.001
Gender				
Male	−.00448	0.00040	−11.1	<.001
State				
California	−.00715	0.00173	−4.1	<.001
Connecticut	−.00696	0.00203	−3.4	<.001
Florida	−.00680	0.00179	−3.8	<.001
Maine	.02120	0.00234	9.1	<.001
Maryland	−.00687	0.00191	−3.6	<.001
Montana	.02590	0.00225	11.5	<.001
New Mexico	.00682	0.00200	3.4	<.001
New York	−.00615	0.00177	−3.5	<.001
South Dakota	.01970	0.00258	7.6	<.001
Vermont	.01930	0.00287	6.7	<.001
West Virginia	.02070	0.00213	9.7	<.001
Wyoming	.01480	0.00265	5.6	<.001
Age (years)				
18-24	.00914	0.00070	13.1	<.001
25-44	.01110	0.00069	16.1	<.001
Education				
Not (yet) BA	.00329	0.00073	4.5	<.001
Not (yet) High School	.01020	0.00079	12.9	<.001
Ethnic affinity				
Asian-American	−.00637	0.00095	−6.7	<.001
Hispanic	.01330	0.00060	22.3	<.001
Other	−.00549	0.00057	−9.6	<.001

In a second linear regression model (described within the section, “Methods”), we predicted the SAI using gender, age group, ethnic affinity, education level, as well as MHP, HDI, and MHI as independent variables. In contrast to the first model, we replaced the binary indicator variables for each state with their corresponding MHP, HDI, and MHI values. This model had an adjusted R^2=0.56 (P<.001). The GVIF for each of the variables ranged from 1.01 to 1.70, and the GVIF^(1/[2×df]), ranged from 1.00 to 1.07, again indicating the absence of multicollinearity in the linear regression model. Table 4 lists the significant variables (based on their P value).

Though the adjusted R^2 for the second model was higher than for the first model, we still wanted to examine the role played by a state’s mental health issues and infrastructure or overall development. So, we fit a third linear regression model using only gender, age group, ethnic affinity, and education as independent variables, that is, not using either of state or MHP or HDI as predictors. We observed a reduced adjusted R^2 of 0.42, compared with 0.56 for the previous model with MHP, HDI, and MHI included.

Relationship of SAI With Online Activity and Offline Health Indicators

Toward our third and final research objective, we first, we performed correlation tests for MHP, HDI, and MHI with SAI, and Table 5 shows the results. On the basis of these results, we can confirm that each of MHP, HDI, and MHI has an association with SAI. In the case of MHP, presence of more mental health institutions is *positively* associated with SAI. However, an increase in the HDI is *negatively* associated with SAI. For MHI, we observe that prevalence of mental health issues in different states is negatively correlated with SAI.

XSL•FO
RenderX

Table 4. Significant variables of Linear Regression Model 2 (dependent variable: SAI [schizophrenia awareness index]; independent variables: gender, MHP, HDI, MHI, age, education, and ethnic affinity).

Variable	Estimate (beta)	Std error	*t* statistic	*P* (>| *t* |)
(Intercept)	.02181	0.00535	4.1	<.001
Gender				
Male	−.00183	0.00080	−2.3	<.001
MHP[a]	.00199	0.00016	12.3	<.001
HDI[b]	−.00364	0.00055	−6.6	<.001
MHI[c]	−.02740	0.01266	−2.2	<.001
Age (years)				
18-24	.01463	0.00123	11.9	<.001
25-44	.01642	0.00124	13.3	<.001
Education				
Not (yet) BA	.00717	0.00119	6.0	<.001
Not (yet) HS	.01448	0.00126	11.5	<.001
Not (yet) MA or PhD	.00507	0.00119	4.3	<.001
Ethnic affinity				
Asian-American	−.01052	0.00148	−7.1	<.001
Hispanic	.01887	0.00108	17.5	<.001
Other	−.00414	0.00096	−4.3	<.001

[a]MHP: number of mental health institutions per 100,000 population.

[b]HDI: human development index.

[c]MHI: percentage of reported mental health issues in each US state.

Table 5. Pearson product-moment correlation of schizophrenia awareness index (SAI), with number of mental health institutions per 100,000 population (MHP), human development index (HDI), and volume of mental health issues (MHI) in each US state.

Metric	MHP vs SAI	HDI vs SAI	MHI vs SAI
Pearson *r*	.238	−.145	−.279
95% CI	0.206-0.270	−0.179 to −0.112	−0.516 to 0.001
t statistic	14.07	−8.44	−1.99
P value	<.001	<.001	.05

Table 6. Pearson product-moment correlation of schizophrenia awareness index (SAI), with Google Trends results of search of schizophrenia-related marker interests for each US states.

Google Trend search	Schizophrenia	Psychosis	Schizoaffective disorder	Hallucination	Paranoid schizophrenia
Pearson *r*	.31	.17	.28	.14	.46
Degree of freedom	48	48	40	43	36
95% CI	0.04-0.55	−0.11 to 0.43	−0.02 to 0.54	−0.16 to 0.41	0.17-0.68
t statistic	2.29	1.17	1.87	0.91	3.13
P value	.03	.25	.069	.34	.003

Next, for comparing state-wise SAI and the state-wise statistics of the marker interests given by Google Trends, we computed the correlation between the percentile Google Search of our marker interests for each state with the corresponding SAI. Table 6 lists the results. These results show that there is some harmony between the two online activity detection platforms (especially for marker interests Schizophrenia and Paranoid Schizophrenia).

XSL•FO
RenderX

Discussion

Principal Findings

Our work shows that data obtained from pervasive social technologies like Facebook may be employed to track public health awareness around schizophrenia-related concerns. This form of tracking needs little active participation and intrusion, and can be performed at a fine temporal, spatial, and demographic granularity at little resource cost. This makes our proposed index a promising mechanism for monitoring the impact of mental health awareness campaigns in near real-time, instead of through retrospective approaches such as surveys that are prone to hindsight bias of responders as well as experimenter demand effects. Prior work [34] reveals that the advertisement estimates are updated weekly, enabling this type of near real-time monitoring; we include a link [50] as a prototype application for the purpose. Additionally, our index could be used to craft more tailored awareness programs, targeting specific demographic groups with low levels of schizophrenia awareness.

Our results raise some interesting discussion points related to population-scale awareness of schizophrenia. First, due to the unavailability of gold standard data on schizophrenia awareness in the general population, we were not able to validate our proposed index with direct correlation metrics. However, our statistical examinations of SAI with three population-scale variables, availability of mental health services (or presence of mental health institutions), the HDI, and the proportion of mental health issues reported, show that the levels of schizophrenia awareness given by our approach bears relationship to offline variables of population well-being in different states of the United States.

On the above note, a priori, it is not clear if there should be a positive or a negative association between the presence of mental health institutions and our proposed SAI. A negative link could indicate that limited access to mental health institutions forces residents to go online to obtain information, whereas a positive link could indicate that more access helps to promote visibility and awareness. In our data, we find a robust positive link between the SAI and the MHP (see Tables 4 and 5). Similarly, both a positive association between the HDI and our SAI, as a result of better education and general interest in health topics, or a negative association, due to the availability of traditional information sources, are plausible hypotheses. In addition, association between SAI and the MHI could be either due to more number of mental issues leading to higher awareness, or due to higher awareness leading to decreased number of issues in mental health. Our results indicate a negative link between SAI and the MHI (see Tables 4 and 5). Further research is needed to derive causal explanations behind the observed directionality of these correlational relationships between SAI and population well-being variables.

Comparison With Prior Work

Although there is a lack of prior work employing social media data to assess schizophrenia awareness, we situate our findings in the light of clinical, psychiatric, and public health research on mental health and schizophrenia awareness. A notable finding given by analyses enabled by our index is the heightened awareness among females compared with males. It is known from prior work that gender and age influence intentions to seek professional psychological help [51]. Females exhibit more favorable intentions to seek help from mental health professionals than males, likely due to their positive attitudes concerning psychological openness [52, 53]. This may explain our observed higher SAI in females over males.

Our results further indicate age-related differences in schizophrenia awareness among females and males. Specifically, SAI for females is highest in the age group 25-44 years, whereas for males, it is highest in the 18-24 age group. It is known that typically schizophrenia develops in late adolescence and early adulthood, and its clinical onset is observed to be later for females compared with males [54]. Consequently, it is possible that the awareness of the condition will correlate to the timeframe individuals are typically known to be affected by the illness.

Furthermore, we found Hispanics to be the ethnic affinity with the highest SAI (1.57 vs 1.03 for all ethnic affinities, $P<.001$). However, according to a study on the demographics of mental health service users [55], Hispanics are the ethnicity with the lowest use of mental health services. The same study also found that they are the group that is most likely to mention “prejudice and discrimination” as a reason for not using mental health services. This discrepancy may indicate that this ethnic group is appropriating online sources (namely, Facebook) more extensively to gain information regarding schizophrenia, possibly because they provide a prejudice, bias, and stigma-free mechanism to gain mental health literacy and information.

Of note, we find that the overall schizophrenia awareness given by our index is 1.03, which is close to the prevalence of 1.1% given by the National Institute of Mental Health (NIMH) [56, 57]. One possible explanation behind this observed alignment could be that the set of people who express schizophrenia interest on Facebook (eg, by “liking” different pages on the topic), also tend to be the ones who are themselves suffering or have suffered from schizophrenia or related psychotic disorders. However, further research through validated self-reported or clinical assessment of mental health is required to reveal to what extent schizophrenia awareness derived from online data sources actually maps to schizophrenia prevalence.

Limitations

Our analysis relies on data provided by Facebook. Though this data is large-scale in its size and its granularity, it lacks transparency concerning how exactly Facebook collects this data. In particular, it is only partly clear how a user’s “interests” are inferred. This reliance on a “black box” is not desirable and related limitations have been discussed in the context of the now obsolete Google Flu service [58]. One consequence of hidden algorithmic details is that we observed certain fluctuations and inconsistencies. As an example, Figure 1 for the age group of 18-24 years shows a SAI value of 1.11, whereas the value for each gender separately in the same age group is higher (1.21 for female and 1.28 for male). The reason for this is that the advertisement audience estimate for the specification regardless of gender gave a lower count (410,000 out of

XSL•FO
RenderX

37,000,000) than the two counts separately (230,000 out of 19,00,0000 for female and 230,000 out of 18,000,000 for male). Though these inconsistencies were generally small, they show that the numbers one obtains might not be as reliable as one might assume.

But, even if we knew how exactly Facebook infers a user's interests and how they perform their audience estimates, there would still be the question of what this *means*. Does the interest in a disease indicate that the user is suffering from it? Does it indicate that someone in the user's social circle is affected by it? Or, does it just indicate general awareness of a topic? Future work can help to shed light on this question by looking more closely at the link between interest levels and prevalence rates. Relatedly, although we found statistically significant relationships between SAI and offline variables like state-wise number of MHP or HDI, further research is needed to understand the implications of these relationships.

Though this work uses Facebook data, other platforms such as Twitter also support targeted advertising and provide an audience estimate. Over time we expect the reach and the accuracy of such "digital census" tools to improve, supporting a wide range of public health monitoring capabilities. But on top of the passive use for data monitoring, targeted advertising or respondent driven sampling techniques can also be used for recruiting study or survey participants [59, 60].

Despite the growing use of social media and its potential as a source of population level data, it is important to note that large parts of the population still do not use it. Relying on data from social media to inform policy decisions can hence lead to a biased view on the world, caused by the digital divide [61]. In fact, even if a particular group is on social media, data analytic algorithms might not perform adequately for members of this group. As an example, the detection of schizophrenia-related interests might perform worse for minority languages where less training data is available, a case of "algorithmic bias" [62]. Similarly, not all individuals report their demographic attributes on Facebook, or include sufficient signals through their Facebook activity that may lend toward advertisements estimates computed by Facebook or the awareness measures given by our approach. We acknowledge a self-selection bias in our data and therefore the assessments made.

Finally, the work described so far only considers the aspect of *monitoring* schizophrenia awareness, but not *improving* it. However, the platform we are using to obtain our data is ultimately designed for exactly that getting a particular message out to a highly targeted audience. Although running a campaign with the goal of attracting millions of advertisement clicks might be beyond the practical feasibility of typical public health campaign budgets, a smaller pilot campaign, in the form of a randomized control trial, may be used to assess the right kind of message to use on other media. Furthermore, the Facebook advertisement platform can be leveraged to reach out to a variety of users, through a survey so as to also validate and calibrate the levels of awareness of schizophrenia among different groups. These constitute promising directions for future research.

Conclusions

In summary, though more work on validation remains, we believe that our work indicates that there is clear potential in using unobtrusively collected data from social media for monitoring awareness of stigmatized health conditions such as schizophrenia and potentially others as well.

Acknowledgments

De Choudhury was partly supported through an NIH grant #1R01GM11269701.

Conflicts of Interest

None declared.

References

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]
2. Kane JM, Correll CU. Past and present progress in the pharmacologic treatment of schizophrenia. J Clin Psychiatry 2010 Sep;71(9):1115-1124 [FREE Full text] [doi: 10.4088/JCP.10r06264yel] [Medline: 20923620]
3. Barbato A. Schizophrenia and Public Health. Geneva: World Health Organization; 1998.
4. Marcus SC, Olfson M. Outpatient antipsychotic treatment and inpatient costs of schizophrenia. Schizophr Bull 2008 Jan;34(1):173-180 [FREE Full text] [doi: 10.1093/schbul/sbm061] [Medline: 17578893]
5. Compton M, Broussard B. Conceptualizing the Multifaceted Determinants of the Duration of Untreated Psychosis. CPSR 2011 Feb 01;7(1):1-11. [doi: 10.2174/157340011795945865]
6. Franz L, Carter T, Leiner AS, Bergner E, Thompson NJ, Compton MT. Stigma and treatment delay in first-episode psychosis: a grounded theory study. Early Interv Psychiatry 2010 Feb;4(1):47-56 [FREE Full text] [doi: 10.1111/j.1751-7893.2009.00155.x] [Medline: 20199480]

XSL•FO
RenderX

7. Kane JM, Robinson DG, Schooler NR, Mueser KT, Penn DL, Rosenheck RA, et al. Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the NIMH RAISE early treatment program. Am J Psychiatry 2016 Apr 1;173(4):362-372. [doi: 10.1176/appi.ajp.2015.15050632] [Medline: 26481174]
8. Corrigan PW, Druss BG, Perlick DA. The impact of mental illness stigma on seeking and participating in mental health care. Psychol Sci Public Interest 2014 Oct;15(2):37-70. [doi: 10.1177/1529100614531398] [Medline: 26171956]
9. Schomerus G, Schwahn C, Holzinger A, Corrigan PW, Grabe HJ, Carta MG, et al. Evolution of public attitudes about mental illness: a systematic review and meta-analysis. Acta Psychiatr Scand 2012 Jun;125(6):440-452. [doi: 10.1111/j.1600-0447.2012.01826.x] [Medline: 22242976]
10. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. Med J Aust 1997 Feb 17;166(4):182-186. [Medline: 9066546]
11. Jorm A, Barney LJ, Christensen H, Highet NJ, Kelly CM, Kitchener BA. Research on mental health literacy: what we know and what we still need to know. Aust N Z J Psychiatry 2006 Jan;40(1):3-5. [doi: 10.1111/j.1440-1614.2006.01734.x]
12. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
13. De Choudhury M. Role of social media in tackling challenges in mental health. In: Proceedings of the 2nd International Workshop on Socially-aware Multimedia. 2013 Presented at: the 2nd International Workshop on Socially-aware Multimedia; October 21-25, 2013; Barcelona p. 49-52. [doi: 10.1145/2509916.2509921]
14. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
15. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
16. Dredze M. How social media will change public health. IEEE Intell Syst 2012 Jul;27(4):81-84. [doi: 10.1109/MIS.2012.76]
17. 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. [doi: 10.1371/journal.pone.0083672]
18. 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 p. 47-56. [doi: 10.1145/2464464.2464480]
19. Culotta A. Estimating county health statistics with twitter. 2014 Presented at: the 32nd annual ACM conference on Human factors in computing systems; April 26–May 1, 2014; Toronto, Ontario p. 1335-1344. [doi: 10.1145/2556288.2557139]
20. Abbar S, Mejova Y, Weber I. You tweet what you eat: Studying food consumption through twitter. 2015 Presented at: In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18–23, 2015; Seoul, Republic of Korea p. 3197-3206.
21. Chunara R, Smolinski MS, Brownstein JS. Why we need crowdsourced data in infectious disease surveillance. Curr Infect Dis Rep 2013 Aug;15(4):316-319 [FREE Full text] [doi: 10.1007/s11908-013-0341-5] [Medline: 23689991]
22. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457(7232):1012-1014. [doi: 10.1038/nature07634]
23. Gittelman S, Lange V, Gotway CC, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res 2015;17(4):e98 [FREE Full text] [doi: 10.2196/jmir.3970] [Medline: 25895907]
24. Chunara R, Bouton L, Ayers JW, Brownstein JS. Assessing the online social environment for surveillance of obesity prevalence. PLoS One 2013 Apr;8(4):e61373 [FREE Full text] [doi: 10.1371/journal.pone.0061373] [Medline: 23637820]
25. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. PNAS 2013;110(15):5802-5805. [doi: 10.1073/pnas.1218772110]
26. 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]
27. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
28. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
29. Birnbaum ML, Candan K, Libby I, Pascucci O, Kane J. Impact of online resources and social media on help-seeking behaviour in youth with psychotic symptoms. Early Interv Psychiatry 2016 Oct;10(5):397-403. [doi: 10.1111/eip.12179] [Medline: 25200743]
30. Ayers J, Westmaas JL, Leas EC, Benton A, Chen Y, Dredze M, et al. Leveraging big data to improve health awareness campaigns: a novel evaluation of the great American smokeout. JMIR Public Health Surveill 2016;2(1):e16 [FREE Full text] [doi: 10.2196/publichealth.5304] [Medline: 27227151]
31. Hui A, Wong P, Fu KW. Evaluation of an Online Campaign for Promoting Help-Seeking Attitudes for Depression Using a Facebook Advertisement: An Online Randomized Controlled Experiment. JMIR Ment Health 2015;2(1):e5 [FREE Full text] [doi: 10.2196/mental.3649] [Medline: 26543911]

XSL•FO
RenderX

32. Facebook. Schizophrenia Awareness URL: https://www.facebook.com/schizowareness/ [accessed 2017-03-28] [WebCite Cache ID 6pJQnwmjK]
33. Facebook. What is interests targeting? URL: https://www.facebook.com/unsupportedbrowser [accessed 2017-03-28] [WebCite Cache ID 6pJRnqbyF]
34. Facebook. Marketing API URL: https://www.facebook.com/unsupportedbrowser [accessed 2017-03-28] [WebCite Cache ID 6pJRpII4R]
35. Facebook. Marketing API - targeting specs URL: https://www.facebook.com/unsupportedbrowser [accessed 2017-03-28] [WebCite Cache ID 6pJS5qQZ8]
36. Simonite T. MIT Technology Review. Facebook's Like buttons will soon track your web browsing to target ads URL: https://www.technologyreview.com/s/541351/facebooks-like-buttons-will-soon-track-your-web-browsing-to-target-ads/ [accessed 2017-03-28] [WebCite Cache ID 6pJQpC38b]
37. Valentino-DeVries J, Singer-Vine J. The Wall Street Journal. 2012. They Know What You're Shopping For URL: https://www.wsj.com/articles/SB10001424127887324784404578143144132736214 [accessed 2017-03-28] [WebCite Cache ID 6pJS8AyQV]
38. Facebook. Marketing API - reach estimate URL: https://www.facebook.com/unsupportedbrowser [accessed 2017-03-28] [WebCite Cache ID 6pJS2oXBz]
39. Facebook. Advert preferences URL: https://www.facebook.com/unsupportedbrowser [accessed 2017-03-28] [WebCite Cache ID 6pJSAYhuu]
40. Facebook. Adverts Manager URL: https://business.facebook.com/ads/manager/creation/creation/ [WebCite Cache ID 6lDaVzh3q]
41. Forbes. URL: http://www.forbes.com/sites/kathleenchaykowski/2016/11/11/facebook-to-ban-ethnic-affinity-targeting-for-housing-employment-credit-related-ads/#8f5b5cda7477 [WebCite Cache ID 6n2VdNvyP]
42. Google Trends. URL: https://www.google.co.in/trends/ [accessed 2016-12-26] [WebCite Cache ID 6n2Vu6w8M]
43. Substance Abuse and Mental Health Services Administration. Mental health facilities data URL: http://www.samhsa.gov/data/mental-health-facilities-data-nmhss [accessed 2017-03-28] [WebCite Cache ID 6pJSZxJuY]
44. United States Census Bureau. Population and housing unit estimates URL: https://www.census.gov/popest/ [accessed 2017-03-28] [WebCite Cache ID 6pJSeOGr3]
45. Measure of America. Human development index URL: http://www.measureofamerica.org/human-development/ [WebCite Cache ID 6lDaPPJbc]
46. United Nations Development Program. Human development index URL: http://hdr.undp.org/en/content/human-development-index-hdi [WebCite Cache ID 6l0e9Wlxl]
47. Measure of America. HD index and American HD index URL: http://www.measureofamerica.org/human-development/#human [WebCite Cache ID 6l0eSfseV]
48. Kaiser Family Foundation. Percent of adults reporting poor mental health status URL: http://kff.org/state-category/health-status/mental-health/ [accessed 2016-12-20] [WebCite Cache ID 6mubInYJd]
49. Pew Research Center. Pew Social Media Update URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [WebCite Cache ID 6n9ttjxhn]
50. Schizophrenia awareness index tool. URL: http://koustuv.com/computeSAI.html [WebCite Cache ID 6n9rgy5Ip]
51. Mackenzie CS, Gekoski WL, Knox VJ. Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. Aging Ment Health 2006 Nov;10(6):574-582. [doi: 10.1080/13607860600641200] [Medline: 17050086]
52. Wide J, Mok H, McKenna M, Ogrodniczuk JS. Effect of gender socialization on the presentation of depression among men: a pilot study. Can Fam Physician 2011 Feb;57(2):e74-e78 [FREE Full text] [Medline: 21642709]
53. Gibbons RJ, Thorsteinsson EB, Loi NM. Beliefs and attitudes towards mental illness: an examination of the sex differences in mental health literacy in a community sample. PeerJ 2015;3:e1004 [FREE Full text] [doi: 10.7717/peerj.1004] [Medline: 26413429]
54. Verdoux H, Geddes JR, Takei N, Lawrie SM, Bovet P, Eagles JM, et al. Obstetric complications and age at onset in schizophrenia: an international collaborative meta-analysis of individual patient data. Am J Psychiatry 1997 Sep;154(9):1220-1227. [doi: 10.1176/ajp.154.9.1220] [Medline: 9286180]
55. Substance AbuseMental Health Services Administration. Racial/Ethnic differences in mental health service use among adults URL: http://www.samhsa.gov/data/sites/default/files/MHServicesUseAmongAdults/MHServicesUseAmongAdults.pdf [accessed 2017-03-28] [WebCite Cache ID 6pJPMYafU]
56. National Institute of Mental Health. Schizophrenia prevalence URL: https://www.nimh.nih.gov/health/statistics/prevalence/schizophrenia.shtml [accessed 2017-03-28] [WebCite Cache ID 6pJPLwR8w]
57. Wu EQ, Shi L, Birnbaum H, Hudson T, Kessler R. Annual prevalence of diagnosed schizophrenia in the USA: a claims data analysis approach. Psychol Med 2006 Nov;36(11):1535-1540. [doi: 10.1017/S0033291706008191] [Medline: 16907994]
58. 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]

XSL•FO
RenderX

59. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
60. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
61. Kaplan A, Haenlein M. Users of the world, unite! the challenges and opportunities of Social Media. Business Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
62. Hajian S, Bonchi F, Castillo C. Algorithmic bias: from discrimination discovery to fairness-aware data mining. In: Association for Computing Machinery. 2016 Presented at: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13, 2016; San Francisco p. 2125-2126.

Abbreviations

API: application program interface
GVIF: generalized variance inflation factor
HDI: human development index
HIV: human immunodeficiency virus
MHI: percentage of reported mental health issues in each US state
MHP: number of mental health institutions per 100,000 population
NIMH: National Institute of Mental Health
SAI: schizophrenia awareness index
SAMHSA: substance abuse and mental health services administration
VIF: variance inflation factor

Edited by R Calvo; submitted 15.10.16; peer-reviewed by I Choi, T Althoff, R Kennedy; comments to author 20.11.16; revised version received 07.01.17; accepted 24.01.17; published 08.05.17.

Please cite as:
Saha K, Weber I, Birnbaum ML, De Choudhury M
Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates
J Med Internet Res 2017;19(5):e156
URL: http://www.jmir.org/2017/5/e156/
doi:10.2196/jmir.6815
PMID:28483739

©Koustuv Saha, Ingmar Weber, Michael L Birnbaum, Munmun De Choudhury. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Automated Text Messaging as an Adjunct to Cognitive Behavioral Therapy for Depression: A Clinical Trial

Adrian Aguilera[1,2], PhD; Emma Bruehlman-Senecal[1], PhD; Orianna Demasi[3], BSc; Patricia Avila[1,2], BA

[1]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[2]Zuckerberg San Francisco General Hospital, Department of Psychiatry, University of California, San Francisco, San Francisco, CA, United States

[3]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, United States

Corresponding Author:
Adrian Aguilera, PhD
School of Social Welfare
University of California, Berkeley
120 Haviland Hall, MC 7400
Berkeley, CA, 94720
United States
Phone: 1 5106428564
Fax: 1 5106436126
Email: aguila@berkeley.edu

Abstract

Background: Cognitive Behavioral Therapy (CBT) for depression is efficacious, but effectiveness is limited when implemented in low-income settings due to engagement difficulties including nonadherence with skill-building homework and early discontinuation of treatment. Automated messaging can be used in clinical settings to increase dosage of depression treatment and encourage sustained engagement with psychotherapy.

Objectives: The aim of this study was to test whether a text messaging adjunct (mood monitoring text messages, treatment-related text messages, and a clinician dashboard to display patient data) increases engagement and improves clinical outcomes in a group CBT treatment for depression. Specifically, we aim to assess whether the text messaging adjunct led to an increase in group therapy sessions attended, an increase in duration of therapy attended, and reductions in Patient Health Questionnaire-9 item (PHQ-9) symptoms compared with the control condition of standard group CBT in a sample of low-income Spanish speaking Latino patients.

Methods: Patients in an outpatient behavioral health clinic were assigned to standard group CBT for depression (control condition; n=40) or the same treatment with the addition of a text messaging adjunct (n=45). The adjunct consisted of a daily mood monitoring message, a daily message reiterating the theme of that week's content, and medication and appointment reminders. Mood data and qualitative responses were sent to a Web-based platform (HealthySMS) for review by the therapist and displayed in session as a tool for teaching CBT skills.

Results: Intent-to-treat analyses on therapy attendance during 16 sessions of weekly therapy found that patients assigned to the text messaging adjunct stayed in therapy significantly longer (median of 13.5 weeks before dropping out) than patients assigned to the control condition (median of 3 weeks before dropping out; Wilcoxon-Mann-Whitney z=−2.21, P=.03). Patients assigned to the text messaging adjunct also generally attended more sessions (median=6 sessions) during this period than patients assigned to the control condition (median =2.5 sessions), but the effect was not significant (Wilcoxon-Mann-Whitney z=−1.65, P=.10). Both patients assigned to the text messaging adjunct (B=−.29, 95% CI −0.38 to −0.19, z=−5.80, P<.001) and patients assigned to the control conditions (B=−.20, 95% CI −0.32 to −0.07, z=−3.12, P=.002) experienced significant decreases in depressive symptom severity over the course of treatment; however, the conditions did not significantly differ in their degree of symptom reduction.

Conclusions: This study provides support for automated text messaging as a tool to sustain engagement in CBT for depression over time. There were no differences in depression outcomes between conditions, but this may be influenced by low follow-up rates of patients who dropped out of treatment.

(J Med Internet Res 2017;19(5):e148) doi:10.2196/jmir.6914

XSL•FO
RenderX

KEYWORDS

depression; text messaging; cognitive behavioral therapy; mhealth; mental health; Latinos

Introduction

Background

Cognitive behavioral therapy (CBT) is an efficacious treatment for depression [1] delivered via various mediums (eg, individual, group, telephone, Internet) and with diverse populations [2,3]. Low-income and Latino populations can benefit from CBT for depression, but they utilize services at lower rates [4] and have shown lower levels of engagement (homework completion and attendance) once in treatment [5]. Research has shown that increased dosage of psychotherapeutic treatment via increased engagement leads to improved outcomes [6]. Given this literature, research exploring more effective methods of improving attendance of psychotherapy and sustaining engagement with treatment is needed, especially within low income, ethnic minority populations where the problem of nonattendance and dropout is particularly pronounced [7].

Mobile Messaging to Improve Engagement

Given the association between engagement and improved depression outcomes [6,8,9], interventions that increase engagement in psychotherapy could improve the effectiveness of CBT for depression, particularly in public sector settings. Mobile health (mHealth) tools such as text messaging can increase engagement in psychotherapeutic interventions in a number of ways. Text messaging is low cost and pervasive across socioeconomic and demographic groups in the United States. The delivery of text messages to patients in treatment can prompt engagement in CBT homework, increasing the application of skills learned in therapy sessions, and leading to improved outcomes and treatment adherence [10-12]. The most comprehensive review of text messaging and mental health studies recently concluded that texting is viewed positively and improves adherence and symptom measurement in treatments [12]. Automated text messages during treatment can also serve to “stay on patients’ radars” and make patients feel supported and close to the group, strengthening therapeutic alliance, and increasing the likelihood that they will attend session and reengage with psychotherapy after a period of absence [13,14]. Text messages can further be used to send direct reminders to attend sessions and to take medications. Additionally, data obtained from text message inquiries can be visualized to help behavioral health clinicians provide higher quality, more personalized care. By periodically reviewing graphical representations of feedback on patient progress before session, clinicians can identify key events and address any clinically relevant events during or between sessions [15-17]. As a result of reminding patients about sessions, making patients feel more supported, and making sessions more immediately relevant, mHealth adjuncts to treatment may increase the number of psychotherapy sessions that patients attend and the likelihood that patients will reengage with psychotherapy after a period of absence.

Although studies suggest that patients find mobile technology adjuncts to treatment acceptable and useful [13,15], the impact of text messaging as an adjunct to depression treatment has shown mixed results. One study found that nonautomated text messaging support, in addition to telephone-based psychotherapy, was not related to outcomes (eg, reductions in depressive symptoms or psychotherapy attendance) [18]. In another study, patients with mood and anxiety disorders were randomized to receive psychotherapy appointment reminder texts or no appointment reminders. Although receipt of reminder texts failed to decrease overall rates of psychotherapy nonattendance, patients receiving these texts were less likely to be categorized by their therapists as having prematurely dropped out of psychotherapy [19]. A third study sent messages to women with postpartum depression, but did not report the impact of messaging on outcomes [20]. Agyapong [21] sent supportive text messages to individuals in an inpatient alcohol use disorder and comorbid depression program, and found that patients receiving supportive text messages experienced lower posttreatment Beck Depression Inventory (BDI) scores than those receiving treatment-as-usual, but the effects did not hold 3 months after text messages were terminated. Although studies have sought to assess differences in key symptom outcomes, they have less often assessed impacts on engagement in mental health interventions. Even if outcomes are not significantly improved with the addition of an mHealth adjunct, engagement may improve, thereby providing more benefit to more people by decreasing attrition.

Aim of This Study

The aim of this study was to test whether a text messaging adjunct (mood monitoring text messages, treatment-related text messages, and a clinician dashboard to display patient data) increases engagement and improves clinical outcomes in a group CBT treatment for depression. Specifically, we aim to assess whether the text messaging adjunct led to an increase in group therapy sessions attended, an increase in duration of therapy attended, and reductions in Patient Health Questionnaire-9 item (PHQ-9) symptoms compared with the control condition of standard group CBT in a sample of low-income Spanish speaking Latino patients. We expect that receiving text messages would result in higher attendance and improved depression outcomes.

Methods

Recruitment

Patients utilizing outpatient services at an urban public hospital were referred to a behavioral health clinician by their primary care provider when there were concerns about depression due to qualitative symptom expression or a positive screen based on the PHQ-9 [22], a commonly used depression measure in primary care. Patients were seen by a behavioral health clinician following their primary care visit or contacted by phone if seen after hours. Patients were considered eligible for group therapy for depression if they had a PHQ-9 score of 10 or above at the time of initial assessment by the behavioral health clinician. Exclusion criteria for group treatment were active suicidal

XSL•FO
RenderX

ideation with a plan and active, severe psychosis. The behavioral health clinician provided a brief behavioral intervention and a referral to group CBT if the patient met the above criteria.

This study used a nonrandomized design to allocate patients into a texting intervention or nontexting control group. Patients were administered consent, baseline questionnaires, and enrolled in the study during each patient's first group therapy session attended, therefore data for patients eligible but not interested in or declining treatment is unavailable. We intended to conduct a randomized controlled trial (RCT), but we encountered organizational and patient level barriers similar to barriers in other projects in low-income public sector settings [23]. At the organizational level, study materials, such as a paper or an online randomization table, were not always accessible to clinicians. At the patient level, scheduling challenges often precluded patients from attending at the day and time they were assigned. Since the intervention was an adjunct to standard care, it was not possible to deny treatment if a patient was not able to attend the group to which they were initially randomized and were instead allowed to attend the group that fit best with their schedule. Additionally, referring clinicians assigned patients to groups based on the need to balance group size, which may impact group dynamics [24].

Patients were neither incentivized to start treatment, attend sessions, nor complete surveys. Neither the therapists and patients nor research assistants were blinded since they participated in the delivery of treatment and data collection. All procedures and materials were approved by the University of California, San Francisco Institutional Review Board Committee.

Intervention

Group Cognitive Behavioral Therapy

Patients referred to group CBT were allocated to one of the two study conditions: group CBT without the text messaging adjunct or group CBT with the text messaging adjunct. Patients in both study conditions participated in a weekly manualized cognitive behavioral group treatment based on an adapted version of the Building Recovery by Improving Goals, Habits, and Thoughts (BRIGHT) manual for depression [25]. The treatment manual was developed in English and Spanish for use in public sectors settings and has been found to be an efficacious treatment for depression in this population [26]. The manual is divided into four, four-week modules largely focusing on cognitive restructuring (thoughts), behavioral activation (activities), interpersonal relationships (people), and healthy behaviors (health). Treatment was delivered by two therapists at a time (total of two clinical psychologists, and two licensed clinical social workers). Therapists in both conditions delivered culturally sensitive care, leading the groups in Spanish, and all had years of experience providing therapy to low-income Latinos.

Psychotherapy was structured as two continuously running groups (a texting and a control group), and was designed to last 16 continuous weeks for each patient, with group sessions offered weekly. Patients were admitted to the groups on a rolling basis to minimize wait times. Group size in any given week varied from as few as 1 patient to as many as 8, with a median group size of 3 patients. Although the manual was structured for 16 weeks of therapy, some patients attended for more than 16 sessions, or longer than a 16-week period, if they began mid-module or had an extended absence in the middle of treatment. However, our analyses focus only on the 16-week period after initial participation in treatment, as this time frame represented potential completion of all content within the BRIGHT manual.

Both conditions used the same BRIGHT manual and differed only in the exclusion or inclusion of paper-based, weekly mood monitoring worksheets and skill-building homework projects. While the nontexting manual included the paper-based homework sheets, patients in the texting condition were instead sent a series of text messages for mood tracking and skill-building in order to complete "real-world" practice.

Text Messaging Adjunct

Patients were instructed on how to reply to the messages upon enrollment. Patients in the texting condition received up to five types of automated text messages (detailed in Figure 1): (1) a daily mood rating prompt (and feedback 20% of the time based on their mood ratings), (2) a daily message supplementing live therapy content (ie, a thematic module-based message), (3) optional daily medication reminders, (4) a weekly reminder to attend psychotherapy, and (5) a monthly opt-out message to terminate message delivery, if desired. Messages were automatically delivered to patients on a predetermined schedule, with the exception of patients deciding whether and when to receive medication reminders. All text messages were delivered in Spanish (see Figure 1).

The daily mood rating prompt was an integral part of the intervention designed to promote mood-monitoring and increase mood state awareness, a critical element of depression treatment [27]. Automated mood rating messages also allowed for the collection of real-time data in each patient's daily environment for a clinician to review. Patients were given feedback on their mood responses a random 20% of the time they sent a mood response (see Figure 1).

An additional daily text message was delivered reinforcing the theme of the therapy session that week (ie, a thematic module-based message). The content for this message was developed from the BRIGHT manual [25] and focused generally on cognition, self-monitoring, behavioral activation, interpersonal interactions, and healthy behaviors affecting mood.

Patients in the texting condition received up to two types of reminder messages—optional daily medication reminders and weekly reminders to attend group psychotherapy appointments, the content of which is described in Figure 1. Additionally, patients in both conditions were reminded to attend therapy by phone when they began and thereafter only when group sessions were canceled, such as during holidays.

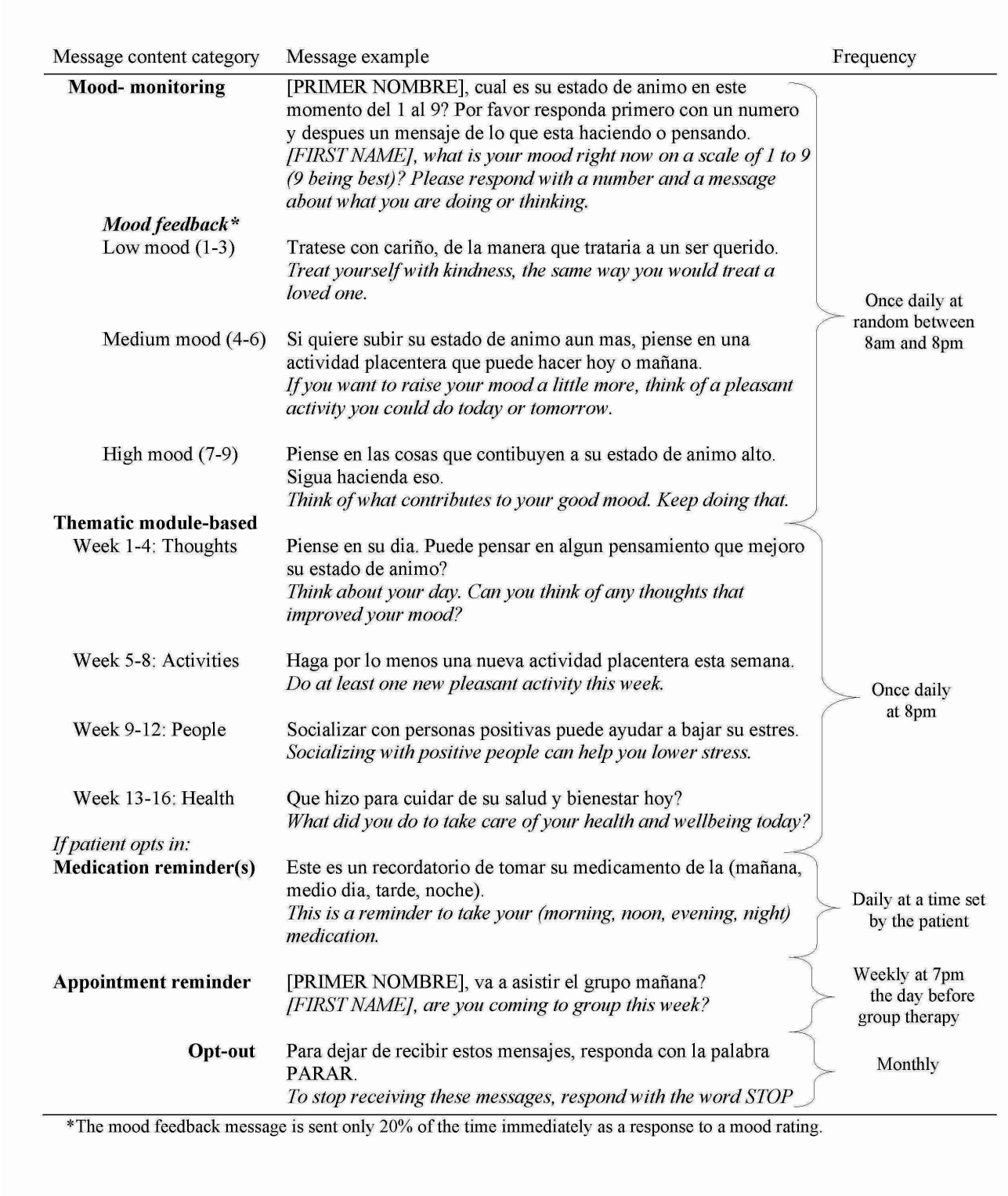
Messages were delivered through an automated text messaging platform, HealthySMS developed for the study by the first author (Multimedia Appendix 1). HealthySMS was designed as a tool for clinicians delivering therapeutic interventions to schedule automated and real-time message delivery. The

XSL•FO
RenderX

platform allows clinicians to track patient progress during treatment and monitor responses to the text message prompts in between sessions (Figure 2). In particular, patient mood data from HealthySMS was projected onto a whiteboard at the beginning of each therapy session to assess patients' mood states in the past week and to apply the tools of therapy to specific events with an emphasis on low and high points.

Messages were sent to patients after terminating therapy for 6 additional months or until they opted out. The content delivered after treatment was equivalent to the content during treatment differing only in the randomization of message order. Patients were instructed that they could opt-out of text messaging at any time by texting the word "STOP" or "PARAR," in Spanish, or by notifying a staff member.

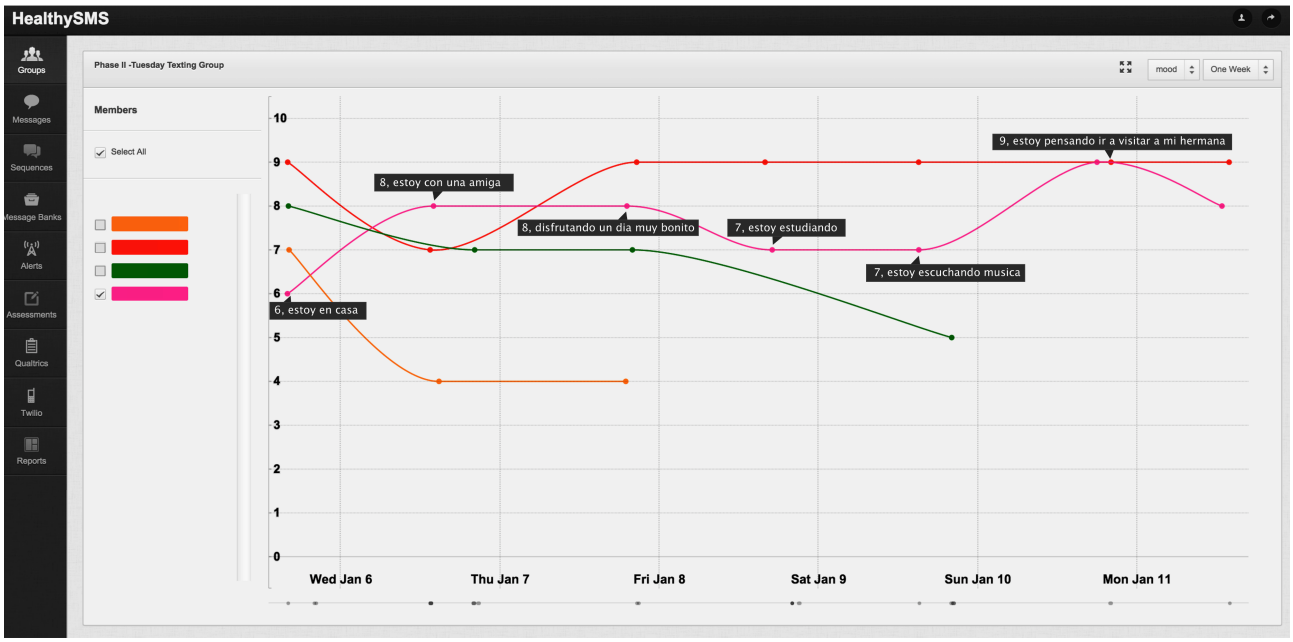
Figure 1. Sample text messages received by patients in the texting condition during depression treatment translated to English.



*The mood feedback message is sent only 20% of the time immediately as a response to a mood rating.

XSL•FO
RenderX

Figure 2. Sample mood graph from HealthySMS used to review in between session mood.



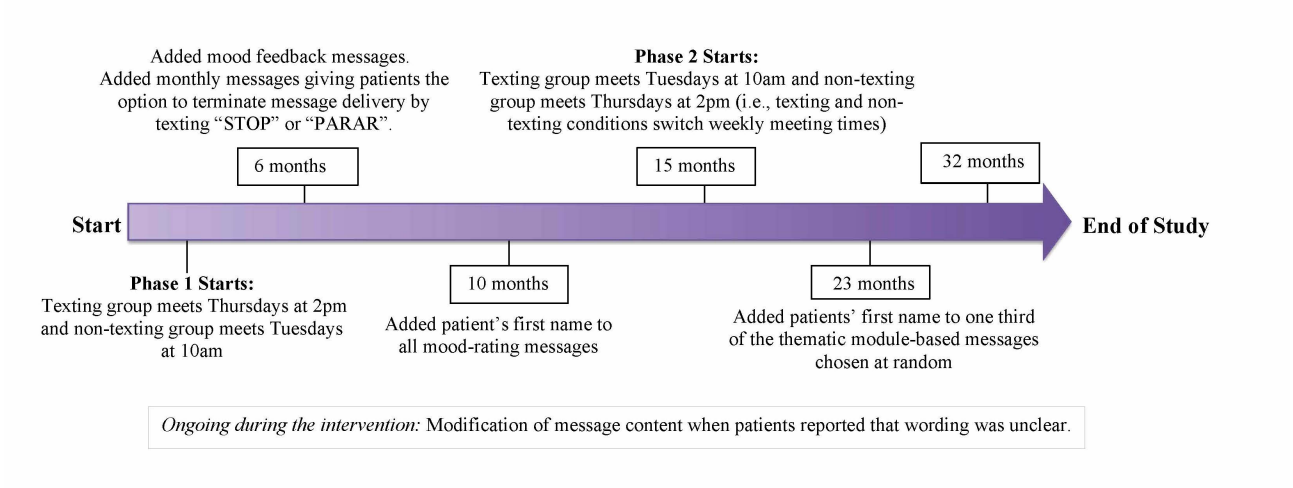
Phases and Iterations of the Intervention

Group therapy sessions held earlier in the day may be more difficult for patients to attend due to depression symptoms, patient characteristics, or employment, therefore two phases (Phase 1 and Phase 2) of the study were conducted in order to switch the day of week and time of day at which the texting and control condition groups took place. In Phase 1 of the study, the texting group (n=21) was held on Thursday afternoons at 2pm and the control group (n=25) was held on Tuesday mornings at 10am (see Figure 3). In Phase 2, the conditions switched meeting times, with the texting group meeting on Tuesday mornings (n=24), and the control group meeting on Thursday afternoons (n=15). Patients attended psychotherapy in only one of the two phases of the study, so each patient attended psychotherapy on only one day of the week, and was exposed to only one treatment condition. Although the same primary therapist (a clinical psychologist) conducted all groups, the cotherapist in both phases of the study differed.

Iterations were also made to improve the text messaging program (Figure 3) based on ongoing feedback from patients and knowledge acquired from emerging mHealth research aimed at improving usability [28]. None of the changes impacted the core elements of CBT or purpose of the study. This approach allowed for the continuous incorporation of knowledge acquired during the intervention, increasingly considered important in research involving behavioral intervention technologies [29,30].

There were no adverse events reported during the study. The HealthySMS platform scanned messages for indications of suicidal ideation by identifying keywords (eg, suicide, kill, die, jump, bridge, and so on) and would notify the principal investigator if any keywords appeared. In total, 24 instances of alerts were triggered, but none of the messages indicated actual suicidal ideation.

Figure 3. Iterations made to the text message content during the intervention.



XSL•FO
RenderX

Measures

Participants completed a baseline questionnaire at their first therapy session that collected demographic data, familiarity with mobile phone use, and measures assessing depression and anxiety symptoms, medication adherence, and depression self-efficacy. For the purposes of this study, only demographic and mobile phone use variables, as well as depression symptoms, are reported.

Attendance

Attendance was recorded at each scheduled group session. Attendance was coded as 1 if the patient was present at the session and 0 if they did not attend. Two key measures of attendance were derived: (1) the total number of psychotherapy sessions patients attended in the first 16 weeks of therapy and (2) the number of weeks patients stayed in treatment before dropping out, up the 16th session of treatment. We included attendance data up until the 16th session of treatment, as this time frame represented potential completion of all content within the treatment manual. The total number of psychotherapy sessions attended was calculated by taking the simple sum of the sessions patients attended over the 16-session period. The number of weeks patients stayed in treatment was calculated by counting the number of sessions that elapsed between the first and the last session the patient attended in this same time frame. A patient was considered to have dropped out of therapy when he or she failed to return to any future therapy session offered.

Depressive Symptoms

Depressive symptoms were measured using the 9-item Patient Health Questionnaire [22,31]. The reliability and validity of the PHQ-9 has been demonstrated in Latino samples [32]. The PHQ-9 was administered at baseline and at every session that a patient attended. Therapists and a research assistant aided patients in completing the questionnaires, first by using paper and pencil measures and later by using a digital equivalent on an iPad. The PHQ-9 was not administered when patients missed a therapy session.

Analytic Plan

Given the inability to randomize to condition, we first tested whether the texting and control conditions differed on any key characteristics at baseline, including depressive symptom severity, familiarity with mobile technologies, and basic demographic characteristics. We then tested the central research questions—whether the conditions differed in their total number of sessions attended, time until dropout, and degree of depressive symptom recovery over the first 16 weeks of treatment.

Condition differences in total sessions attended and weeks in treatment before dropout from psychotherapy were tested using Wilcoxon-Mann-Whitney *U* test for independent samples with patients stratified by study phase. This nonparametric approach was chosen to avoid normality assumptions on the underlying distributions, given that the distributions of the attendance variables were nonnormal. Because there was an unequal distribution of patients to condition across the two study phases, we stratified by phase to conservatively control for any differences that may have been confounding between the study phases. Statistical analyses were performed using the asymptotic assumption on the test statistics [33].

Condition differences in depressive symptom recovery were tested using linear mixed models. The data consisted of repeated measures of the PHQ-9 (up to 16, corresponding to the 16 weeks of therapy), nested within patients. To model dependencies in the same patient's PHQ-9 scores across time, we used a two-level hierarchical model, modeling patient-specific (ie, random) intercepts and patient-specific slopes for week of therapy. The model included condition, week of therapy, and the interaction of condition and week of therapy as predictors, and weekly PHQ-9 scores as the outcome. The condition×week of therapy term allowed for an assessment of whether the texting condition experienced greater depressive symptom improvement over the course of therapy than the control condition. As with the attendance outcomes, we also controlled for study phase to conservatively control for confounds linked to this variable. Analyses utilized all data points for participants (j) and occasions (i), where neither the response y_{ij} (ie, weekly PHQ-9 scores) nor the covariates x_{ij} were missing. For patients who dropped out of therapy before their 16th session, all and only those data collected before the date of attrition were included in these models.

Initial analyses utilized an intent-to-treat approach, including all patients, with the exception of patients who had participated in prior group CBT studies in the clinic. These "returning" patients were excluded from the analyses due to concern that their prior familiarity with the treatment and texting protocol might bias their data. We compared these intent-to-treat findings with results of "active texters." Active texters refers only to patients who utilized the texting adjunct at least once (we excluded the data of patients assigned to the texting condition who either did not receive any texts, or did not respond to any texts received during the first 16 weeks of therapy). To differentiate these analyses from the intent-to-treat analyses, we refer to them as "active-texting" analyses. Active texting analyses were conducted to better isolate the potential effect of the texting adjunct with actual users of the intervention. Patients included in the control condition remained constant over both the intent-to-treat and the active-texting analyses. All statistical analyses were performed using Stata—the attendance analyses were implemented with the vanelteren package for conducting stratified Wilcoxon-Mann-Whitney *U* test, and the depression analyses were implemented via the xtmixed command.

Results

Participants

A total of 91 patients were enrolled in the study between January 2014 and August 2016, of which 48 were allocated to the texting condition and 43 were allocated to the control condition. We excluded 6 patients (3 from each condition) from all analyses because they had participated in prior group CBT studies within this same behavioral health clinic. This exclusion resulted in 85 patients, 45 in the texting condition, and 40 in the control. A total of 6 patients in the texting condition were further excluded from the active-texting analyses, 2 because they sent no text messages during the first 16 weeks of therapy, and

XSL•FO
RenderX

another 4 because they did not respond to a single message received during this period. This additional exclusion resulted in 39 active texting patients in the texting condition, and 40 in the control.

Baseline Data

The baseline characteristics of patients (Table 1) did not differ significantly between the two conditions. Patients were predominantly female and middle aged, with a relatively low level of education (81% of patients [68/84] lacked a high school diploma or equivalent). Slightly less than half of all patients reported being in therapy for depression in the past (44%, 37/85) and similar numbers reported taking medication for depression at baseline (45%, 38/85). The majority of patients in both conditions owned a mobile phone (94%, 80/85), and were familiar with how to use text messaging (73%, 57/78).

The following variables have missing data for 1-2 patients—marital status, education, use of SMS in prior month, preferred method of contact.

Outcomes

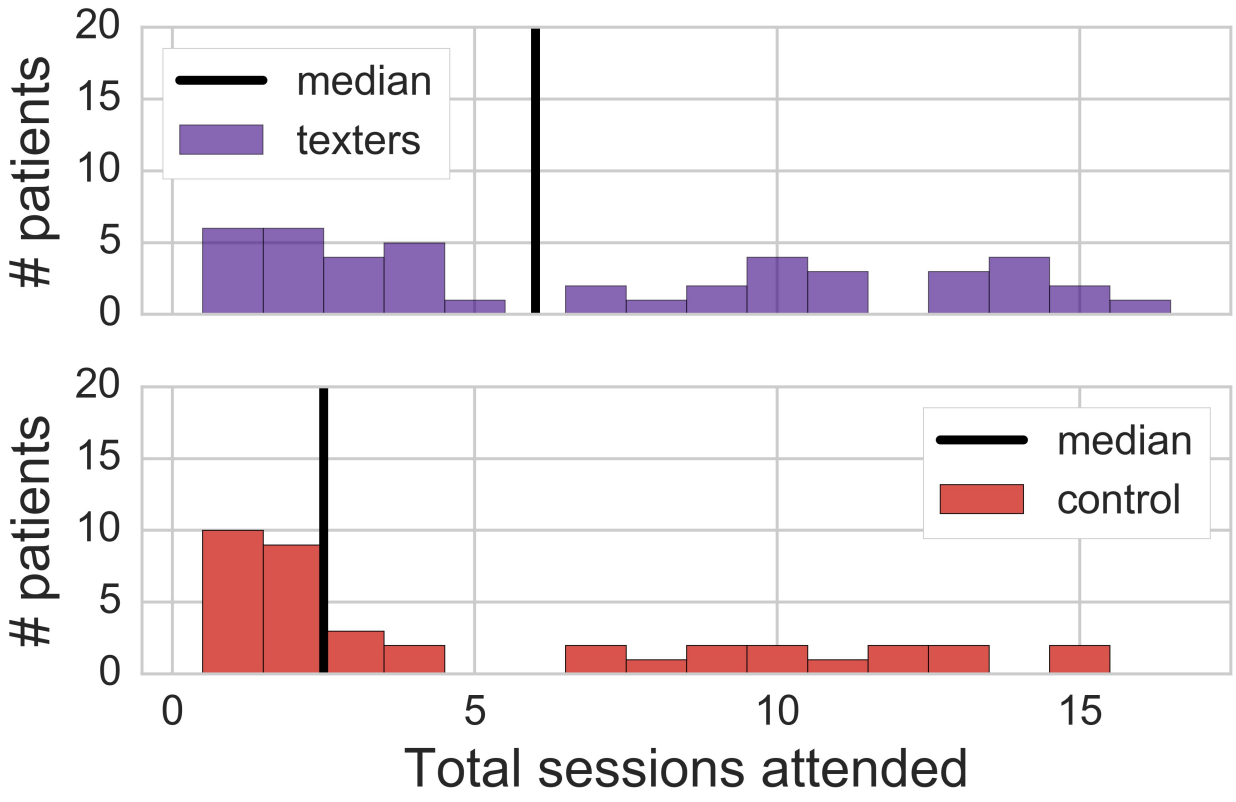
Condition Differences in Attendance Patterns

For the attendance analyses, an additional 3 patients (1 in the intervention and 2 in the control) were excluded, because the study ended before these patients had been offered 16 sessions of therapy. Thus, the total number of sessions attended by these three patients and their time in therapy before dropout was not comparable to that of other patients. This exclusion resulted in 82 patients in the intent-to-treat analyses (44 in the texting condition and 38 in the control) and 76 patients in the active-texting analyses (with 38 patients in each condition).

Total Sessions Attended

Intent-to-treat analyses indicated that patients assigned to the texting condition generally attended more sessions across the first 16 weeks of therapy than patients assigned to the control condition, but the effect was not statistically significant when patients were stratified by phase (Wilcoxon-Mann-Whitney, $z=-1.65$, $P=.10$). The median number of sessions attended by patients assigned to the texting condition was 6, whereas the median number of sessions attended by control patients was 2.5 (Figure 4). The effect was marginally significant when only the data of active texters was compared with that of the control condition (Wilcoxon-Mann-Whitney, $z=-1.91$, $P=.06$), with active texters attending a median of 7.5 sessions in the first 16 weeks of therapy.

Figure 4. Condition differences in total sessions attended. Figures display intent-to-treat analyses.



XSL•FO
RenderX

Table 1. Baseline characteristics in the texting and control conditions. Statistics reflect intent-to-treat analyses.

Characteristics	Texting condition (n=45)	Control condition (n=40)	*P* value
Age (years), mean (SD[a])	51.71 (11.55)	51.83 (11.73)	.96[b]
Gender: female, n (%)	38 (84.44)	29 (72.50)	.18[c]
Highest education, n (%)			.11[c]
No formal education	3 (6.67)	4 (10.26)	
1st to 5th grade	12 (26.67)	7 (17.95)	
6th to 8th grade	12 (26.67)	12 (30.77)	
Some high school	14 (31.11)	4 (10.26)	
High school grad or GED[d]	2 (4.44)	3 (7.69)	
Some college	1 (2.22)	6 (15.38)	
College graduate	1 (2.22)	2 (5.13)	
Graduate or professional school	0 (0.00)	1 (2.56)	
Employment status, n (%)			.73[c]
Full-time	7 (15.56)	4 (10.00)	
Part-time	7 (15.56)	4 (10.00)	
Homemaker	3 (6.67)	3 (7.50)	
Unemployed	12 (26.67)	13 (32.50)	
On disability	10 (22.22)	12 (30.00)	
Retired	4 (8.89)	4 (10.00)	
Other	2 (4.44)	0 (0.00)	
Marital status, n (%)			.76[c]
Single	17 (38.64)	14 (35.00)	
Married or partnered	13 (29.55)	13 (32.50)	
Divorced or separated	7 (15.91)	9 (22.50)	
Widowed	7 (15.91)	4 (10.00)	
Depression measures			
PHQ-9[e], mean (SD)	13.36 (5.96)	13.13 (4.99)	.85[b]
Prior therapy for depression, n yes (%)	21 (46.67)	16 (40.00)	.54[c]
Medication for depression, n yes (%)	22 (48.89)	16 (40.00)	.41[c]
Prior hospitalization for depression, n yes (%)	5 (11.36)	7 (17.50)	.42[c]
Mobile phone ownership, n yes (%)	43 (95.56)	37 (92.50)	.55[c]
Use of text-messaging in prior month, n yes (%)[g]	34 (79.07)	23 (65.71)	.19[c]
Preferred method of contact[f], n (%)			.74[c]
Call	27 (61.36)	26 (66.67)	
Text	6 (13.64)	6 (15.38)	
Depends	11 (25.00)	7 (17.95)	

[a]SD: standard deviation.

[b]Indicates that a *t* test was used to test for condition differences.

[c]Indicates that a chi-square test was used to test for condition differences.

[d]GED: general educational development.

XSL•FO
RenderX

[e]PHQ-9: Patient Health Questionnaire-9 item.

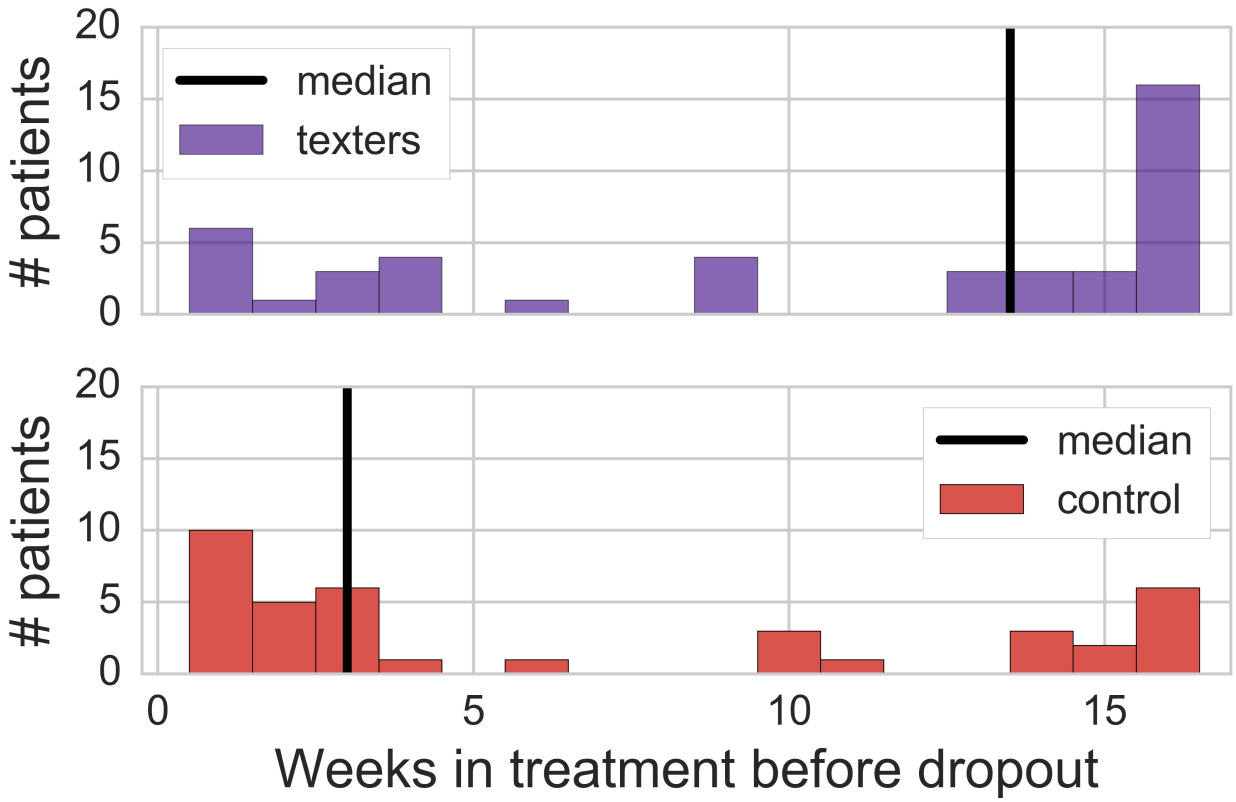
[f]Preferred method of contact refers to the question "In general, if someone needs to reach you, do you prefer that they call or that they text you?"

[g]Question asked only to the patients who reported owning a mobile phone.

Time Until Dropout

Intent-to-treat analyses indicated that patients assigned to the texting condition stayed in therapy significantly longer before dropping out than patients assigned to the control condition when stratified by phase (Wilcoxon-Mann-Whitney, z=−2.21, P=.03). Patients assigned to the texting condition stayed in therapy for a median of 13.5 weeks before dropping out, whereas patients in the control conditions stayed in therapy for a median of only 3 weeks before dropping out (Figure 5). Results were substantively similar when contrasting the duration data of active texters (median of 14 weeks until dropout) to that of patients in the control condition (Wilcoxon-Mann-Whitney, z=−2.28 , P= . 02).

Figure 5. Condition differences in weeks in treatment until patient dropout. Figures display intent-to-treat analyses.



Condition Differences in Depression Symptom Recovery

Intent-to-treat analyses indicated that, as anticipated, depressive symptoms significantly declined in both the texting condition (B=−.29, 95% CI −0.38 to −0.19, z=−5.80, P<.001) and the control condition (B=−.20, 95% CI −0.32 to −0.07, z=−3.12, P=.002) over the first 16 weeks of therapy. These coefficients indicate that, on average, the texting condition's PHQ-9 scores decreased 0.29 points for every week that patients were enrolled in therapy, whereas the control condition's PHQ-9 scores decreased an average of 0.20 points per week. However, the magnitude of depressive symptom improvement in the texting condition was not significantly greater than that of the control (B of the condition×therapy week interaction=−.09, 95% CI −0.25 to 0.07, z=−1.09, P=.27). Similarly, active texting analyses failed to demonstrate greater improvements in depressive symptom recovery among active texters as compared with the control condition (B of interaction=−.10, 95% CI −0.27 to 0.06, z=−1.25, P=.21). Descriptive statistics on PHQ-9 scores by condition corresponding to each the 16 psychotherapy sessions are provided in Table 2.

XSL•FO
RenderX

Table 2. Means, standard deviations, and sample size for the Patient Health Questionnaire-9 item (PHQ-9) by condition across psychotherapy. Statistics reflect intent-to-treat analyses.

Psychotherapy session	Texting condition Mean PhQ-9 (SD[a])	n	Control condition Mean PhQ-9 (SD)	n
1 (baseline)	13.36 (5.96)	45	13.13 (4.99)	39
2	9.90 (6.15)	21	10.60 (4.71)	20
3	8.83 (5.31)	23	9.83 (5.60)	18
4	6.94 (5.15)	17	9.07 (4.35)	15
5	9.44 (5.85)	18	9.90 (4.33)	10
6	7.32 (5.61)	19	7.71 (5.31)	14
7	7.56 (5.18)	18	7.85 (5.80)	13
8	6.35 (3.44)	17	10.53 (5.55)	15
9	7.09 (4.12)	23	8.73 (3.29)	11
10	6.83 (6.08)	18	7.93 (6.13)	15
11	7.05 (4.50)	22	7.44 (4.88)	9
12	6.38 (3.84)	13	6.33 (4.85)	9
13	7.84 (6.24)	19	6.80 (4.92)	10
14	5.24 (3.40)	17	9.10 (5.67)	10
15	8.07 (5.74)	15	9.83 (6.55)	6
16	6.00 (4.20)	18	12.67 (6.40)	9

[a]SD: standard deviation.

Discussion

Principal Findings

This study investigated whether a mobile phone-based text messaging adjunct to CBT for depression increased engagement and improved clinical outcomes compared with standard CBT without a text messaging adjunct. We found that receiving text messages during a group CBT intervention led to decreased attrition in a sample of Spanish-speaking Latinos with depression. These findings support the use of text messaging and mHealth interventions as adjuncts to psychotherapeutic treatments in order to reduce attrition. Patients in the texting condition may have felt more engaged in the intervention and the messages may have helped them practice skills throughout the week. It is also possible that receiving text messages helped patients feel more supported and more motivated [13], and thus diminished the symptomology inherent of depression by targeting maladaptive thoughts and behaviors [34] that could prevent patients from attending therapy. At the most basic level, it is possible that the simple act of receiving a weekly reminder encouraged patients to resume therapy attendance even after one or several sessions were missed [14]. It is also possible that the review of patient data in HealthySMS allowed clinicians to provide more personalized interventions based on patient mood and text responses that helped keep patients more engaged.

Despite a relatively small sample size, we explored whether there were significant differences in PHQ-9 ratings between conditions over the course of the intervention. We did not find significant differences in PHQ-9 outcomes between the conditions. Since both conditions received an active treatment and the control condition has been found to be effective in previous studies [26], we would need a larger sample size to detect smaller differences in outcomes. It is also possible that people who did not attend were more depressed, but we were unable to assess PHQ-9 ratings when patients did not attend.

Even though we did not find differences in symptoms between conditions, our findings that patients stayed in treatment longer indicate that the text messaging adjunct can promote sustained engagement with an already efficacious treatment. Results from this study may even be applicable to CBT delivered through other modalities such as the Internet, that suffer from high levels of attrition, by bringing the treatment into users' daily lives. However, attrition is higher and engagement with technology is lower without human support [35]. More research is needed to determine an ideal balance between digital and human intervention.

Other key steps in advancing this area of research would be to determine whether there are broader cost savings by sending automated messages and keeping patients in treatment longer. A key selling point for mHealth technologies is the ability to reduce health care costs [36-38]. Although patients staying in psychotherapy longer is seemingly more costly in the short-term, it is possible that receiving a stronger dose of treatment and having more people complete treatment can improve recovery from depression and reduce associated costs such as losses in productivity and other depression-related societal burdens long-term. This study was conducted in group therapy setting, which is already less resource-intensive than individual therapy,

XSL•FO
RenderX

thus the cost reduction associated with a text message adjunct may be greater.

Despite the recent proliferation of interventions using technology-based symptom monitoring via mobile smartphone apps and passive sensing [39], text messages (and other mobile messaging apps) remain a low cost and simple way of collecting data from patients with low technological ability who may otherwise have a difficult time logging symptoms in apps or websites [15,40]. Despite the automation of messages, they also offer a personal connection to another individual, whether it is perceived or real [13].

Generalizability

The study included low-income, Spanish-speaking Latinos. It is unclear how these findings translate to other populations. Research has shown that this population generally has lower engagement in mental health interventions [5], therefore it was appropriate to test the texting adjunct with this population. A strength of our study is that it occurred as part of standard clinical practice in a patient pool with multiple chronic illnesses that is typically less responsive to interventions and is thus more likely to successfully generalize to other clinical settings. It is likely that our findings in this population that is less educated, less technologically savvy, and more difficult to engage may generalize to more diverse and higher socioeconomic status (SES) populations. It is also possible that mHealth and text messaging adjuncts could improve other individual and group psychotherapies from other modalities or focusing on something other than depression, especially if they rely on completion of between session homework and skills practice.

Limitations

This study was not fully randomized, despite our initial intention. Furthermore, the lack of randomization opens up the possibility of third variables being responsible for any group differences. However, given that there were no baseline differences between the groups, the possibility of confounds may be low. Another limitation is that clinicians and data collectors were not blind to conditions, which opens up the possibility of bias. Last and most important, we were unable to assess depressive symptoms on days patients missed treatment. This particularly limits the interpretation of no differences in outcomes between the groups. It is possible that people stopped attending either because they were highly symptomatic or vice versa because they felt recovered. Future studies should ensure that assessments are collected regardless of therapy session attendance.

Future Directions

Future research should more specifically study mechanisms of action to improve the efficiency and effectiveness of mHealth treatment adjuncts. Future studies should also assess the impact of text messaging and mHealth adjuncts on symptoms in a larger sample size to determine if there is clinically significant improvement compared with standard group CBT. They should also make sure to collect symptom data from nonengaged patients. It is possible that gains from therapy can be maintained longer if patients continue to receive text messages that encourage them to practice skills thought to be active ingredients in improvements. Along with research on overall efficacy and effectiveness of mHealth interventions, studies should assess how to utilize incoming data to predict key clinical events. For example, analyses of daily mood data found that lower mood the day before a therapy session resulted in lower likelihood of attendance [41]. These types of analytics can inform just in time interventions to improve mHealth and in person interventions. Finally, cost-effectiveness analyses can help determine the relative value of increasing attendance to psychotherapy sessions and whether that improves outcomes long-term, resulting in lower health care costs overall.

Conclusions

Our study shows that a text messaging intervention used as an adjunct to psychotherapy for depression can improve engagement in treatment. We found that patients who received text messages dropped out later in treatment compared with patients receiving standard CBT treatment. By testing this intervention in a low-SES Latino population, it may also generalize to other populations that have been difficult to engage in mental health treatment. As the focus of translational science moves to improve the implementation of efficacious interventions for the broad benefit of the public's health, mobile interventions as adjuncts to treatment are emerging as valuable tools.

Acknowledgments

AA was supported by National Institutes of Mental Health—K23MH094442 and a Robert Wood Johnson New Connections Award. Publication made possible in part by support from the Berkeley Research Impact Initiative (BRII) sponsored by the UC Berkeley Library. We would like to thank Julia Bravin, Omar Contreras, Sandra Larios, PhD, Marta Perez, LCSW, and Laura Pullen, LCSW, for their work on this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Example of clinician dashboard from HealthySMS.

[JPEG File, 1MB - jmir_v19i5e148_app1.jpeg]

XSL•FO
RenderX

References

1. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [Medline: 23870719]
2. Miranda J, Azocar F, Organista KC, Dwyer E, Areane P. Treatment of depression among impoverished primary care patients from ethnic minority groups. Psychiatr Serv 2003 Feb;54(2):219-225. [doi: 10.1176/appi.ps.54.2.219] [Medline: 12556604]
3. Miranda J, Chung JY, Green BL, Krupnick J, Siddique J, Revicki DA, et al. Treating depression in predominantly low-income young minority women: a randomized controlled trial. JAMA 2003 Jul 2;290(1):57-65. [doi: 10.1001/jama.290.1.57] [Medline: 12837712]
4. Alegría M, Chatterji P, Wells K, Cao Z, Chen C, Takeuchi D, et al. Disparity in depression treatment among racial and ethnic minority populations in the United States. Psychiatr Serv 2008 Nov;59(11):1264-1272 [FREE Full text] [doi: 10.1176/appi.ps.59.11.1264] [Medline: 18971402]
5. Organista KC, Muñoz RF, González G. Cognitive-behavioral therapy for depression in low-income and minority medical outpatients: description of a program and exploratory analyses. Cogn Ther Res 1994 Jun;18(3):241-259. [doi: 10.1007/BF02357778]
6. Kazantzis N, Whittington C, Dattilio F. Meta-Analysis of homework effects in cognitive and behavioral therapy: a replication and extension. Clin Psychol: Science and Practice 2010;17(2):144-156. [doi: 10.1111/j.1468-2850.2010.01204.x]
7. Hans E, Hiller W. A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. Clin Psychol Rev 2013 Dec;33(8):954-964. [doi: 10.1016/j.cpr.2013.07.003] [Medline: 23988455]
8. Neimeyer RA, Kazantzis N, Kassler DM, Baker KD, Fletcher R. Group cognitive behavioural therapy for depression outcomes predicted by willingness to engage in homework, compliance with homework, and cognitive restructuring skill acquisition. Cogn Behav Ther 2008;37(4):199-215. [doi: 10.1080/16506070801981240] [Medline: 18608311]
9. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behavior Therapy 2016;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002]
10. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The relationship between homework compliance and therapy outcomes: an updated meta-analysis. Cognit Ther Res 2010 Oct;34(5):429-438 [FREE Full text] [doi: 10.1007/s10608-010-9297-z] [Medline: 20930925]
11. Hollon S. Does cognitive therapy have an enduring effect? Cognit Ther Res 2003;27(1):71-75. [doi: 10.1023/A:1022538713914]
12. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mhealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
13. Aguilera A, Berridge C. Qualitative feedback from a text messaging intervention for depression: benefits, drawbacks, and cultural differences. JMIR Mhealth Uhealth 2014 Nov 05;2(4):e46 [FREE Full text] [doi: 10.2196/mhealth.3660] [Medline: 25373390]
14. Ogrodniczuk JS, Joyce AS, Piper WE. Strategies for reducing patient-initiated premature termination of psychotherapy. Harv Rev Psychiatry 2005;13(2):57-70. [doi: 10.1080/10673220590956429] [Medline: 16020021]
15. Aguilera A, Muñoz RF. Text messaging as an adjunct to CBT in low-income populations: a usability and feasibility pilot study. Prof Psychol Res Pr 2011 Dec 1;42(6):472-478 [FREE Full text] [doi: 10.1037/a0025499] [Medline: 25525292]
16. Aguilera A, Schueller SM, Leykin Y. Daily mood ratings via text message as a proxy for clinic based depression assessment. J Affect Disord 2015 Apr 1;175:471-474 [FREE Full text] [doi: 10.1016/j.jad.2015.01.033] [Medline: 25679202]
17. Lambert MJ, Harmon C, Slade K, Whipple JL, Hawkins EJ. Providing feedback to psychotherapists on their patients' progress: clinical results and practice suggestions. J Clin Psychol 2005 Feb;61(2):165-174. [doi: 10.1002/jclp.20113] [Medline: 15609358]
18. Furber G, Jones GM, Healey D, Bidargaddi N. A comparison between phone-based psychotherapy with and without text messaging support in between sessions for crisis patients. J Med Internet Res 2014;16(10):e219 [FREE Full text] [doi: 10.2196/jmir.3096] [Medline: 25295667]
19. Clough BA, Casey LM. Using SMS reminders in psychology clinics: a cautionary tale. Behav Cogn Psychother 2014 May;42(3):257-268. [doi: 10.1017/S1352465813001173] [Medline: 24495365]
20. Broom MA, Ladley AS, Rhyne EA, Halloran DR. Feasibility and perception of using text messages as an adjunct therapy for low-income, minority mothers with postpartum depression. JMIR Ment Health 2015;2(1):e4 [FREE Full text] [doi: 10.2196/mental.4074] [Medline: 26543910]
21. Agyapong VI, Ahern S, McLoughlin DM, Farren CK. Supportive text messaging for depression and comorbid alcohol use disorder: single-blind randomised trial. J Affect Disord 2012 Dec 10;141(2-3):168-176. [doi: 10.1016/j.jad.2012.02.040] [Medline: 22464008]
22. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]

XSL•FO
RenderX

23. Rosal MC, White MJ, Borg A, Scavron J, Candib L, Ockene I, et al. Translational research at community health centers: challenges and successes in recruiting and retaining low-income Latino patients with type 2 diabetes into a randomized clinical trial. Diabetes Educ 2010;36(5):733-749. [doi: 10.1177/0145721710380146] [Medline: 20729512]
24. Burlingame GM, McClendon DT, Alonso J. Cohesion in group therapy. Psychotherapy (Chic) 2011 Mar;48(1):34-42. [doi: 10.1037/a0022063] [Medline: 21401272]
25. Miranda J, Woo S, Lagomasino I, Hepner K, Wiseman S, Muñoz R. i4health. 2006. BRIGHT CBT depression group manuals URL: http://i4health.paloaltou.edu/manuals/bright-manuals.html [accessed 2017-04-05] [WebCite Cache ID 6pVAtwdpi]
26. Muñoz RF, Mendelson T. Toward evidence-based interventions for diverse populations: The San Francisco General Hospital prevention and treatment manuals. J Consult Clin Psychol 2005 Oct;73(5):790-799. [doi: 10.1037/0022-006X.73.5.790] [Medline: 16287379]
27. Cohen JS, Edmunds JM, Brodman DM, Benjamin CL, Kendall PC. Using self-monitoring: implementation of collaborative empiricism in cognitive-behavioral therapy. Cogn Behav Prac 2013 Nov;20(4):419-428. [doi: 10.1016/j.cbpra.2012.06.002]
28. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
29. 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]
30. 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]
31. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [Medline: 15550799]
32. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the patient health questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J Gen Intern Med 2006 Jun;21(6):547-552 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00409.x] [Medline: 16808734]
33. Lehmann E, D'Abrera HJ. Nonparametrics: statistical methods based on ranks. San Francisco: Holden-Day; 1975.
34. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression: A Personal Reflection. New York: Guilford Press; 1979.
35. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
36. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
37. Akter S, Ray P. mHealth - an Ultimate Platform to Serve the Unserved. Yearb Med Inform 2010:94-100. [Medline: 20938579]
38. Downer SR, Meara JG, Da Costa AC, Sethuraman K. SMS text messaging improves outpatient attendance. Aust Health Rev 2006 Aug;30(3):389-396. [Medline: 16879098]
39. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
40. Walsh S, Golden E, Priebe S. Systematic review of patients' participation in and experiences of technology-based monitoring of mental health symptoms in the community. BMJ Open 2016 Jun 21;6(6):e008362 [FREE Full text] [doi: 10.1136/bmjopen-2015-008362] [Medline: 27329437]
41. Bruehlman-Senecal E, Aguilera A, Schueller SM. Mobile phone–based mood ratings prospectively predict psychotherapy attendance. Behavior Therapy 2017 Sep;48(5):614-623. [doi: 10.1016/j.beth.2017.01.002]

Abbreviations

BDI: Beck Depression Inventory
CBT: cognitive behavioral therapy
mHealth: mobile health
PCP: primary care provider
PHQ-9: Patient Health Questionnaire-9 item
RCT: randomized controlled trial
SES: socioeconomic status

XSL•FO
RenderX

Edited by R Calvo; submitted 28.10.16; peer-reviewed by S Berrouiguet, S Hoermann; comments to author 07.12.16; revised version received 10.02.17; accepted 13.03.17; published 08.05.17.

Please cite as:
Aguilera A, Bruehlman-Senecal E, Demasi O, Avila P
Automated Text Messaging as an Adjunct to Cognitive Behavioral Therapy for Depression: A Clinical Trial
J Med Internet Res 2017;19(5):e148
URL: http://www.jmir.org/2017/5/e148/
doi:10.2196/jmir.6914
PMID:28483742

©Adrian Aguilera, Emma Bruehlman-Senecal, Orianna Demasi, Patricia Avila. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Web-Based Guided Recommendation Service for a Curated List of Readily Available Mental Health and Well-Being Mobile Apps for Young People: Randomized Controlled Trial

Niranjan Bidargaddi[1,2], BEng (Hons), PhD; Peter Musiat[1,3], PhD; Megan Winsall[1,2], BSc; Gillian Vogl[4], BA (Hons), PhD; Victoria Blake[4], BSocSc; Stephen Quinn[1,5], BMath (hons), MMath, PhD; Simone Orlowski[1,2], BPsyc(Hons); Gaston Antezana[1,2], BPsych (Hons); Geoffrey Schrader[1], MBBS, FRANZCP, PhD

[1]Digital Psychiatry & Personal Health Informatics Group, School of Medicine, Flinders University, Clovelly Park, Australia

[2]Young and Well Cooperative Research Centre, Melbourne, Australia

[3]Section of Eating Disorders, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[4]ReachOut.com, Sydney, Australia

[5]School of Health Sciences, Department of Statistics Data Science and Epidemiology, Swinburne University, Melbourne, Australia

Corresponding Author:
Niranjan Bidargaddi, BEng (Hons), PhD
Digital Psychiatry & Personal Health Informatics Group
School of Medicine
Flinders University
Flinders at Tonsley
1284 South Road
Clovelly Park, 5042
Australia
Phone: 61 872218842
Fax: 61 882012904
Email: niranjan.bidargaddi@flinders.edu.au

Abstract

Background: Mental disorders are highly prevalent for the people who are aged between 16 and 25 years and can permanently disrupt the development of these individuals. Easily available mobile health (mHealth) apps for mobile phones have great potential for the prevention and early intervention of mental disorders in young adults, but interventions are required that can help individuals to both identify high-quality mobile apps and use them to change health and lifestyle behavior.

Objectives: The study aimed to assess the efficacy of a Web-based self-guided app recommendation service ("The Toolbox") in improving the well-being of young Australians aged between 16 and 25 years. The intervention was developed in collaboration with young adults and consists of a curated list of 46 readily available health and well-being apps, assessed and rated by professionals and young people. Participants are guided by an interactive quiz and subsequently receive recommendations for particular apps to download and use based on their personal goals.

Methods: The study was a waitlist, parallel-arm, randomized controlled trial. Our primary outcome measure was change in well-being as measured by the Mental Health Continuum-Short Form (MHC-SF). We also employed ecological momentary assessments (EMAs) to track mood, energy, rest, and sleep. Participants were recruited from the general Australian population, via several Web-based and community strategies. The study was conducted through a Web-based platform consisting of a landing Web page and capabilities to administer study measures at different time points. Web-based measurements were self-assessed at baseline and 4 weeks, and EMAs were collected repeatedly at regular weekly intervals or ad hoc when participants interacted with the study platform. Primary outcomes were analyzed using linear mixed-models and intention-to-treat (ITT) analysis.

Results: A total of 387 participants completed baseline scores and were randomized into the trial. Results demonstrated no significant effect of "The Toolbox" intervention on participant well-being at 4 weeks compared with the control group (P=.66). There were also no significant differences between the intervention and control groups at 4 weeks on any of the subscales of the MHC-SF (psychological: P=.95, social: P=.42, emotional: P=.95). Repeat engagement with the study platform resulted in a

XSL•FO
RenderX

significant difference in mood, energy, rest, and sleep trajectories between intervention and control groups as measured by EMAs (P<.01).

Conclusions: This was the first study to assess the effectiveness of a Web-based well-being intervention in a sample of young adults. The design of the intervention utilized expert rating of existing apps and end-user codesign approaches resulting in an app recommendation service. Our finding suggests that recommended readily available mental health and well-being apps may not lead to improvements in the well-being of a nonclinical sample of young people, but might halt a decline in mood, energy, rest, and sleep.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR): ACTRN12614000710628; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=366145 (Archived by WebCite at http://www.webcitation.org/6pWDsnKme)

(J Med Internet Res 2017;19(5):e141) doi:10.2196/jmir.6775

KEYWORDS

well-being; mental health; young people; online intervention; apps; engagement

Introduction

Background

In Australia, clinical mental disorders are highly prevalent among young adults aged between 16 and 25 years. Approximately one in four suffer from at least one diagnosable mental disorder in the past 12 months [1], and mental disorders account for a quarter of the burden of disease in this age group [2]. With similar prevalence rates globally (eg, in Europe, Africa, the United States, and Asia) adolescent mental health is an international public health challenge [3]. Adolescence is a crucial developmental stage for the individual and disruption to mental health during this stage can have far-reaching effects, and whose full personal and socioeconomic impact often only becomes apparent at a later stage in life [3]. Thus, effective, engaging, and easy to disseminate strategies that reduce the multiplicative impact of these risk factors within young people are needed.

mHealth Apps

Although technology mediated mental health interventions have frequently been praised for their potential and ease of access, previous research into one-size-fits-all intervention has demonstrated that these interventions have limited appeal and that they have failed to gain traction within health care [4]. It is possible that this limitation can be overcome by delivering interventions using mediums and resources young people are already interacting with and that are tailored to their circumstances. Mobile health (mHealth) apps have great potential for the prevention and early intervention of many physical and mental health problems. To date, there are approximately 165,000 health apps available for Android and iOS mobile phones and tablet devices, approximately 10% of which are mobile apps for mental health problems [5]. However, it has been shown that many easily available mHealth apps are of dubious quality and do not follow evidence-based principles [5]. For example, research into currently available apps for individuals with bipolar disorder found that the majority were not developed in line with best practice clinical guidelines or self-management principles currently used in the treatment of bipolar disorder. Most apps also did not contain source citations or privacy policies, making it difficult for users to assess app quality [6]. An evaluation of mobile apps for mindfulness highlighted that many apps often claim to be for a particular purpose or provide a particular intervention when they in fact do not [7]. Those mobile mental health apps that are based on evidence-based principles and have demonstrated efficacy often have been developed as part of research studies and are not available publically [8]. This suggests that there is a gap between evidence-based research and readily available existing mHealth interventions in the open market. Therefore, to utilize the public health potential of these existing mHealth apps, it is important to identify those apps of high quality and guide individuals in finding interventions that meet their need and likely work.

Rating Apps

With regard to identifying high quality apps, most people use app store ratings as a marker for app quality, as indicated by the correlation between app user ratings and their popularity on the marketplace [9]. However, these ratings mainly reflect subjective experiences from a usability and aesthetics perspective, and not whether the apps are designed with appropriate strategies necessary to improve health outcomes [10]. Obtaining more objective evaluations of mobile apps with regard to their quality is hampered by the fact that even with regard to simple criteria, such as the degree of personalization, the funding source of related research, or data import and export capabilities, interrater reliability is poor [11]. In an attempt to overcome the limitations of user ratings for mobile apps, Stoyanov et al [10] created the mobile app rating scale (MARS), a questionnaire to assess the quality of health apps on the domains engagement, functionality, aesthetics, information, and subjective quality. Whereas this scale is a more objective marker of app quality than app store ratings, it is also resource intensive and requires thorough assessment. As a result, application of this approach on a large scale is still in its infancy due to the large volume of mHealth apps.

With regard to identifying effective apps, research suggests that theory-driven health interventions, that is, interventions employing evidence-based behavior change techniques (BCTs) are more effective than interventions that are not theory-driven [12]. The implementation of such strategies within mobile apps is influenced by how users interact with mobile apps [13]. Mobile phone apps are generally optimized to the way people

XSL•FO
RenderX

engage with such phones and as such tend to implement only some strategies with functionalities that are brief and easy to use. As a result, a single app alone is unlikely to contain all necessary strategies for mental health, and furthermore identical strategies might be implemented in different apps with varying aesthetics and ease of use. Thus, characterizing strategies implemented within an app will be a crucial step to identify effective apps. To ensure that ethical values in health care are met, this characterization process could be facilitated by clinicians or researchers by reviewing the scientific literature, searching apps stores, reviewing app descriptions and reviews, and piloting the app themselves [14]. However, it is important to recognize that individuals are less likely to engage with interventions that implement effective strategies but poor aesthetics and usability, thus resulting in ineffective outcomes [15,16].

One way of overcoming these challenges is to create a repository of high-quality mHealth apps and guide users in the process of identifying effective and engaging ones. The "Beacon" website is one such resource developed in Australia that categorizes, reviews, and rates websites and mobile apps for mental and physical disorders [17]. Recent findings suggest participants are willing to use several apps when they are recommended a range of custom-selected apps with different behavioral strategies [13]. The challenges of this approach have been highlighted in the UK's National Health Service's attempt of creating a curated app repository for patients with chronic health conditions. Many of the apps were found to transmit sensitive data without the knowledge of the user [18] or provide clinically questionable advice [19] that resulted in the app library to be closed. In addition, the majority of apps are generally identified and downloaded by users directly from apps stores and the keywords people use when searching for specific health apps do not necessarily yield the most appropriate or effective app [20]. Instead they might be reflective of these words appearing in place like app name, text used in description of the apps combined with the user rated popularity of the apps, none of which alone are markers of quality. Developing a nuanced understanding of search patterns for mental health and well-being problems will be crucial to facilitate access to apps of high quality.

Impact Assessment

Methods to measure the impact of app usage may require different approaches, given the way people tend to interact with apps, usually for short periods of time, on a regular or irregular basis. An alternative to traditional questionnaire-based measures of mental health and well-being, ecological momentary assessment (EMA) [21] may be a suitable means of detecting short-term changes with regard to parameters, such as mood or sleep on a day-to-day basis.

In this paper, we report on findings from a waitlist randomized controlled trial (ACTRN 12614000710628) [22] which was designed to test the efficacy of a guided recommendation service for readily available mobile mental health apps for young people aged 16-25 years. Whereas our primary outcome measure was the well-validated Mental Health Continuum-Short Form (MHC-SF), we employed ecological momentary assessments to track mood, sleep, and energy.

Methods

Overview

We conducted a Web-based parallel-arm randomized controlled trial comparing "The Toolbox," a guided app recommendation service, to a waiting list control group. Web-based measurements were assessed at baseline and 4 weeks, and ecological momentary assessments were collected repeatedly at regular weekly intervals or ad hoc when participants interacted with the study platform. Details of study design, intervention and control conditions, outcome measures, and sample sizes are reported extensively in the previously published study protocol [22]. A brief overview of the study is outlined in the following section. The study received ethical approval by the Social and Behavioural Research Ethics Committee of Flinders University (registration number 64780) and is registered in the Australian New Zealand Clinical Trials Registry (ACTRN: 12614000710628). It also gained ethical approval for recruitment by the Department of Education and Child Development of South Australia (DECD CS/14/511-23).

Recruitment

Participants were recruited from the general young adult (16-25 years) population across Australia, with access to a computer or mobile phone and the Web. Preexisting mental health conditions were not considered as exclusion conditions for this study. Several Web-based and community strategies, either paid or unpaid, were utilized for recruitment. The recruitment message was formulated around overall well-being and health (and not on illnesses): Examples of such messages included:

> *Want to improve your energy and fitness? Find out what your wellbeing looks like and use apps to achieve your goals.*
>
> *Better health & fitness: Monitor your wellbeing, set goals, & access health & fitness apps.*
>
> *Summer fun taking its toll? Track your wellbeing & download apps for mind+body.*

Web-based paid advertisements were placed on Facebook, Twitter, YouTube, and Google AdWords from November 19, 2014 to March 12, 2015. The keywords for the advertisements were defined in collaboration with a reference group representing the target population to ensure their validity and relevance. Examples of keywords included fitness, stress, relationships, balance, and goals. A total of 12 advertisements were placed across the previously mentioned platforms, with an average duration of 21 days per advertisement. The paid strategy also included recruitment through a recruitment agency for clinical trials. The agency referred individuals in the target demographic to the study website over a period of 2 months (July 8, 2015 to September 2, 2015). In addition to paid advertisements, links to the study site were provided to 39 organizations and educational institutions frequently visited by young people from different backgrounds in Australia (most notably the partner organizations of the Young and Well Cooperative Research Centre) to integrate into their websites

XSL•FO
RenderX

and promote via their social media channels (Facebook, Twitter). Community-based organizations such as schools, universities, sporting clubs, and local councils in one rural region of South Australia were approached and asked to help promote the study. Promotional packages comprising a video, information, and instructions on how to access the Online Wellbeing Centre (OWC) were distributed to 32 institutions and community contacts and presented to potential participants.

Participants with informed consent and aged between 16 and 25 years were included, and parental consent was obtained if participants were recruited from community organizations and were below 18 years old. Using unique links, data was collected to objectively identify the recruitment source for each participant. The yield per strategy and characteristics of participants between channels are reported elsewhere [23].

Procedures

The study was conducted through an OWC platform consisting of a landing Web page and capabilities to manage consent, sign up, randomize, administer study measures at different time points, monitor engagement, and provide feedback to users in a meaningful graphic display. Study advertisements linked participants to the landing page of the OWC which contained a brief overview of the study, detailed participant information sheet, and a Web-based consent form.

After completing the Web-based consent form, participants completed a registration form. The OWC software randomized the participants and sent them their login details, either via email or SMS (short message service). Participants logged into the OWC to complete study measurements and if they were assigned to the intervention arm, a link to the intervention website (The Toolbox) was accessible through the OWC immediately. Participants in the control group accessed "The Toolbox" 4 weeks after registration. During the study period, participants from both the intervention and control arm received weekly SMS or email prompts at a time chosen by them during registration, encouraging them to log in to the OWC. The prompts contained a unique link which when clicked logged them in and took them to a page to complete their EMAs. After completing these assessments, they were directed to the OWC homepage, which contained "The Toolbox" access link for the intervention group, and generic well-being advice for the control group. They also received prompts to log in to the OWC to complete study measures at 4 weeks or until they completed.

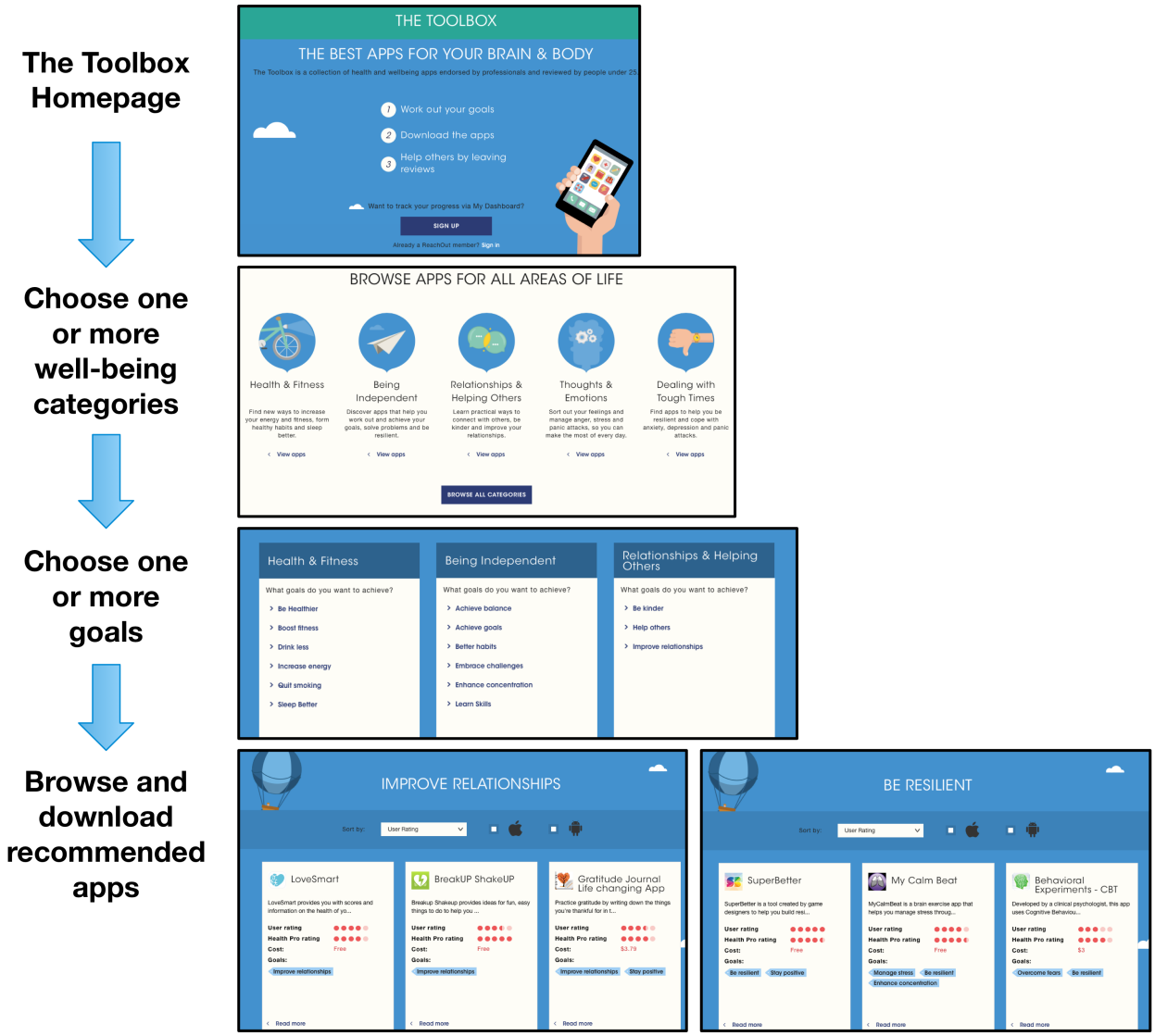
Intervention

The intervention was a personalized app recommendation service called "The Toolbox," available through the ReachOut.com website [24]. The content and structure of the Web-based intervention was determined by young adults' perspectives on well-being and expectations with Web-based interventions. End-user studies were conducted with Australian young adults to investigate how young adults conceptualize well-being. Data were collected via user experience workshops with young people aged 15-21 years. Key findings from the workshop influenced the structure and content of the developed Web-based intervention. The workshops with young people were analyzed, resulting in a nuanced understanding of young adults' conceptualization of health and wellness, and an empirical knowledge of concrete behaviors and actions in their daily lives that they associate with well-being. Data from the workshops were subsequently synthesized into 27 key action areas or goals, and categorized into 6 overarching key themes: health and fitness (15 apps), being independent (8 apps), relationships and helping others (3 apps), thoughts and emotions (18 apps), and dealing with tough times (14 apps). Of these apps, 31 were available for free, 12 apps were paid with costs of up to Aus $6.49, and 3 apps either offered a free lite version or were available for free on one of the platforms, but not on the other. The process of selecting apps to populate "The Toolbox" consisted of two stages. First, a contextual review of available apps was conducted, followed by a review of these apps according to the MARS [10] by professionals and young people. For the contextual review, a list of key search terms was created (see Multimedia Appendix 1), which was drawn from a conceptual well-being model of promoting resilience and flourishing developed for the Young and Well Cooperative Research Centre, as well as qualitative input gained in workshops with young people. These terms were then used to conduct an audit of existing well-being mobile apps available on Google Play and the Apple App Store in late 2013. Only apps that (1) appeared in the first 200 search results, (2) were under Aus $5, (3) were available for Android or iOS, and (4) were deemed appropriate for 13-25 year olds, were included in the rating process. During the rating process, irrelevant apps were removed as well as those not meeting minimum functionality and aesthetic requirements. Remaining apps were rated by researchers using the MARS for both effectiveness and usability and only the highest scoring apps were selected for additional rating by a mental health expert. Apps that contained valid information and were deemed not harmful for young people were selected for "The Toolbox" and additionally rated on the MARS by at least two young people. The final curated list of 46 readily available apps, categorized according to the 27 goals, were put together into a Web resource called "The Toolbox," with an average of 3.62 (SD 3.05) apps per goal (see Multimedia Appendix 2). "The Toolbox" is a responsive website hosted by Reachout.com. Participants first choose the areas they want to focus on, guided by an interactive quiz and subsequently receive recommendations for particular apps to download and use based on their preferences (see Figure 1). For each recommended app, additional information is provided, including the MARS score and reviews by both health professionals and end users on what they liked and did not like, along with costs and links to download from the app store (see Figure 2).

Participants assigned to the intervention arm upon completion of their baseline measures were displayed a Web link which gave them immediate access to "The Toolbox." Over the 4-week study period, participants were sent weekly reminders (via email or SMS) advising them to visit "The Toolbox" at least once, take the quiz, and use the recommended apps. The use of "The Toolbox" website and the recommended third-party apps constitutes the intervention in this study. Participants were aware at all times that the researchers assumed no responsibility for the content and/or functionality of these apps.

XSL•FO
RenderX

Figure 1. Flow of "The Toolbox" website, including home, well-being category selection, goals selection, and recommended apps pages.



Control

Participants in the control group were advised that they were on a waiting list for 4 weeks before they would be given a link to access "The Toolbox." At 4 weeks after completing the baseline measures, participants were provided access to "The Toolbox" via the OWC.

Measures

The primary outcome was a self-reported measure of well-being assessed using the 14-item MHC-SF that measures subjective psychological, emotional, and social well-being. Secondary outcomes were EMAs of 3 questions: How are you feeling today? How is your energy level today? How well did you sleep last night? (see Figure 3). Participants completed primary outcome measures on the Web at baseline and 4 weeks. The EMAs were completed each time participants logged into the OWC during the study period. The log file from the Web app during the trial period was gathered to derive engagement with the study platform. The EMAs of participants were obtained for up to 6 months' postcompletion of trial.

XSL•FO
RenderX

Figure 2. Example of information related to each app available on "The Toolbox" website, including app overview, link to download, user and professional ratings, and app reviews.

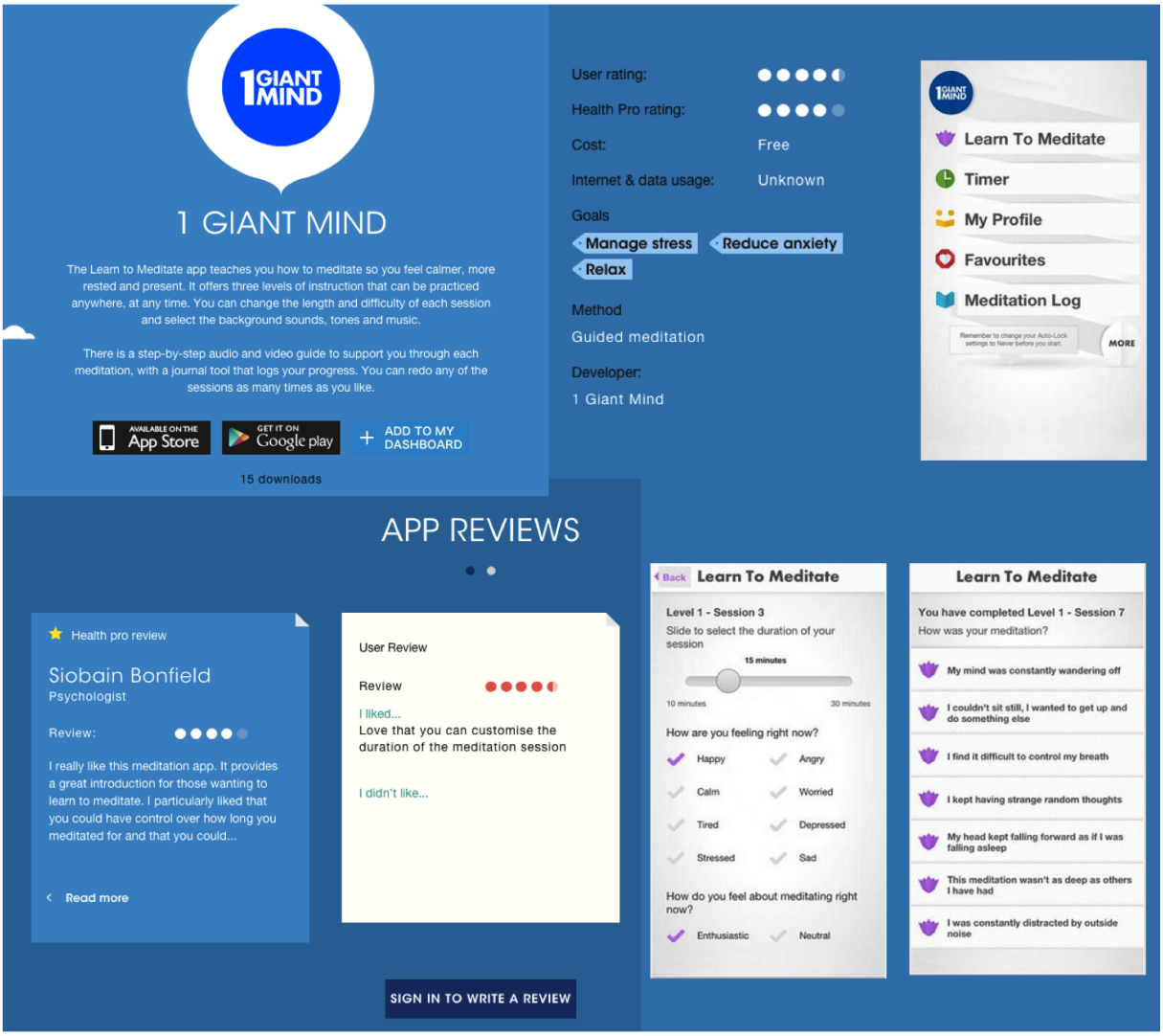
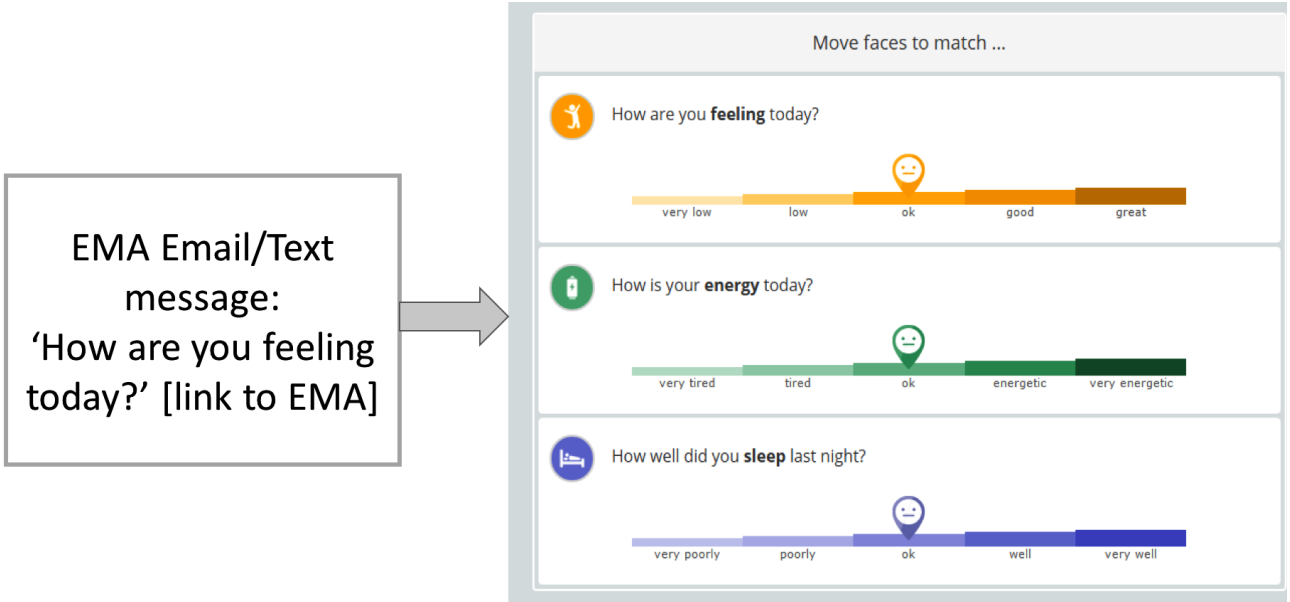


Figure 3. Text message and ecological momentary assessment (EMA).



XSL•FO
RenderX

Statistical Analysis

Differences in attributes between groups (Table 1) and in attrition versus not (Table 2) were assessed using chi-squared tests, *t* test or Wilcoxon rank-sum tests as appropriate. The primary analysis was based on intention to treat, and missing values from all randomized participants were imputed with 50 samples redrawn from the original data. The primary outcome was analyzed using linear regression (Table 3) with well-being score at 4 weeks as the dependent variable. The independent variables were well-being, measured at baseline, and group assignment. A multivariable intention to treat linear regression sensitivity analysis was also conducted, as well as the same analysis, using observed data only. To investigate how the momentary assessment of mood, sleep, rest, and energy were influenced by the intervention (Table 4), the trajectories of momentary assessment measures were examined using random effects mixed modeling. The independent variables were group assignment, engagement with the study platform (coded as number of logins), the product term of group assignment, and engagement, and potential confounders were age, gender, prior application use (coded as a binary variable), baseline energy, baseline mood, and whether or not an app was downloaded. Subject was entered into the model as a random effect to account for correlated readings within an individual. Differences between groups were assessed using interaction terms. Similar multivariable linear regression analyses were conducted with the MHC-SF subscales as outcomes, with covariates listed as before. To examine whether engagement with the study platform was associated with changes in the EMA measures, a linear regression model was run with postintervention EMA measures as the outcome. All models contained an additional term representing the number of logins. For energy, mood, and rest, the other covariates were listed before in the sensitivity analyses. For sleep the other covariates were baseline sleep and group assignment due to the small number of observations. For these regressions, postintervention measurements for EMAs were taken as the measurement that occurred between 30 and 45 days, with the earliest one after 30 days. All results are reported with 95% CI and *P* values. A *P* value <.05 (2-tailed) was taken to be significant. All analyses were performed using Stata version 13.1 (StataCorp).

Results

Flow of Participants

Figure 4 shows the flow of participants. A total of 476 people were consented and signed up on the Web. Of these, 387 completed baseline scores and were randomized into the control (n=195) and active (n=192) arm of the trial.

Participant Characteristics

The demographic characteristics and baseline scores are shown in Table 1.

Table 1. Baseline characteristics.

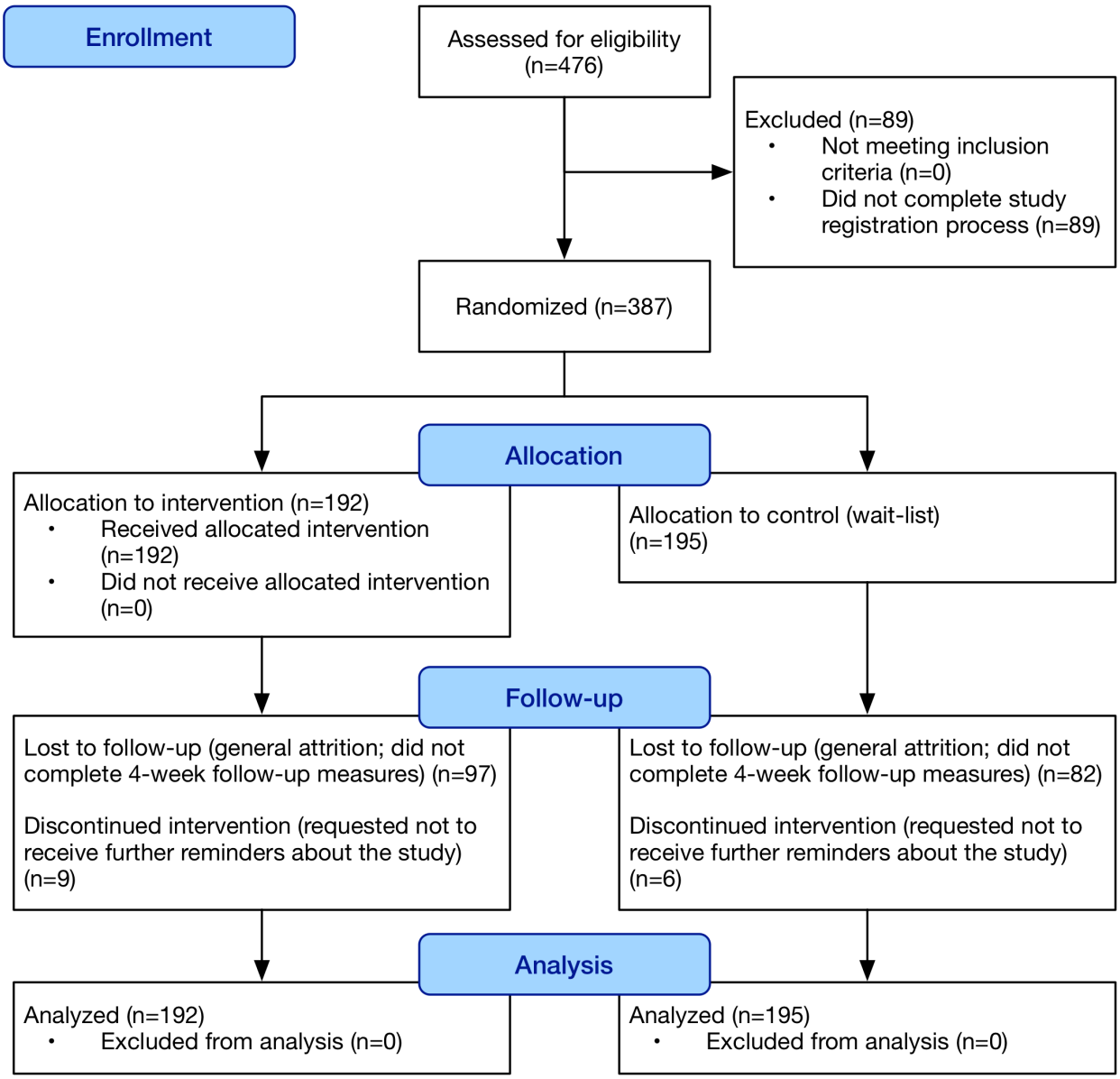
Participant characteristics	Control	Intervention	Total	Statistics	*P* value
Age in years, median (interquartile range)	23 (20-25)	23 (20-25)	23 (20-25)	z=0.67	.51
Female gender, n (%)	152 (78.4)	143 (75.3)	295 (76.8)	χ^2_1=0.5	.47
Prior app usage, n (%)	72 (36.9)	65 (33.9)	137 (35.4)		
MHC-SF[a], median (interquartile range)	41 (28-51)	39 (27-51)	40 (27.5-51)	t_{358}=0.16	.44
Subscale: emotional, median (interquartile range)	10 (7-12)	10 (8-12)	10 (7-12)	z=0.24	.81
Subscale: social, median (interquartile range)	13 (7-17)	12 (6-17)	12 (6-17)	z=0.83	.41
Subscale: psychological, median (interquartile range)	18 (11-22)	18 (11-23)	18 (11-23)	z=0.03	.98
EMA[b] "rest," median (interquartile range)	50 (36-65)	50 (35-64)	50 (36-54)	z=0.03	.98
EMA "mood," median (interquartile range)	50 (40-70)	50 (40-70)	50 (40-70)	z=0.03	.98
EMA "energy," median (interquartile range)	50 (40-60)	50 (30-60)	50 (40-60)	z=0.03	.98

[a]MHC-SF: Mental Health Continuum-Short Form.

[b]EMA: ecological momentary assessment.

XSL•FO
RenderX

Figure 4. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of study participants during study enrolment, allocation, follow-up, and analysis.



Attrition

Attrition measured as the failure to respond to the primary outcome measurements at 4 weeks postrandomization was 45.1% in the control group (88/195) versus 55.2% in the active group (106/192), *P*=.047. In the control group, the mean baseline MHC-SF for those who responded at 4 weeks, versus those who did not was 40.2 versus 37.6 (*P*=.25). In the intervention group, the mean baseline MHC-SF for those who responded at 4 weeks, versus those who did not was 36.7 versus 42.3 (*P*=.12).

A comparison of characteristics between completers (ie, those that reported data on the primary outcome at 4 weeks) and noncompleters revealed no significant differences across all demographics at baseline, apart from mood as assessed through EMA (Table 2).

XSL•FO
RenderX

Table 2. Baseline characteristics for those who provided data at four weeks versus not.

Participant characteristics[a]	Noncompleters	Completers	Statistics	*P* value
Age in years, median (interquartile range)	23 (20-25)	23 (20-25)	z=0.05	.96
Female gender, n (%)	14 (77.0)	148 (76.7)	χ^2_1=0.0	.95
Prior app usage, n (%)	69 (35.6)	68 (35.2)	t_{358}=0.50	.94
MHC-SF[a], median (interquartile range)	41 (27-50)	40 (28-52)	t_{358}=0.50	.78
Subscale: emotional, median (interquartile range)	10 (7-12)	10 (8-12)	z=0.82	.41
Subscale: social, median (interquartile range)	12 (6-17)	12 (7-17)	z=0.49	.62
Subscale: psychological, median (interquartile range)	19 (11-22)	18 (12-23)	z=0.14	.89
EMA[b] "rest," median (interquartile range)	50 (35.5-63)	50 (36-65)	z=0.18	.86
EMA "mood," median (interquartile range)	50 (40-70)	60 (50-70)	z=2.13	.03
EMA "energy," median (interquartile range)	50 (30-60)	50 (40-60)	z=0.66	.51
EMA "sleep," median (interquartile range)	420 (375-480)	435 (360-525)	z=0.22	.85

[a]MHC-SF: Mental Health Continuum-Short Form.

[b]EMA: ecological momentary assessment.

Primary Analysis

The mean (SD) observed MHC-SF scores at 4 weeks for the control and active groups were 38.6 (SD 15.4) and 42.0 (SD 16.8), respectively. In the primary intention-to-treat (ITT) analysis, those in the intervention group experienced an improvement of 0.63 (95% CI −2.26 to 3.53) in MHC-SF score relative to the control group, but this was not significant, *P*=.66. In multivariable sensitivity intention to treat analysis the difference was almost identical, 0.64 (95% CI −2.27 to 3.54), *P*=.66. In a further multivariate sensitivity analysis with observed data only, there was also no difference between groups, 1.10 (95% CI −1.68 to 3.89), *P*=.44. In a completers only analysis there was also no difference between groups in MHC-SF scores 1.17 (95% CI −1.98 to 3.53), *P*=.40.

Subscales of Mental Health Continuum-Short Form (MHC-SF)

There were no significant differences between groups in any of the subdomain scores, psychological 0.57 (95% CI −0.67 to 1.81), *P*=.95, social 0.46 (95% CI −0.68 to 1.59), *P*=.42, and emotional −0.02 (95% CI −0.72 to 0.68), *P*=.95.

Analyses of Ecological Momentary Assessments

For all EMA measures, the control group decreased significantly per login in contrast to the active group which showed no significant change in scores over time. Thus, the difference between groups per login was also significant (Table 4).

Table 3. Multivariable intention to treat analyses (adjusted for group assignment, age, gender, prior app use, energy, mood, and whether or not an app has been downloaded).

Outcome Measure	Value beta (95% CI)	*P* value
MHC-SF[a]	.64 (−2.27 to 3.54)	.66
Psychological	.57 (−0.67 to 1.81)	.95
Social	.46 (−0.68 to 1.59)	.42
Emotional	−.02 (−0.72 to 0.68)	.95

[a]MHC-SF: Mental Health Continuum-Short Form.

XSL•FO
RenderX

Table 4. Changes per login in ecological momentary assessments (adjusted for group assignment, age, gender, prior app use, whether or not an application has been downloaded).

Ecological Momentary Assessments	Control beta (95% CI), *P* value	Intervention beta (95% CI), *P* value	Difference between groups beta (95% CI), *P* value
Mood	−.25 (−0.40 to −0.11), *P*=.001	.15 (−0.04 to 0.33), *P*=.13	.40 (0.16-0.63), *P*=.001
Energy	−.20 (−0.33 to −0.07), *P*=.003	.11 (−0.06 to 0.29), *P*=.19	.31 (0.10-0.52), *P*=.004
Rest	−.19 (−0.36 to −0.08), *P*=.001	.12 (−0.04 to 0.29), *P*=.15	.31 (0.11-0.52), *P*=.002
Sleep	−2.34 (−3.16 to −1.52), *P*<.001	−.46 (−1.66 to 0.74), *P*=.15	1.88 (0.43-3.34), *P*<.001

Engagement-Response Analysis

There was no evidence to suggest that a beneficial effect was associated with the number of logins in any of the measures, mood 0.15 (95% CI −0.58 to 0.87), *P*=.69, energy 0.08 (95% CI −0.56 to 0.72), *P*=.81, rest −0.13 (95% CI −0.91 to 0.64), *P*=.73, and sleep −5.29 (95% CI −11.96 to 1.54), *P*=.12.

Discussion

Principal Findings

The aim of this study was to assess the efficacy of "The Toolbox" Web-based well-being intervention in a young adult population. Results from the randomized controlled trial demonstrated no significant benefit on well-being (as assessed using the MHC-SF) at 4 weeks compared with the control group. There were no significant differences between the active and control groups at 4 weeks on any of the subscales of the MHC-SF either. The trial results also suggest that the impact of receiving weekly texts and the opportunity to monitor and visualize sleep, mood, and energy led to repeated logins in both control and active groups. In addition, participants in the control group reported a significant decline in mood, energy, rest, and sleep as assessed with EMAs with an increasing number of logins, whereas the intervention group showed no change. Thus, repeat engagement with the intervention might halt decline in mood, energy, rest, and sleep, without resulting in significant changes in well-being as assessed by the MHC-SF at fixed points. It remains unclear as to whether this interaction can be attributed to using "The Toolbox" intervention or using the study platform and its repeated assessments. The magnitude of change in the control group was very small, which may explain why no change in well-being as measured by the MHC-SF was observed.

Comparison With Previous Work

The lack of effect on MHC-SF well-being scores observed in this intervention are comparable with results from similar Web-based intervention trials, conducted in older (mean age 43.2 years) [25] and under 16-years-old school-based samples [26]. Across both studies, positive benefits were shown in depression scales but benefits assessed using mental well-being scales themselves were minuscule and nonsignificant both immediately after the intervention and at follow-up. Compared with these studies, the intervention in our study was targeted at a general young adult sample (mean age 23 years) with a broad inclusion criterion that did not exclude participants based on symptom screening, which closely resembles the real-world setting of intervention delivery through the Reachout.com website. Considerably more females than males participated in this study, which is in concordance with the majority of research into mental health and well-being interventions. In part, the higher proportion may be attributed to the higher prevalence of mental disorders in females in this age group [27]. However, differences in help-seeking behavior between males and females likely account for the majority of this difference [28].

In our study, instead of administering depression scales, we assessed symptoms of depression at multiple time points through momentary assessments. Despite the differences in type of assessment, we detected comparable benefits on depression as evident by significant improvement in mood trajectories in the intervention group. One plausible explanation from these findings is that modest improvements detected in mental well-being might actually be a reduction in depression symptoms that have collinearity with mental well-being [29]. This raises questions about sensitivity of mental well-being scales to detect change and if Web-based interventions can change mental well-being as an independent construct in the absence of mental illness.

Compared with past studies, the intervention in this study is unstructured, in the form of a collection of curated list of mobile app resources accessible through a self-guided hub, as well as monitoring tools to engage participants and provide feedback irrespective of app use. Disseminating a curated list of mental health and well-being apps for depression and anxiety alone has been recently demonstrated to yield better app uptake [13]; however, ours is the first study to investigate effectiveness of such an intervention. In order to ensure optimal app recommendations, different components and strategies within apps that serve as active intervention ingredients must thus be identified [30] and aligned with end user needs. The intervention in this study included a broad range of curated apps (n=46) with an algorithm that assigned these apps to one of the 27 action areas used in the app selection quiz, that were identified based on young adults' conceptualization of well-being through codesign activities. Since the spread of apps were not uniform across all action areas, the collection of apps might not have been optimal to be suitable and effective for all individuals. We also observed a slightly higher baseline symptomology in participants dropping out from the intervention, which might be caused by the mismatch between apps and individual health circumstances that were not factored in the matching algorithm.

The use of specific apps over a 1-month period may not have been sufficient to induce significant changes in well-being as measured by a global mental health scale, such as the MHC-SF.

XSL•FO
RenderX

Instead, our results suggest that app usage may affect momentary moods and behaviors more easily measured by EMAs. It may be that EMAs provide a more accurate measure of the day-to-day impact of app usage. To date, there are few studies of well-being interventions utilizing momentary assessments as outcome measures, although their superiority to traditional questionnaire measures has recently been reported [31].

Limitations

There are several limitations to our study. There was a high attrition rate of almost 50% subject randomized, which is not unusual in Web-based interventions. However, the results of the primary analysis were consistent with the sensitivity analyses. Interestingly, we found a higher rate of attrition in the intervention group compared with the control group, a finding which was also reported by Bolier et al [25]. The reason for the greater attrition rate in the intervention group is unclear and, along with many factors associated with the high attrition rates in Web-based interventions, requires further study [32]. Given that intervention and study dropout are often linked in Web-based interventions, it is possible that overall attrition was higher in the intervention group because participants immediately had access to the intervention and thus had no incentive to participate in the 4-week assessment. Alternatively, it is possible that the content, functionality, and aesthetics of some apps may have changed during the short time from when they were added to "The Toolbox" to when they were accessed by participants, as is common on mobile app marketplaces [33], thus not meeting, or differing significantly from their expectations. All of the recommended apps remained available on the app stores for the duration of the study.

Due to the nature of the intervention, it was not possible to quantify the download and use of the apps recommended by "The Toolbox." Thus, there was no direct measure of intervention adherence. This was accounted for by using ITT analysis; however, ultimately it was not possible to determine whether the lack of effect on well-being was due to nonadherence or due to a lack of effectiveness of "The Toolbox" and its recommended apps. This is an inherent problem when studying the effectiveness of third-party intervention that can only be overcome by retaining full ownership of the intervention tool, as was the case in Lattie et al [13]. The study was also limited by a relatively heterogeneous sample of participants recruited using varying strategies, although this could be both a strength and limitation as it replicates app recommendation interventions in real-life setting. The other major limitation was the lack of longitudinal follow-up data; however, it is unlikely longitudinal effects would be found when no benefits were observed at follow-up (4 weeks), which is when most changes are expected.

Conclusions and Recommendations for Future Research

Whereas there are several randomized controlled trials of the efficacy of Web-based services to improve mental health, previous studies have been conducted in adults with symptoms of anxiety and depression [25,34,35]. In comparison, this was the first study to assess the effectiveness of a well-being intervention designed to recommend the use of readily available mobile apps in a sample of young adults. The design of the intervention utilized expert rating of existing apps and end-user codesign approaches, resulting in an app recommendation service.

Our findings cast doubt on the effectiveness of mobile apps for well-being and mental health in a nonclinical population of young adults. This intervention included a self-guided optimal selection of apps. Further work could focus on developing algorithms to automate the process of determining optimal apps for an individual, taking into account active ingredients in apps, personal characteristics, engagement, and needs. Intervening with the right combination of quality apps is critical to realizing benefits of over 30,000 mental health related apps available in the app store. The instruments used for assessing mental well-being in this study may not have been sensitive enough to detect change caused by app usage. Future research should focus on refining the construct so that it is sensitive to change even when symptoms of depression or mental illness are absent. In addition, consideration should be given to the measurement of the specific behaviors targeted by particular apps as well as to overall constructs, such as well-being.

Acknowledgments

This research is part of a collaborative project cofunded by the Young and Well Cooperative Research Centre, Flinders University, and Country Health SA. The Young and Well CRC was established under the Australian Government's Cooperative Research Centres Program. "The Toolbox: the best apps for your brain and body" has been designed by ReachOut.com.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of key search terms used to conduct an audit of existing well-being mobile apps available on Google Play and the Apple App Store, during the app selection process for "The Toolbox" intervention.

[PDF File (Adobe PDF File), 37KB - jmir_v19i5e141_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2

List of apps contained within “The Toolbox,” classified according to their overarching key themes and action areas or goals.

[PDF File (Adobe PDF File), 41KB - jmir_v19i5e141_app2.pdf]

Multimedia Appendix 3

CONSORT E-HEALTH checklist form.

[PDF File (Adobe PDF File), 1MB - jmir_v19i5e141_app3.pdf]

References

1. AIHW. 2011. Young Australians: their health and wellbeing 2011 URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737419259 [accessed 2016-10-06] [WebCite Cache ID 6l34aBcbj]
2. Australian Institute of Health and Welfare. AIHW. 2016. Australian Burden of Disease Study: impact and causes of illness and death in Australia URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129555476 [accessed 2016-10-06] [WebCite Cache ID 6l34reT3F]
3. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
4. Whitfield G, Williams C. If the evidence is so good-why doesn't anyone use them? A national survey of the use of computerized cognitive behaviour therapy. Behav Cognit Psychother 1999;32(1):57-65. [doi: 10.1017/S1352465804001031]
5. IMS Institute for Healthcare Informatics. IMShealth. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2015. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance URL: http://www.imshealth.com/files/web/IMSH%20Institute/Reports/Patient%20Adoption%20of%20mHealth/IIHI_Patient_Adoption_of_mHealth.pdf [accessed 2016-10-06] [WebCite Cache ID 6l35OXf16]
6. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
7. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
8. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
9. Chia P, Yamamoto Y, Asokan N. Is This App Safe?: A Large Scale Study on Application Permissions and Risk Signals. 2012 Apr 16 Presented at: Proceedings of the 21st International Conference on World Wide Web; April 16-20, 2012; Lyon, France p. 311-320 URL: http://dl.acm.org/citation.cfm?id=2187879 [doi: 10.1145/2187836.2187879]
10. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
11. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. JMIR Mhealth Uhealth 2016;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
12. Painter JE, Borba CP, Hynes M, Mays D, Glanz K. The use of theory in health behavior research from 2000 to 2005: a systematic review. Ann Behav Med 2008 Jun;35(3):358-362. [doi: 10.1007/s12160-008-9042-y] [Medline: 18633685]
13. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, et al. Uptake and usage of IntelliCare: a publicly available suite of mental health and well-being apps. Internet Interv 2016 May;4(2):152-158 [FREE Full text] [doi: 10.1016/j.invent.2016.06.003] [Medline: 27398319]
14. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
15. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
16. Comello ML, Qian X, Deal AM, Ribisl KM, Linnan LA, Tate DF. Impact of game-inspired infographics on user engagement and information processing in an eHealth program. J Med Internet Res 2016 Sep 22;18(9):e237 [FREE Full text] [doi: 10.2196/jmir.5976] [Medline: 27658469]
17. Christensen H, Murray K, Calear AL, Bennett K, Bennett A, Griffiths KM. Beacon: a web portal to high-quality mental health websites for use by health professionals and the public. Med J Aust 2010 Jun 07;192(11 Suppl):S40-S44. [Medline: 20528708]
18. 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]

XSL•FO
RenderX

19. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
20. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AX, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
21. Firth J, Torous J, Yung AR. Ecological momentary assessment and beyond: the rising interest in e-mental health research. J Psychiatr Res 2016 Sep;80:3-4 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.05.002] [Medline: 27236099]
22. Antezana G, Bidargaddi N, Blake V, Schrader G, Kaambwa B, Quinn S, et al. Development of an online well-being intervention for young people: an evaluation protocol. JMIR Res Protoc 2015;4(2):e48 [FREE Full text] [doi: 10.2196/resprot.4098] [Medline: 25929201]
23. Musiat P, Winsall M, Orlowski S, Antezana G, Schrader G, Battersby M, et al. Paid and unpaid online recruitment for health interventions in young adults. J Adolesc Health 2016 Sep 20;59(6):662-667. [doi: 10.1016/j.jadohealth.2016.07.020] [Medline: 27663927]
24. ReachOut. The Toolbox: The Best Apps for your Brain & Body URL: https://au.reachout.com/sites/thetoolbox [accessed 2016-10-06] [WebCite Cache ID 6l37cwKDW]
25. Bolier L, Haverman M, Kramer J, Westerhof GJ, Riper H, Walburg JA, et al. An Internet-based intervention to promote mental fitness for mildly depressed adults: randomized controlled trial. J Med Internet Res 2013;15(9):e200 [FREE Full text] [doi: 10.2196/jmir.2603] [Medline: 24041479]
26. Burckhardt R, Manicavasagar V, Batterham PJ, Miller LM, Talbot E, Lum A. A web-based adolescent positive psychology program in schools: randomized controlled trial. J Med Internet Res 2015;17(7):e187 [FREE Full text] [doi: 10.2196/jmir.4329] [Medline: 26220564]
27. Ang RP, Lim KM, Tan A, Yau TY. Effects of gender and sex role orientation on help-seeking attitudes. Curr Psychol 2004 Sep;23(3):203-214. [doi: 10.1007/s12144-004-1020-3]
28. Mackenzie CS, Gekoski WL, Knox VJ. Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. Aging Ment Health 2006 Nov;10(6):574-582. [doi: 10.1080/13607860600641200] [Medline: 17050086]
29. Keyes CL. Mental illness and/or mental health? Investigating axioms of the complete state model of health. J Consult Clin Psychol 2005 Jun;73(3):539-548. [doi: 10.1037/0022-006X.73.3.539] [Medline: 15982151]
30. van Genugten L, Dusseldorp E, Webb TL, van Empelen P. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? a meta-analysis. J Med Internet Res 2016;18(6):e155 [FREE Full text] [doi: 10.2196/jmir.4218] [Medline: 27268104]
31. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123. [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
32. 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]
33. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016 Aug 09;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
34. Christensen H, Batterham P, Mackinnon A, Griffiths KM, Kalia HK, Kenardy J, et al. Prevention of generalized anxiety disorder using a web intervention, iChill: randomized controlled trial. J Med Internet Res 2014;16(9):e199 [FREE Full text] [doi: 10.2196/jmir.3507] [Medline: 25270886]
35. Kramer J, Conijn B, Oijevaar P, Riper H. Effectiveness of a web-based solution-focused brief chat treatment for depressed adolescents and young adults: randomized controlled trial. J Med Internet Res 2014;16(5):e141 [FREE Full text] [doi: 10.2196/jmir.3261] [Medline: 24874006]

Abbreviations

ANZCTR: Australian New Zealand Clinical Trials Registry
BCT: behavior change technique
DECD: Department of Education and Child Development of South Australia
EMA: ecological momentary assessment
ITT: intention-to-treat
MARS: mobile app rating scale
MHC-SF: Mental Health Continuum-Short Form
OWC: Online Wellbeing Centre
SMS: short message service

XSL•FO
RenderX

Edited by G Eysenbach; submitted 09.10.16; peer-reviewed by J Torous, R Kothari, S Schueller; comments to author 03.11.16; revised version received 12.12.16; accepted 18.02.17; published 12.05.17.

Please cite as:
Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, Orlowski S, Antezana G, Schrader G
Efficacy of a Web-Based Guided Recommendation Service for a Curated List of Readily Available Mental Health and Well-Being Mobile Apps for Young People: Randomized Controlled Trial
J Med Internet Res 2017;19(5):e141
URL: http://www.jmir.org/2017/5/e141/
doi:10.2196/jmir.6775
PMID:28500020

©Niranjan Bidargaddi, Peter Musiat, Megan Winsall, Gillian Vogl, Victoria Blake, Stephen Quinn, Simone Orlowski, Gaston Antezana, Geoffrey Schrader. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 12.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Attachment Style and Internet Addiction: An Online Survey

Christiane Eichenberg[1], PhD; Markus Schott[1], MSc; Oliver Decker[2], PhD; Brigitte Sindelar[1], PhD

[1]Sigmund Freud PrivatUniversität Wien, Vienna, Austria

[2]Universtität Leipzig, Leipzig, Germany

Corresponding Author:
Christiane Eichenberg, PhD
Sigmund Freud PrivatUniversität Wien
Kelsenstr 2
Vienna,
Austria
Phone: 43 1 7984098
Email: eichenberg@sfu.ac.at

Abstract

Background: One of the clinically relevant problems of Internet use is the phenomenon of Internet addiction. Considering the fact that there is ample evidence for the relationship between attachment style and substance abuse, it stands to reason that attachment theory can also make an important contribution to the understanding of the pathogenesis of Internet addiction.

Objective: The aim of this study was to examine people's tendency toward pathological Internet usage in relation to their attachment style.

Methods: An online survey was conducted. Sociodemographic data, attachment style (Bielefeld questionnaire partnership expectations), symptoms of Internet addiction (scale for online addiction for adults), used Web-based services, and online relationship motives (Cyber Relationship Motive Scale, CRMS-D) were assessed. In order to confirm the findings, a study using the Rorschach test was also conducted.

Results: In total, 245 subjects were recruited. Participants with insecure attachment style showed a higher tendency to pathological Internet usage compared with securely attached participants. An ambivalent attachment style was particularly associated with pathological Internet usage. Escapist and social-compensatory motives played an important role for insecurely attached subjects. However, there were no significant effects with respect to Web-based services and apps used. Results of the analysis of the Rorschach protocol with 16 subjects corroborated these results. Users with pathological Internet use frequently showed signs of infantile relationship structures in the context of social groups. This refers to the results of the Web-based survey, in which interpersonal relationships were the result of an insecure attachment style.

Conclusions: Pathological Internet use was a function of insecure attachment and limited interpersonal relationships.

(J Med Internet Res 2017;19(5):e170) doi:10.2196/jmir.6694

KEYWORDS

Internet; addictive behavior; surveys and questionnaires; Rorschach test

Introduction

Background

Nowadays the Internet is a central aspect of everyday life. There are many opportunities for online dating, problems can be discussed in thematically appropriate forums, and doctors can be consulted [1,2]. Even university education is shifting toward the net. Due to the interactive options afforded, the Internet has firmly established itself within everyday life [3]. However, negative aspects of the Internet are increasingly brought to the attention of the public [4]. For instance, certain Web-based self-help forums on problematic areas such as suicidality [5], self-injurious behavior [6], cyberbullying [7], or negative effects on health-related Internet use [8], are controversially discussed both within the media and among professionals.

Internet Addiction

One of the clinically relevant problems Internet use may yield is Internet addiction. Even though most people use the Internet every day without problems, recent figures highlight that excessive Internet use is a significant problem [9]. As a unified concept of the disorder and related diagnostic tools are not yet available [10], prevalence estimates demonstrate high levels of variance with values ranging from 1.5-11.6%. A review article

XSL•FO
RenderX

found an overall prevalence of 3-5% for computer game and Internet addiction.

With regard to the many behaviors associated with the Internet, the question arises as to whether "Internet addiction" concerns a singular disorder, or whether the only commonality among various Internet-related behavior disorders is the use of the same medium (for the concept of specific Internet addiction, see Davis [11]). Accordingly, excessive Internet use can refer to different Web-based services, such as Web-based computer games [12], gambling [13], sexual content [14], or communicative apps such as chat or social networks [15] that differ in their "addictive" potential. Of particular concern for addiction behaviors are Web-based role-play games and sexual content [16].

Another question relating to the independence of the disorder "Internet addiction" has been discussed for many years [17]: Is "Internet addiction" an independent clinical disorder or rather a concomitant symptom of some other mental illness? Neurobiological research contributes to the answer by showing the equivalence of substance-related addictions and behavioral addictions [18]. This finding is mainly based on the neurobiological correlations between substance-related and behavioral addictions [19]. These results are part of a growing body of research that advocates Internet addiction as a behavioral addiction.

Behavioral addiction is a relatively new term for excessive behavior, which depicts characteristics of psychological dependence [20,21]. The term "behavioral addiction" marks the behavior of the user, and not the medium as the pathological object. In the latest edition of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), behavioral addictions have been explicitly included; the chapter previously named "Substance-Related Disorders" was renamed into "Addiction and Related Disorders." The diagnostic criteria of a behavioral addiction are based on the criteria known for substance dependencies, that is, development of tolerance, withdrawal symptoms, unsuccessful attempts to reduce consumption, neglecting other areas, and use despite negative consequences [22]. Although Internet addiction does not constitute an autonomous clinical entity, Internet Gaming Disorder is listed in the appendix of DSM-5.

The inclusion in Section III of DSM-5 outlines the importance of Internet-related addictions, emphasizing the need for further research in this area. Thus, further studies investigating the etiology and pathogenesis are necessary. Furthermore, it is crucial to identify variables and dispositions that determine the etiology of Internet addiction in order to develop effective therapeutic measures.

Attachment Theory

In line with integrative explanations for substance use disorders, complex etiopathogenic models are also employed for Internet addiction. For instance, Wölfling et al [23] developed an integrative model with an emphasis on learning theory and neurobiological mechanisms in the context of personality traits. Additionally, Internet addiction can be explained using cognitive-behavioral [24], psychodynamic approaches [25], as well as cultural and social considerations [26]. In this context, attachment theory can also make an important contribution to understanding preconditions associated with the development of behavior addictions, due to the widely documented relationship between attachment experiences and substance dependency [27]. Similar mechanisms underlying substance dependence are therefore both feasible and plausible for Internet addictions. For example, bypassing situations where patterns of attachment are activated by excessively spending time on the Internet or replacing negative relationship experiences by rewarding Web-based activity [28]. Overall, as a social medium and the relationship component contained therein, the Internet—compared with other addictive substances—provides even more possibilities to manage deficient attachment and relationship patterns. For example, online interaction in social networks, chats, and forums can dampen feelings of social isolation of people with an uncertain attachment style. In particular, the possibility of anonymous communication over the Internet plays a vital role in compensating social isolation via online contacts and relationships [29].

Research Question

On the basis of the discussed findings, the aim of this study was to investigate the relationship between attachment style, motives for use, used services, and Internet addiction. It was hypothesized that, in comparison with securely attached users, significantly more insecurely attached users would display an Internet addiction. Furthermore, it was assumed that users with an insecure attachment style would differ in their motives for use compared with users with a secure attachment style. It was also hypothesized that users with an Internet addiction would report different motives for use compared with users without an Internet addiction. Finally, it was assumed that users with an insecure attachment style would use Web-based services more often than users with a secure attachment style.

Methods

Design

A Web-based survey was carried out. The Web-based questionnaire was distributed both on Facebook and on 15 thematically different forums (ranging from parent-, travel-, computer- to craft- and comic platforms) to obtain the most heterogeneous sample possible. The survey period lasted for 6 weeks. For descriptive inferential data analysis and hypothesis testing, SPSS version 20 (IBM, Somers, NY, USA) was used.

To rule out methodological artifacts in the first quantitative self-assessment study, a second qualitative study was conducted. In light of the limitations given in any self-evaluation study, the prior aim of this methodological triangulation study was to gain data that for sure could not be influenced by intentional or unintentional bias. Therapists in Austrian clinics as well as the University Clinic of Mainz were contacted to recruit possible participants for the Rorschach study.

Sociodemographic Data

Age, gender, relationship status, and the duration of the existing partnership, highest level of education as well as current job situation were assessed. Furthermore, information about duration and frequency of Internet usage was collected.

XSL•FO
RenderX

Bielefelder Partnership Expectations Questionnaire

The Bielefelder partnership expectations questionnaire was used to assess the attachment style of participants. This inventory consists of 31 items that are rated on a 5-point Likert scale ranging from 0 (*completely disagree*) to 4 (*completely agree*).

The questionnaire evaluates the following five attachment styles: *secure*, *conditionally secure*, *avoidant-closed*, *ambivalent-clingy*, and *ambivalent-closed*. The reliability of the scales (Cronbach alpha=.77 to .89) is satisfactory.

The Bielefeld questionnaire is different from others in two ways: (1) attachment style is operationalized as configurations of scale scores, which allow qualitative distinctions in terms of functioning and (2) five empirically identified attachment styles are distinguished. Nonetheless validation of the classifications with a German translation of the "Adult Attachment Scale (AAS)" yielded good results [30].

Online Addiction Scale

The online addiction scale is a diagnostic tool consisting of 14 items about Internet addiction that are rated on a 5-point Likert scale with a maximum possible total score of 27 points. Cutoff values differentiate three different user types: *normal* (<7 points), *problematic* (>7 points and <13 points), and *pathological* (>13 points).

In addition, the questionnaire assesses how frequently participants used the following 8 different Web-based services: Web-based games, shopping, chatting in forums, writing emails, Web-based sex services, Web-based gambling, Web-based communities, and information retrieval. The frequency of use is reported on a 4-point scale ranging from 0 (*never*) to 3 (*very often*).

Reliability, validity, and utility of the instrument have been confirmed showing a good internal consistency of .88 and homogeneity of .34. An exploratory factor analysis yielded a one-way solution confirming factorial validity [31].

Cyber Relationship Motive Scale (CRMS-D)

The Cyber Relationship Motive Scale is a self-evaluation of user's relationship motives for going online. Survey respondents are prompted to rate how well each of the 27 possible items applied to their motives for Internet usage on a 5-point scale ranging from 0 (*strongly agree*) to 4 (*strongly disagree*): anonymity, opportunity to meet new people, simple communication, curiosity, emotional support, social contact, escape from the real-world, finding love or a sexual partner.

A confirmatory factor analysis was conducted yielding a goodness-of-fit index of .90. On the basis of the factor loadings, acceptable validity could be determined [32].

Rorschach Inkblot Test

The Rorschach Inkblot Test is a performance-based personality test. The test consists of 10 inkblot stimuli: 5 are achromatic and 5 include chromatic colors. Examinees look at each inkblot and say what it looks like or what it might be. Examinees can give one or more responses per inkblot. Following test administration, Rorschach responses are coded and tallied to form main variables, such as "Situational Stress," "Affective Features," "Interpersonal Perception," or "Self-Perception." The interpretation for each Rorschach variable is guided by interpretive paragraphs that are sequentially arranged in the test manual. The Rorschach variables are given a cutoff score that indicates which interpretive paragraph to choose. To determine the degree to which the results statistically deviate from the norm, the examiner must compare each of the variables with the relevant descriptive statistics that are reported in large normative tables [33].

The Rorschach test can be described as "immune" to any form of bias. This immunity is due to the fact that the evaluation of the Rorschach test is not carried out by an interpretation of the subject's responses; instead responses are coded by established criteria. Although those established criteria differ based on the evaluation system applied, they generally follow the same basics: acquisition, the experience, and content [34]. This complex evaluation strategy is by no means transparent for subjects; as such any attempt at influencing the interpretation is impossible. This criterion is impossible to be met by any questionnaire or any narrative projective test (like Thematic Apperception Test) but only by a performance-based test analyzing the performances of the client quantitatively [34,35].

As the Rorschach Inkblot Method (RIM) had provoked numerous discussions about its reliability, validity, and utility, it has received a maybe more intensive level of scrutiny than any other personality test, summarized in a meta-analysis and reproved in an independent blue ribbon panel. Taken together, studies on reliability and validity of the RIM showed the same or even more valid results for the RIM as for other well-validated inventories [36,37].

Results

Sample

During the survey period, the Web-based questionnaire was accessed N=1009 times. Around 39,74% (401/1009) of participants did not proceed past the start page, another 28.35% (286/1009) had dropped out by page 3. However, as only a further few participants dropped out at later stages of the survey, the dropout rate can be deemed acceptable [38].

Overall, the questionnaire was completed 249 times, constituting 24.86% of total page views. After checking for plausibility of answers, 4 records were removed, resulting in a total sample of 245 participants (168 female, 77 male) aged 16-61 years (mean 29.6, SD 9.17).

At the time of the survey, 78.8% (193/245) were in a partnership, with an average relationship duration of mean 77.2 months (SD 101.21). Less than half (40.81%) of subjects (100/245) were employed full-time or part-time, one-quarter of the sample consisted of students (62/245, 25.31%), a further 5.71% (14/245) were also working in addition to studying, a small fraction of the total sample were unemployed (16/245, 6.53%), few participants (16/245, 6.53%) were self-employed, and 2.86% were trainees (7/245). In total, 30 subjects (12.2%) reported "other" in terms of employment.

XSL•FO
RenderX

The Rorschach Inkblot Test was conducted with a small sample of 16 voluntary male subjects. Although 8 participants showed abusive Internet usage and 3 of them met the criteria for Internet addiction, a control group of 8 subjects with nonabusive Internet usage was recruited. Participants were aged between 18 and 47 years (mean 31 years). Subjects were primarily students, full-time employees, or self-employed.

Duration and Frequency of Internet Use

On average, study participants had used the Internet for over 10 years (mean 10.91, SD 3.92). The daily use of the Internet for private purpose amounted to an average of mean 4.35 h (SD 4.27), ranging from a minimum of 0.1 h up to 21 h a day. A majority of the sample (93.1%, 228/245) used the Internet daily, only 6.9% (17/245) went online 2-3 times per week. None of the subjects stated that they used the Internet once a week, once a month, or less.

Online Services

Subjects used the Internet mainly for online shopping (mean 2.32, SD 0.72). Other popular services included online research (mean 2.32, SD 0.72) and social platforms (mean 2.02, SD 1.11), emails (mean 2.02, SD 0.85), as well as chats and forums (mean 1.55, SD 1.06). In contrast, games (mean 0.70, SD 0.99), sexual content (mean 0.51, SD 0.79), and gambling (mean 0.12, SD 12.43) were less popular.

Online Relationship Motives

Simple communication (mean 3.47, SD 1.02) and the opportunity to find new friends (mean 3.20, SD 0.95) were the dominant motives for Internet use. Emotional support (mean 2.37, SD 0.94), anonymity (mean 2.04, SD 1.3), and escapism (mean 2.03, SD 1.17) appear to be motives of medium importance. The lowest, sexual (mean 1.59, SD 1.07), and social-compensatory motives (mean 1.69, SD 1.14), were less frequently reported as online relationship motives.

Attachment Style

More than half (50.6%, 124/245) of participants showed an insecure attachment style (*ambivalent-closed*, *ambivalent-clingy*, and *avoidant-closed*), whereas almost as many subjects showed a secure attachment style (*conditionally secure* and *secure*). Therefore, secure and insecure attachment style was equally represented in the sample (Table 1).

Table 1. Frequency distribution of the variable attachment style.

Attachment Style	Frequency, n (%)
Secure attachment	121 (49.4)
Secure	25 (10.2)
Conditionally secure	96 (39.2)
Insecure attachment	124 (50.6)
Avoidant-closed	50 (20.4)
Ambivalent-clingy	36 (14.7)
Ambivalent-closed	38 (15.5)
Overall	245 (100)

Internet Addiction

Only 1.2% of subjects (3/245) could be classified as addicted Internet users. To allow for appropriate statistical analysis, the categories “pathological” and “problematic use” were merged to the variable *Internet addiction tendency* with 30 subjects. A majority (87.3%, 214/245) of users showed normal Internet use.

Attachment Style and Internet Addiction

It was hypothesized that, in comparison with securely attached users, significantly more insecurely attached users would display an Internet addiction. Descriptive data revealed that the majority of participants with a tendency for Internet addiction (n=24) were categorized as insecurely attached, and only a few (n=6) were securely attached. An opposite trend can be detected when considering subjects without Internet addiction tendencies. More subjects categorized as securely attached (n=115) showed no tendencies for an Internet addiction than insecurely attached subjects (n=99).

A chi-square test was conducted. Unsurprisingly, it was found that insecurely attached subjects differ significantly from securely attached subjects in their tendency for Internet addiction (χ^2_1=12.0, *P*=.003). Furthermore, differences regarding Internet addiction tendencies between the five attachment categories were significant (χ^2_4=27.09, *P*=.004). Most strikingly, an ambivalent attachment style was associated with Internet addiction. More than two-third (n=21) of all subjects, who showed Internet addiction tendencies (n=30), could be allocated to this category, although occupation of this category was generally rather low (n=52).

Attachment Style and Motives for Use

A univariate test of between-subjects effects (analysis of variance, ANOVA) was conducted to investigate the relationships between attachment style and motives for Internet use. A significant result was found for two motives, that is, *anonymity* (F_4=2.82, *P*=.02) and *emotional support* (F_4=3.16, *P*=.03). For subjects with an insecure attachment, the motives “emotional support” (mean 2.66) and “anonymity” (mean 2.36) were significantly more important than for subjects with secure attachment (emotional support, mean 2.03; anonymity, mean 1.67). To check for significant differences between the three

XSL•FO
RenderX

attachment styles, post hoc single comparisons were performed. Bonferroni correction was used to adjust the significance threshold for multiple comparisons. There were significant differences between users with the attachment style *ambivalent-closed* and *conditionally secure* (P <.001) and *secure* (*P*=.02) with respect to the motives anonymity and *emotional support*. Ambivalent-closed participants showed significantly higher values in the motives *anonymity* (mean 2.84) and *emotional support* (mean 2.56) than *conditionally secure* (anonymity, mean 2.10; emotional support, mean 1.68) or *secure* users (anonymity, mean 1.79; emotional support, mean 1.63).

Internet Addiction and Online Relationship Motives

It was checked whether people with Internet addiction tendencies differed significantly in their online relationship motives compared with normal Internet users. A comparison of means revealed significant differences in the online relationship motives *anonymity* (t_{159}=−4.42, *P*=.003), *simplified communication* ($t_{39.72}$=−3.38, *P*=.006), *emotional support* (t_{159}=−3.74, *P*<.001), *social compensation* ($t_{27.72}$=−2.13, *P*=.04), and *escapism* (t_{159}=−4.88, *P*<.001).

For all motives, participants with Internet addiction tendencies had higher mean values (see Figure 1).

Attachment Style and Internet Services

A multivariate analysis of variance (MANOVA) was performed to examine the relationship between attachment style and Web-based services. Using the conservative Pillai trace, no significant main effect for the factor attachment style could be found. Thus, the hypothesis that the use of various Web-based services and apps is associated with different binding styles can be rejected.

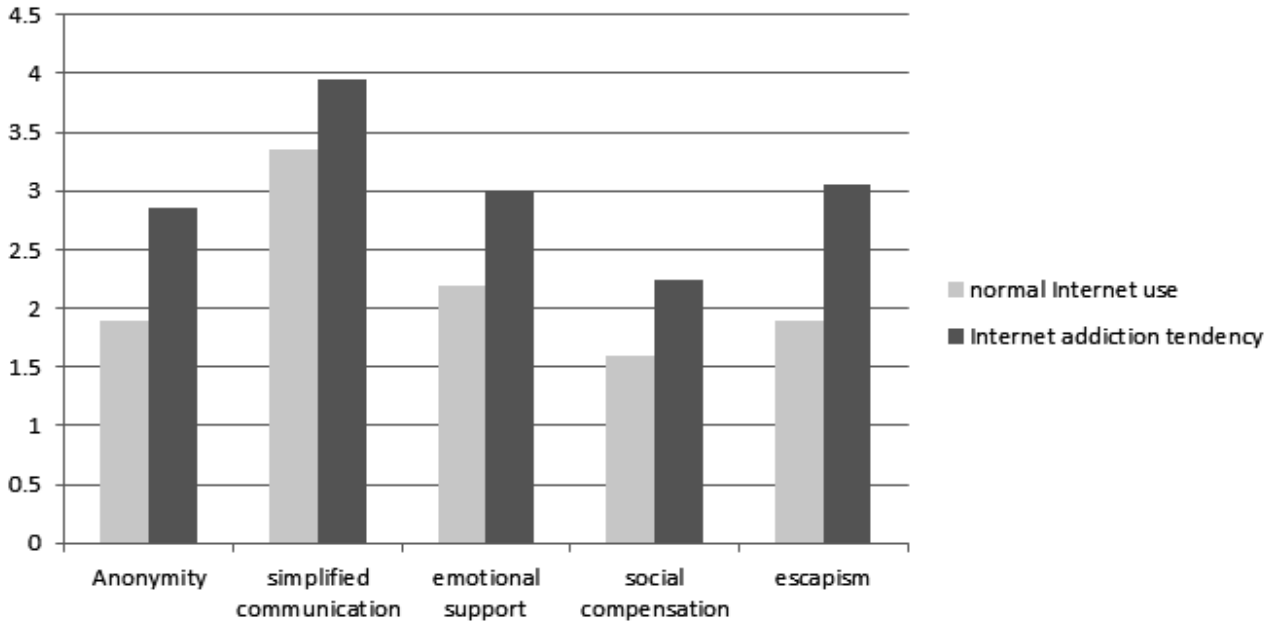
Internet Addiction Tendencies and Internet Services

There were significant differences between groups with and without Internet addiction tendencies with regard to the use of Web-based sex services ($t_{32.219}$=−3.20, *P*=.002, *d*=0.84) and the use of chats and forums (t_{242}=−2.09, *P*=.04, *d*=0.40). The comparison of means showed that both services were more frequently used by subjects with Internet addiction tendencies (forums or chats: mean 2.93, SD 1.01; Web-based sex services: mean 2.1, SD 1.12) than by modest users (forums or chats: mean 2.5, SD 1.06; Web-based sex services: mean 1.43, SD 0.70).

Rorschach Inkblot Test

The analysis of the Rorschach protocols gave the following results: patients who reported a tendency to pathological Internet use showed a significantly lower degree of sociability (as measured by H% [percentage of human figures], *P*=.05, Mann-Whitney *U* test), whereas a higher level of accuracy of perception (measured on F+% [form level], *P*=.05, Mann-Whitney *U* test) occurred in the control group of patients who denied a tendency to pathological Internet use. Examination of the homogeneity of variance using Levene test proved the parameters of the percentage of intermediate figure responses (“S-responses”) that are seen as an indicator for the potential of aggression, as significant (*P*<.001). In the group of abusive patients, the minimum percentage of S-responses was at 0% and the maximum at 38.6%, whereas in the group of inconspicuous Internet users, the minimum was at 10.64% and the maximum was at 25%. Therefore, the group of abusive Internet users can be regarded not only as particularly aggression-inhibited, but also as aggressive patients. Of particular interest in this context is the common assumption that the online world can both dissipate and fuel aggression.

Figure 1. Internet addiction and online relationship motives.



XSL•FO
RenderX

Discussion

Principal Findings

Although a growing body of scientific literature highlights that Internet addiction is a serious health problem, there have been no etiopathological studies to support this claim. Since an association between secure or insecure attachment and substance dependence is well-documented, the aim of this study was to examine whether people differ in their Internet addiction tendencies in regards to their attachment style.

The assumption that insecure attached people show higher Internet addiction tendencies could be confirmed. Ambivalent attachment styles were particularly associated with pathological Internet usage tendencies. For ambivalent-closed attached people, the motives *anonymity* and *social support* were significantly more important than for secure and conditionally secure users. Furthermore, participants with Internet addiction tendencies usage identified anonymity, emotional support, escapism, and social compensation as important relationship motives for Internet use.

However, attachment styles were not associated with the Web-based services used. Nonetheless, there was a difference in the use of Web-based services with regards to the Internet addiction tendencies. In particular, participants with Internet addiction tendencies engaged significantly more in Web-based chats and forums than normal users. Thus, the possibility of Web-based communication seems to play a prevalent role in the context of pathological Internet use.

It appears therefore that, in comparison with Web-based services used, motivational factors relating to online relationships can best explain this finding. In contrast, the relationship between the Web-based services used and the tendency toward Internet addiction can be attributed to the individual addictive potential of the particular Web-based service in question.

Overall, the findings of the Rorschach Inkblot Test are in line with the results of the Web-based survey, in which abusive Internet use presents itself as a function of insecure attachment and impaired interpersonal relationships, that is, the preference for chats and forums by study participants with tendencies toward Internet addiction as a result of an infantile relationship that has not reached the level of group compatibility. Chats and forums offer fictional contact with a group, while communication between users remains in virtual space, which in turn significantly weakens the experience of group presence. Thus, communication in chat-rooms or forums is within the framework of a group constellation, however, always experienced at the level of primary intersubjectivity, that is, within the relationship between I and You. Considering interpersonal skills as an expression of attachment capability [39], it is striking that the intersubjectivity of abusive Internet users has not reached the level of maturity of secondary intersubjectivity [40,41].

These results illustrate the escapist and social-compensatory function associated with pathological Internet usage. Although ambivalent-closed users demonstrate difficulties with acceptance and opening up to others, a parallel desire to connect with others also exists [42]. The media-ecological framework model according to Döring [3] and the uses and gratifications [43] approach can help to understand these findings. On this basis, it can be assumed that the Internet offers specific opportunities for interaction that people with a high Internet addiction tendency do perceive as sufficient forms of communication. Consequently, they report Web-based communication to be more easily accessible and understandable. Apparently, these people appear to be able to use the medium to their advantage, they seem to be able to build a relationship world online in which they can experience emotional support and balance social restrictions. At the same time, the network offers the possibility to temporarily remove a burden from a distressed reality. The factor “anonymity” therefore takes a position of particular importance. The anonymity of the Internet allows for a new presentation of the self, whereby ambivalent-closed users are able to compensate for fears associated with acceptance. At the same time, however, the anonymity provides a platform for online disinhibition [44], which may increase the willingness to open up to others. It appears that it is predominantly users with an ambivalent-closed style attachment who use the Internet in order to compensate for “real” deficits, demonstrating therefore the clearest trend toward pathological Internet use [45]. The importance of relationship motives for this group supports the hypothesis that a socially compensatory component is key for a high Internet addiction tendency.

The satisfaction of these motives is in line with the needs relevant in the context of media usage postulated in Schramm and Hasebrink [46]. Above all, social needs, as well as the search for relaxation and escapism can be satisfied online. At the same time, the fulfillment of an existing need for social affection in the context of an otherwise unfavorable binding pattern plays a particular important role. At this point, it becomes clear to what extent the cognitive-behavioral model of the Internet addiction is appropriate in this context. Although on the one hand side attachment styles [47] can have an unfavorable influence on the self-concept, on the other side online social support and social compensation can act as a reinforcing stimulus. This can result in cognitions suggested by Davis [11], which view the self and others only online pleasantly. Strengthening of this experience might result in an abusive and even addictive Internet usage behavior.

Overall, this study allows for important conclusions about the background of pathological Internet usage to be made. As a consequence, the presented findings are indicatory for future research regarding the etiopathology of Internet addiction.

Limitations

Methodological issues limit the validity of the research results. For example, the lack of participants classified as “addicted” marks a key problem of this study. Therefore, conclusions can only be drawn for problematic, subclinical Internet use. In light of the generally low prevalence of the disorder of only 3-5% [9], it is recommended to conduct another study with a specially selected sample. It is also important to point out that users of Web-based games, gambling, and sex services were not adequately represented in the sample. This one-sided sample composition can be explained by solely recruiting participants

XSL•FO
RenderX

on forums and social networks. Moreover, it would be important to find out to what extent abusive and pathological Internet use differs (ie, regarding comorbidity, used Web-based services and further disease progression). A long-term monitoring of users with Internet addiction tendencies would be required to reveal whether attachment style can act as a disposition and consequently facilitate the development of an Internet addiction. As such, adequate preventive measures could then be applied at an early stage.

In all Web-based surveys, the sample composition constitutes another issue limiting the value of the found results [48]. It may be that Internet users, who are striving to relativize the negative image of Internet dependency, predominantly responded to the survey. A related concern is the limited possibilities to gain information in a self-assessment process typical for Web-based studies. However, it should be noted that while the statistical evaluation can be regarded as objective, the analysis of values and their interpretation cannot be assumed for the data itself [49], due to the fact that every form of self-evaluation has an intentional or unintentional bias. This can include self-deception, simulation or dissimulation, or indeed socially desirable answers [41].

For the Rorschach test, it has to be taken into account that projective techniques are differing from structured tests in stimulus and response and therefore remain "problematic instruments from a psychometric standpoint" [50]. In projective techniques, the stimuli used are more ambiguous than in structured tests. Although items of structured personality tests based on self-observation also bear a certain degree of ambiguity, for example, by using the term "often," that can be interpreted in various ways, projective techniques, in general, are providing a much wider freedom of response and consecutively provoking a much wider response variety in nature and number going together with a complex procedure of scoring. Therefore, they are much more vulnerable by interpreter´s scope of accuracy.

Implications for Therapy

A tendency to pathological Internet use was associated with an insecure attachment style and limited interpersonal relationships; it therefore seems plausible to take therapeutic measures to help patients address real-life deficits, that is, this can be achieved by a therapeutic relationship with the therapist as a "substitute attachment figure" [28], or in a group therapy, where the therapeutic community can also provide corrective relationship experiences [51], and finally the results of the Rorschach test highlight the need for secondary and tertiary intersubjectivity in patients [52].

Conflicts of Interest

None declared.

References

1. Czajka S, Mohr S. Internetnutzung in privaten Haushalten in Deutschland. Wirtschaft und Statistik 2009;6:553-560.
2. InitiativeD21. 2012. (N)onliner Altlas- Eine Topographie des digitalen Grabens durch Deutschland URL: http://www.initiative21.de/portfolio/nonliner-atlas [accessed 2017-04-20] [WebCite Cache ID 6prVXIzpx]
3. Döring N. Sozialpsychologie des Internet: Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen. 2. vollständig überarbeitete Ausg. Göttingen: Hogrefe; 2003.
4. Eichenberg C, Kühne S. Einführung Onlineberatung und -therapie: Grundlagen, Interventionen und Effekte der Internetnutzung. München: UTB; 2014.
5. Sueki H, Eichenberg C. Suicide bulletin board systems comparison between Japan and Germany. Death Stud 2012 Jul;36(6):565-580. [Medline: 24563935]
6. Eichenberg C, Schott M. An empirical analysis of Internet message boards for self-harming behavior. Arch Suicide Res 2016;15:1-15. [Medline: 27845618]
7. Perren S, Corcoran L, Cowie H. Tackling cyberbullying: review of empirical evidence regarding successful responses by students, parents, and schools. Int J Conf Violence 2012;6:283-292. [doi: 10.4119/UNIBI/ijcv.244]
8. Eichenberg C, Wolters C. Gesundheitsängste und Internet? Cyberchondrie, ein modernes Symptom? Neurotransmitter 2013;7:28-32.
9. Müller KW, Wölfling K. Pathologische computerspiel- und internetnutzung: der forschungsstand zu phänomenologie, epidemiologie, diagnostik und komorbidität. Suchtmed 2010;12:1-11.
10. Ko CH, Yen JY, Yen CF. The association between internet addiction and psychiatric disorder: a review of the literature. Eur Psychiatry 2013;27:1-8. [doi: 10.1016/j.eurpsy.2010.04.011] [Medline: 22153731]
11. Davis RA. A cognitive-bahvioral model of pathological internet use. Comput Human Behav 2001;17:187-195. [doi: 10.1016/S0747-5632(00)00041-8]
12. Festl R, Scharkow M, Quandt T. Problematic computer game use among adolescents, younger and older adults. Addiction 2013 Mar;108(3):592-599. [doi: 10.1111/add.12016] [Medline: 23078146]
13. Gainsbury S. Internet gambling: Current research findings and implications. New York: Springer Science & Business Media; 2012.
14. Dhuffar MK, Griffiths MD. A systematic review of online sex addiction and clinical treatments using CONSORT evaluation. Curr Addict Rep 2015;2:163-174. [doi: 10.1007/s40429-015-0055-x]

XSL•FO
RenderX

15. Kuss DJ, Griffiths MD. Online social networking and addiction--a review of the psychological literature. Int J Environ Res Public Health 2011 Sep;8(9):3528-3552 [FREE Full text] [doi: 10.3390/ijerph8093528] [Medline: 22016701]
16. Meerkerk GJ, Van Den Eijnden R, Garretsen H. Predicting compulsive internet use: it's all about sex!. Cyberpsychol Behav 2006 Feb;9(1):95-103. [doi: 10.1089/cpb.2006.9.95] [Medline: 16497122]
17. Eichenberg C, Klemme A, Theimann T. Internetsuchtin neues störungsbild? ein überblick zu neueren befunden. Psychomed 2003;2:100-105.
18. Ko CH, Liu GC, Hsiao S, Yen J, Yang M, Lin W, et al. Brain activities associated with gaming urge of online gaming addiction. J Psychiatr Res 2009 Apr;43(7):739-747. [doi: 10.1016/j.jpsychires.2008.09.012] [Medline: 18996542]
19. Weinstein A, Lejoyeux M. New developments on the neurobiological and pharmaco-genetic mechanisms underlying internet and videogame addiction. Am J Addict 2015 Mar;24(2):117-125. [doi: 10.1111/ajad.12110] [Medline: 25864599]
20. Burkett JP, Young LJ. The behavioral, anatomical and pharmacological parallels between social attachment, love and addiction. Psychopharmacology (Berl) 2012 Nov;224(1):1-26 [FREE Full text] [doi: 10.1007/s00213-012-2794-x] [Medline: 22885871]
21. Grüßer-Sinopoli SM, Thalemann CN. Verhaltenssucht: Diagnostik, Therapie, Forschung. Bern: Huber; 2006.
22. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th edition. Arlington: American Psychiatric Publishing; 2006.
23. Wölfling K, Jo C, Bengesser I. Computerspiel-und Internetsucht. Ein kognitiv-behaviorales Behandlungsmanual. Stuttgart: Kohlhammer; 2013.
24. King DL, Delfabbro PH. The cognitive psychology of internet gaming disorder. Clin Psychol Rev 2014 Jun;34(4):298-308 [FREE Full text] [doi: 10.1016/j.cpr.2014.03.006] [Medline: 24786896]
25. Essig T. The addiction concept and technology: diagnosis, metaphor, or something else? a psychodynamic point of view. J Clin Psychol 2012 Nov;68(11):1175-1184. [doi: 10.1002/jclp.21917] [Medline: 22976153]
26. Dau W, Hoffmann JDG, Banger M. Therapeutic interventions in the treatment of problematic internet use? experiences from Germany. In: Montag C, Reuter M, editors. Internet Addiction. Switzerland: Springer International Publishing; 2015:183-217.
27. Borhani Y. substance abuse and insecure attachment styles: a relational study. Lux 2013;2(1):1-13 [FREE Full text] [doi: 10.5642/lux.201301.04]
28. Schindler A. Bindung und Sucht. In: Urban M, Hartmann HP, editors. Bindungstheorie in der Psychiatrie. Göttingen: Vadenhoeck & Ruprecht; 2005:96-106.
29. Te Wildt BT, Putzig I, Vikicevic A. Störungen von Selbsterleben und Beziehungsverhalten bei Menschen mit Internetabhängigkeit. Sucht 2011;57:17-26.
30. Höger D, Buschkämper S. Der bielefelder fragebogen zu partnerschaftserwartungen. J Individ Differ 2002;1:83-98.
31. Wölfling K, Müller KW, Beutel ME. Skala zum Onlinesuchtverhalten (OSV-S). In: Muecken D, Teske A, Rehbein F, editors. Prävention, Diagnostik und Therapie von Computerspielabhängigkeit. Lengerich: Pabst Science Publishers; 2010.
32. Wang CC, Chang YT. Cyber relationship motivescale development and validation. Soc Behav Pers 2010;38:289-300.
33. Mihura JL, Meyer GJ, Dumitrascu N, Bombel G. The validity of individual Rorschach variables: systematic reviews and meta-analyses of the comprehensive system. Psychol Bull 2013 May;139(3):548-605. [doi: 10.1037/a0029406] [Medline: 22925137]
34. Rorschach H. Psychodiagnostik. Bern: Presses Universitaires de France; 1948.
35. Bohm E. Lehrbuch der Rorschach-Psychodiagnostik für Psychologen, Ärzte und Pädagogen. Bern: Verlag Hans Huber; 1990.
36. Rosenthal R, Hiller JB, Bornstein RF, Berry DT, Brunell-Neuleib S. Meta-analytic methods, the Rorschach, and the MMPI. Psychol Assess 2001 Dec;13(4):449-451. [Medline: 11793890]
37. Society for Personality Assessment. The status of the rorschach in clinical and forensic practice: an official statement by the board of trustees of the society for personality assessment. J Pers Assess 2005;85(2):219-237. [doi: 10.1207/s15327752jpa8502_16] [Medline: 16171424]
38. Sax LJ, Gilmartin SK, Bryant AN. Assessing response rates and nonresponse bias in web and paper surveys. Res High Educ 2003;44:409-432.
39. Bowlby J. The Making and Breaking of Affectional Bonds. New York: Routledge; 2005.
40. Rieken B, Sindelar B, Stephenson T. Individualpsychologische Entwicklungstheorie und Krankheitslehre. In: Psychoanalytische Individualpsychologie in Theorie und Praxis: Psychotherapie, Pädagogik, Gesellschaft (German Edition). Wien: Springer; 2011.
41. Fonagy P. Attachment Theory and Psychoanalysis. New York: Other Press LLC; 2001.
42. Müller K. Spielwiese Internet. Sucht ohne Suchtmittel. Berlin, Heidelberg: Springer; 2013.
43. Aelker L. Der uses-gratifications ansatz. In: Krämer N, Schwan S, Unz D, Suckfüll M, editors. Medienpsychologiechlüsselbegriffe und Konzepte. Stuttgart: Kohlhammer; 2008:17-22.
44. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]

XSL•FO
RenderX

45. Schuhler P. Bindungsdynamische Sichtweise. In: Perty J, editor. Dysfunktionaler und Pathologischer PC-Gebrauch. Göttingen: Hogrefe; 2010:57-64.
46. Schramm H, Hasebrink U. Fernsehnutzung und Fernsehwirkung. In: Mangold R, Vorderer P, Bente G, editors. Lehrbuch der Medienpsychologie. Göttingen: Hogrefe; 2004:465-492.
47. Bartholomew K, Horowitz LM. Attachment styles among young adults: a test of a four-category model. J Pers Soc Psychol 1991;61(2):226-244.
48. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):e13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
49. Bowlby J. The Making and Breaking of Affectional Bonds. New York: Routledge; 2005.
50. Lilienfeld SO, Wood JM, Garb HN. The Scientific Status of Projective Techniques. Psychol Sci Public Interest 2009;1(2):27-66.
51. Schindler A. Bindung und Sucht. In: Hartmann HP, Urban M, editors. Bindungstheorie in der Psychiatrie. Göttingen: Vadenhoeck & Ruprecht; 2005:96-106.
52. Schuhler P, Vogelsang M, Perty J. Pathologischer PC-/Internetgebrauch: krankheitsmodell, diagnostische und therapeutische ansätze. Psychotherapeut 2009;54:187-192.

Abbreviations

AAS: Adult Attachment Scale
ANOVA: analysis of variance
CRMS-D: Cyber Relationship Motive Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
MANOVA: multivariate analysis of variance
RIM: Rorschach Inkblot Method

Edited by G Eysenbach; submitted 25.09.16; peer-reviewed by J Tayaben, J Apolinário-Hagen; comments to author 03.11.16; revised version received 20.01.17; accepted 08.02.17; published 17.05.17.

Please cite as:
Eichenberg C, Schott M, Decker O, Sindelar B
Attachment Style and Internet Addiction: An Online Survey
J Med Internet Res 2017;19(5):e170
URL: http://www.jmir.org/2017/5/e170/
doi:10.2196/jmir.6694
PMID:28526662

©Christiane Eichenberg, Markus Schott, Oliver Decker, Brigitte Sindelar. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 17.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Toward the Design of Evidence-Based Mental Health Information Systems for People With Depression: A Systematic Literature Review and Meta-Analysis

Fabian Wahle[1*], MSc; Lea Bollhalder[2*], MA; Tobias Kowatsch[3], PhD; Elgar Fleisch[2], PhD

[1]Center for Digital Health Interventions, Department of Management, Technology and Economics, ETH Zürich, Zürich, Switzerland

[2]Institute of Technology Management, University of St Gallen, St Gallen, Switzerland

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

[*]these authors contributed equally

Corresponding Author:
Tobias Kowatsch, PhD
Center for Digital Health Interventions
Institute of Technology Management
University of St Gallen
Dufourstrasse 40a, Bu¨ro 1-236
St Gallen, 9000
Switzerland
Phone: 41 712247244
Fax: 41 712247244
Email: tobias.kowatsch@unisg.ch

Abstract

Background: Existing research postulates a variety of components that show an impact on utilization of technology-mediated mental health information systems (MHIS) and treatment outcome. Although researchers assessed the effect of isolated design elements on the results of Web-based interventions and the associations between symptom reduction and use of components across computer and mobile phone platforms, there remains uncertainty with regard to which components of technology-mediated interventions for mental health exert the greatest therapeutic gain. Until now, no studies have presented results on the therapeutic benefit associated with specific service components of technology-mediated MHIS for depression.

Objective: This systematic review aims at identifying components of technology-mediated MHIS for patients with depression. Consequently, all randomized controlled trials comparing technology-mediated treatments for depression to either waiting-list control, treatment as usual, or any other form of treatment for depression were reviewed. Updating prior reviews, this study aims to (1) assess the effectiveness of technology-supported interventions for the treatment of depression and (2) add to the debate on what components in technology-mediated MHIS for the treatment of depression should be standard of care.

Methods: Systematic searches in MEDLINE, PsycINFO, and the Cochrane Library were conducted. Effect sizes for each comparison between a technology-enabled intervention and a control condition were computed using the standard mean difference (SMD). Chi-square tests were used to test for heterogeneity. Using subgroup analysis, potential sources of heterogeneity were analyzed. Publication bias was examined using visual inspection of funnel plots and Begg's test. Qualitative data analysis was also used. In an explorative approach, a list of relevant components was extracted from the body of literature by consensus between two researchers.

Results: Of 6387 studies initially identified, 45 met all inclusion criteria. Programs analyzed showed a significant trend toward reduced depressive symptoms (SMD –0.58, 95% CI –0.71 to –0.45, $P<.001$). Heterogeneity was large ($I2\geq76$). A total of 15 components were identified.

Conclusions: Technology-mediated MHIS for the treatment of depression has a consistent positive overall effect compared to controls. A total of 15 components have been identified. Further studies are needed to quantify the impact of individual components on treatment effects and to identify further components that are relevant for the design of future technology-mediated interventions for the treatment of depression and other mental disorders.

(J Med Internet Res 2017;19(5):e191) doi:10.2196/jmir.7381

XSL•FO
RenderX

KEYWORDS

literature review; mental health; design feature; depression; information systems

Introduction

Over the last decade, numerous technology-mediated treatments for mental health disorders have been developed and tested in controlled trials. They form a subset of what the World Health Organization in 2005 coined "mental health information system" (MHIS). A MHIS "is a system for collecting, processing, analyzing, disseminating, and using information about a mental health service and the mental health needs of the population it serves" [1]. Although such a system does not necessarily need to rely on computerization, evidence from recent years suggests that technology-mediated MHIS holds vast opportunities in terms of much-needed scalability while ensuring treatment effectiveness. This was shown by a number of reviews and meta-analyses on computerized and Internet-delivered MHIS for mental health disorders in general [2-4] and for depression in particular [5-8].

Despite this success, it remains unclear what guides the design of MHIS and the choice of components that support existing evidence-based mental health interventions. Existing research postulates a variety of such components that show an impact on utilization of technology-mediated services and treatment outcome in general [9]. Morrison et al [10] more specifically assessed the effect of isolated components on the results of Web-based MHIS interventions and introduced that "there has been relatively little formal consideration of how differences in the design of an intervention (ie, how the content is delivered) may explain why some interventions are more effective than others." They defined four core interactive system components that may mediate the effects of intervention design on outcomes: (1) social context and support, (2) contacts with intervention, (3) tailoring, and (4) self-management. A study by Whitton et al [11] examined the associations between symptom reduction and use of components across computer and mobile phone platforms for people with depression for one specific computerized intervention. They found that the incorporation of alert-based components, such as reminders and short motivational messages, quotes, or facts, that were sent by email or short message service (SMS) text message showed greater therapeutic gain compared with programs that did not make use of these components. At large, it was found that reminders play a decisive role in the engagement of users in mental health interventions and are a cost-effective approach for engaging users [11-13]. In addition, Landenberger and Lipsey [14] studied the relationship between specific components and the effects of computerized cognitive behavior therapy (CBT) on the recidivism of adult and juvenile offenders. Despite these findings, there remains uncertainty with regard to which components of technology-mediated interventions for mental health exert the greatest therapeutic gain across MHIS targeting people with depression. An analysis reviewing trials of technology-adaptable interventions for the treatment of depression in adults with cognitive impairments is still underway [15].

A recent preliminary literature review by Wahle and Kowatsch [16] aimed at identifying a first set of generic components for the design of MHIS for people with depression and acted as a starting point for this review in further identifying relevant components. Similar to Morrison et al [10], they hypothesized that the channel of delivery (eg, mobile phone-based, Web-based), the degree of peer support, the availability of subsidiary support, the degree of tailoring, and the existence of gamification elements likely have an impact on treatment outcome, independent from the underlying therapeutic approach [16].

This work aims at extending this list in a systematic manner and to seek evidence for the effectiveness of each of the newly identified components. By nature, MHIS represent persuasive systems. Persuasive systems may be defined as "computerized software or information systems designed to reinforce, change, or shape attitudes or behaviors or both without using coercion or deception" [17]. Therefore, we drew on Oinas-Kukkonnen and Harjumaa's generic model of persuasive systems design [18] to identify further potentially relevant components when investigating the influence on the effectiveness of MHIS. Components the authors deemed meaningful that did not strictly follow Oinas-Kukkonnen and Harjumaa's proposed system features were also added.

In summary, this systematic review and meta-analysis aims to add to the current body of literature by providing a systematic update in evaluating the overall effectiveness of technology-mediated treatments for depression, as well as identifying the current set of system components in use, which has not previously been conducted on a systematic review targeting depression treatments.

Methods

An electronic search was conducted in MEDLINE, PsycINFO, and the Cochrane Library. Titles and abstracts of the identified randomized controlled trials (RCTs) were screened using predefined inclusion criteria. We independently assessed the eligibility for inclusion of all potentially relevant studies identified by the search strategy. Any disagreements were resolved by discussion among the authors. Manually screening reference lists for additional studies of relevance and tracing trials was aimed to obtain further studies possibly eligible for inclusion. Included RCTs were categorized by (1) location, (2) total number of patients randomized, (3) target condition (depression or depression comorbid with anxiety), (4) depression severity, (5) age of participants, (6) name and type of intervention, (7) type of comparator, (8) study quality, and (9) MHIS system components (see Data Extraction). A change in validated depression scores was used as the primary outcome. Data were collected from eligible trials and transferred to a data extraction table. Study quality was assessed using the widely used Jadad scale [19], additionally checking each trial for appropriate randomization, blinding of patients, as well as

XSL•FO
RenderX

dropouts and withdrawals (see Assessment of Methodological Quality).

For the implementation of this systematic review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement was used [20]. Methods of the analysis and inclusion criteria were specified in advance and documented in a protocol (can be provided on request). However, the protocol also included the evaluation of other mental disorders. Due to the large number of RCTs identified and the resulting high degree of heterogeneity, it was decided that mental disorders other than depression were not to be evaluated in this systematic review.

Information Sources

Electronic searches were conducted in MEDLINE, PsycINFO, and the Cochrane Controlled Trials Register. Medical Subject Headings (MeSH) and relevant text word terms were used to identify relevant studies (see Search Strategy). The last search was run on September 1, 2016. Reference lists of systematic reviews and articles identified were manually checked for relevant entries.

Search Strategy

Search terms for depression were used to scan all trials registers and databases outlined previously. Additional terms for a range of delivery methods (eg, online, Internet, Web, computer, phone) and terms that specify the type of intervention (eg, cognitive behavioral, psychodynamic, interpersonal, psychoeducation) were applied. Further search terms were utilized to limit the search to studies of therapeutic interventions (eg, therapy, psychotherapy, intervention, treatment) and to RCTs. Figure 1 gives an overview of the terms used in this literature search.

As a consequence of the protocol also including the evaluation of other mental disorders, the search strategy was refined during the course of the review to limit our study to depression. A compilation of the preliminarily defined search terms is given in Figure 1.

Figure 1. Keyword combinations used in the literature search process.

Keyword combinations						
Title (OR)		Abstract/Title (OR)		Abstract/Title (OR)		Full text (OR)
Mental health Mood disorder(s) Mental disorder(s) Mental illness Depression Mood Psychological disorder(s) Depressive disorder(s)	(AND)	Intervention Support Therapy Treatment Psychotherapy Psychodynamic therapy Psychoeducation Cognitive-behavioural Cognitive-behavioural Training Management care Assistance Help Exercise Skills Coping skills	(AND)	Computer Computerised Computerized Internet Web Online Phone Mobile phone Smartphone Ehealth Mhealth e-health m-health Epsychology e-psychology Telecare e-mental health Technologies Technology Information-system(s)	(AND)	Rct Randomized controlled trial(s) Randomised controlled trial(s) randomized randomised

Study Selection

The process of study selection required an eligibility check for each article identified. Eligibility of studies was assessed by reviewing the abstracts of the references identified by the search strategy. Full texts were additionally screened when necessary. In case of doubt, any disagreements and ambiguous articles were discussed among the authors, and eligibility of studies was decided by consensus.

Eligibility Criteria

Type of Study Design

Any parallel-group RCT published in English between January 2000 to September 2016 in a peer-reviewed journal was considered eligible for inclusion in this systematic analysis and synthesis.

Type of Participants

Studies were included if they evaluated adults or adolescents who had any of the following conditions: mild to severe depression, excluding depression co-occurring with non-

XSL•FO
RenderX

Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) (DSM-V) [21] disorders or depression caused by environmental factors (eg, traumatic events), or depression comorbid with anxiety disorders. Studies of participants with other mental health problems (eg, schizophrenia, sleep disorders, personality disorders) were excluded to reduce the risk of heterogeneity.

Types of Interventions

Studies that assessed any form of technology-based intervention for depression were included in this systematic review. To assure a sufficient degree of comparability, a necessity for meta-analyses, we only included interventions in which there existed evidence for comparable outcomes for the treatment of depression. This was decided based on literature and consultation from two licensed psychotherapists. These included (1) psychotherapy (eg, cognitive behavioral therapy [22], interpersonal therapy [23], problem-solving therapy [24], supportive therapy [25]), (2) psychoeducation [26], and (3) exercise/physical activity [27] which showed results on par with a pharmacological treatment. Additional administration of drugs or procedures was allowed. The following channels for service delivery were considered eligible for inclusion: (1) offline delivery, including all interventions that did not require an Internet connectivity to provide care (eg, stand-alone computers); (2) Web-supported delivery, including interventions that made use of the Internet to deliver services (eg, using interactive websites to provide interventions or online self-help forums); and (3) mobile phone, smartphone, and tablet delivery, including all treatment programs that made use of mobile phone or tablet apps. The range of apps ranged from simple message passing to feature-rich multimedia interventions.

To meet the secondary inclusion criteria, all eligible clinical trials (according to the aforementioned eligibility criteria) were then inspected with regard to their technical feasibility. Criteria for studies being classified as technically feasible were the following: (1) to provide a methodologically structured format of care to the participants, the intervention must have adhered to a manual, protocol, or structured approach (with clearly stated processes, program structure, and objectives) and (2) treatment must not have been primarily based on face-to-face interaction, group discussion, or any other form of treatment that required personal interaction. Specific accompanying service configurations, which facilitated interaction and support with the study team and/or peer groups made possible by technology, were eligible for inclusion. In general, mental health care services, such as psychotherapeutic or behavioral interventions, were deemed suitable for the provision in a guided or nonguided format and were thus considered technically feasible. Only trials of interventions that were considered to be technically viable were included in this systematic review.

Types of Endpoints and Outcome Measures

In terms of types of endpoints, RCTs assessing the impact on symptoms of depression were taken into consideration. Our primary outcome measures of interest were symptoms of depression. Trials were eligible for inclusion if they have evaluated the severity of depression pre- and postintervention using one or both of these valid assessment scales: (1) the Beck Depression Inventory (BDI, BDI-1A, or BDI-II) or (2) the Patient Health Questionnaire (PHQ, PHQ-9, or PHQ-2).

The BDI is a widely used psychometric test to assess characteristic attitudes and symptoms of depression. The test consists of 21 multiple-choice self-report questions and is employed by the majority of researchers and health care professionals to measure depression severity [28]. The PHQ is a brief, self-administered assessment tool for screening and diagnosis as well as for selecting and monitoring treatment. It is part of the longer PHQ that integrates *DSM-IV* depression criteria with other leading major depressive symptoms into a concise self-report instrument [29].

Types of Controls

Studies were included if they compared technology-based interventions for depression to either waiting-list control (WLC), treatment as usual (TAU), or any other form of treatment for depression. The RCTs were also deemed eligible if they compared one channel of service delivery to another channel of delivery. Trials were further considered eligible if they analyzed interventions that compared different forms of subsidiary support.

System Components

For each identified component, we provide a rational for inclusion in the Results section. Despite some components appearing to be derived from underlying psychological theory, they were included because they were either enabled or administered by technology. For each study included in the systematic review, we determined the presence of defined system components for later analysis.

We would like to emphasize that our analysis of system components is not comprehensive and was only conducted to the degree possible based on published information in the respective literature included for the meta-review.

Data Extraction

Data collection tables were predeveloped and subsequently refined during the process of data extraction. The following information were collected from every article.

Comorbidity

The occurrence of comorbidity was recorded. A difference was made between depression/depressive symptoms only and depression/depressive symptoms co-occurring with anxiety.

Characteristics of Intervention

The name of the therapy program and the year and location of the study were recorded. Also, information on the duration of the intervention in weeks and follow-up in months was collected. In addition, program structure and format, as well as the number of modules, were recorded. For each intervention, we further gathered any information on the aim of the intervention (inferred from the description of the intervention) and the MHIS channel(s) used (eg, online, mobile phone, computer program).

Characteristics of the Control Condition

Where applicable, all relevant information provided on the control condition was recorded.

XSL•FO
RenderX

Sample Characteristics

After applying the inclusion and exclusion criteria, we collected information on the severity of clinical depression at baseline. As a consequence of differences in the reporting of symptom severity, which was used for the inclusion/exclusion of participants, we categorized studies into one of four severity classes.

Similarly, we recorded age inclusion and exclusion criteria of all studies included in this systematic review and categorized studies into five groups (see Table 1). All data on the total size of the study population as well as on the number of participants in the treatment and control group(s) were collected.

Table 1. Categorization of studies according to baseline depression severity and age of participants.

Item	Rating
Depression severity	
PHQ-9	
0	No depression
1-5	Minimal depression
6-9	Mild depression
10-14	Moderate depression
15-19	Moderately severe
≥20	Severe depression
BDI-II	
0-13	No depression
14-19	Minimal depression
20-28	Mild depression
≥29	Severe depression
Depression category	
0	Not reported
1	Mild/minimal to moderate
2	Moderately severe/severe
3	Moderately severe to severe/major depressive episode
Age category	
0	No age restrictions
1	Adults (>16 years of age)
2	Adolescents (14-24 years of age)
3	Older adults (>50 years of age)
4	Adults without older adults (18-75 years of age)

All Relevant Outcomes

We recorded all relevant outcomes reported on at least one of the following scores: the BDI or PHQ score.

Study Quality

Study quality was assessed according to Jadad et al [19] (see Assessment of Methodological Quality).

Service Components

For a quantified overview, the individual system components were either binary coded or, if applicable, kept in original scale.

Assessment of Methodological Quality

The quality of trials was examined according to the Jadad score [19] and use of intention-to-treat analysis for the available endpoints and practice of a blinded endpoint assessment. Further information can be obtained from the protocol and the quality assessment table (see Table 2).

Table2. Study quality: risk of bias in included studies (N=45). (1) Double blinded? (2) Withdrawals and dropouts reported? (3) Method of randomization reported and appropriate? (4) Method of blinding reported and appropriate? (5) Analysis “intention-to-treat”? (6) Assessment of the endpoint blinded?

XSL•FO
RenderX

Table 2.

Study	1	2	3	4	5	6	Total score	Quality rating
Depression								
Agyapong [30]	0.0	1.0	1.0	0.5	1.0	0.5	5.0	Good
Andersson [31]	0.0	1.0	1.0	0.5	1.0	0.5	5.0	Good
Andersson [32]	0.0	1.0	1.0	0.5	1.0	0.5	5.0	Good
Berger [33]	0.0	1.0	1.0	0.0	1.0	0.5	4.5	Good
Burton [34]	0.0	1.0	1.0	0.5	1.0	0.5	5.0	Good
Carlbring [35]	0.0	1.0	1.0	0.0	1.0	0.5	4.5	Good
Choi [36]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Clarke [37]	0.0	0.5	1.0	0.0	1.0	0.0	3.5	Fair
de Graaf [38]	0.0	1.0	0.0	0.0	1.0	0.0	3.0	Fair
Holländare [39]	0.0	1.0	1.0	0.0	0.0	0.0	2.5	Poor
Høifødt [40]	0.0	1.0	1.0	0.5	1.0	1.0	5.5	Good
Johansson [41]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Johansson [42]	0.0	1.0	1.0	0.5	0.0	0.0	3.5	Fair
Kay-Lambkin [43]	0.0	0.5	1.0	0.5	1.0	1.0	5.0	Good
Kessler [44]	0.0	0.5	1.0	0.5	1.0	0.0	4.0	Fair
Kivi [45]	0.0	1.0	1.0	0.5	0.0	1.0	4.0	Fair
Lappalainen [46]	0.0	1.0	1.0	0.0	0.0	0.0	3.0	Fair
Lappalainen [47]	0.0	1.0	0.0	0.0	0.0	0.0	2.0	Poor
Ly [48]	0.0	1.0	1.0	0.0	1.0	1.0	5.0	Good
Meyer [49]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Meyer [50]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Morgan [51]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Moritz [52]	0.0	1.0	0.0	0.5	1.0	0.5	4.0	Fair
Perini [53]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Phillips [54]	1.0	1.0	1.0	1.0	0.5	1.0	6.5	Good
Preschl [55]	0.0	1.0	1.0	0.0	0.5	0.0	3.5	Fair
Richards [56]	0.0	0.5	1.0	0.0	1.0	0.0	3.5	Fair
Richards [57]	0.0	1.0	1.0	0.0	1.0	0.0	3.5	Fair
Ruwaard [58]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Sheeber [59]	0.0	1.0	1.0	1.0	0.0	1.0	5.0	Good
Spek [60]	0.0	1.0	0.0	0.0	1.0	0.0	3.0	Fair
Ström [61]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Titov [62]	0.0	0.5	1.0	0.0	1.0	0.0	3.5	Fair
Titov [63]	0.0	1.0	1.0	0.5	1.0	0.0	4.5	Good
Vernmark [64]	0.0	1.0	1.0	0.5	1.0	1.0	5.5	Good

XSL•FO
RenderX

Study	1	2	3	4	5	6	Total score	Quality rating
Wagner [65]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Watts [66]	0.0	0.5	1.0	0.0	0.0	0.0	2.5	Poor
Agyapong [30]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Depression and anxiety								
Agyapong [30]	0.0	1.0	1.0	0.5	1.0	1.0	5.5	Good
Andersson [31]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Andersson [32]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair
Berger [33]	0.0	1.0	1.0	0.5	1.0	0.0	4.5	Good
Agyapong [30]	0.0	1.0	0.0	0.0	1.0	0.0	3.0	Fair
Andersson [31]	0.0	1.0	1.0	0.0	0.0	0.0	3.0	Fair
Andersson [32]	0.0	1.0	1.0	0.0	1.0	0.0	4.0	Fair

The quality rating was based on the total score achieved, and studies were categorized into three groups according to their quality scores: (1) good (4.5-7 points), (2) fair (3-4 points), and (3) poor (0-2.5 points). One point was given for every quality criterion met, 0.5 points for an incomplete description of the methodology used, and no points if a quality criterion was not met. As a consequence of only including RCTs in this review, it was expected that every study was described as “randomized” and thus attained at least 1 point on the quality rating score. Achieving a successful blinding in psychotherapy trials is generally considered to be very challenging, and the methods of blinding are seldom described appropriately [67]. As a consequence, we expected that only a minority of studies would reach a score higher than 5 points and consequently set the cut-off scores according to our expectations of study qualities.

Data Synthesis

This systematic review included a broad variety of clinical subpopulations (eg, differences in baseline severity or age) as well as treatment programs and types of comparators. Therefore, the feasibility of conducting a meta-analysis required careful consideration because the calculation of a mean treatment effect across studies could be irrelevant if studies varied significantly with regard to study populations, interventions, comparisons, or methods [68]. The protocol prespecified that if there was an adequate number of comparable studies, a random-effects meta-analysis according to the methodology of DerSimonian and Laird [69] would be conducted for the combined study groups of depression and depression comorbid with anxiety. Tables 3 and 4 provide a summarizing overview of the included studies.

XSL•FO
RenderX

Table 3. Summary of study characteristics, including location, sample size, study name, severity, and age of participants (N=45).

Study	Location	N	Name	Severity	Age
Depression					
Agyapong [30]	Ireland	54	No name	Moderately severe to severe	Adults (≥16 years)
Andersson [31]	Sweden	69	No name	Mild to severe	Adults (≥16 years)
Andersson [32]	Sweden	117	No name	Mild to moderate	Adults (≥16 years)
Berger [33]	Switzerland Germany	76	Deprexis	Mild to severe	Adults (≥16 years)
Burton [34]	Romania Spain Scotland UK	28	Help4Mood	Mild to severe	Adults without older adults (18-75 years)
Carlbring [35]	Sweden	80	Depressions-hjälpen	Mild to moderate	Adults (≥16 years)
Choi [36]	Australia	63	Brighten Your Mood Program	Minimal to moderately severe	Adults (≥16 years)
Clarke [37]	USA	160	No name	NR	Adolescents (14-24 years)
de Graaf [38]	Netherlands	303	Colour Your Life	Mild to moderate	Adults without older adults (18-75 years)
Holländare [39]	Sweden	84	No name	Mild	Adults (≥16 years)
Høifødt [40]	Norway	106	MoodGYM (Norwegian Version)	Moderate to severe	Adults without older adults (18-75 years)
Johansson [41]	Sweden	92	SUBGAP	Mild to severe	Adults (≥ 16 years)
Johansson [42]	Sweden	121	No name	Mild to severe	Adults (≥ 16 years)
Kay-Lambkin [43]	Australia	97	SHADE	Mild to severe	Adults (≥ 16 years)
Kessler [44]	UK	297	No name	Mild to severe	Adults without older adults (18-75 years)
Kivi [45]	Sweden	92	Depressions-hjälpen	Mild to moderate	Adults (≥16 years)
Lappalainen [46]	Finland	39	Good Life Compass	Mild to severe	Adults (≥16 years)
Lappalainen [47]	Finland	38	Good Life Compass	Mild to severe	Adults (≥16 years)
Ly [48]	Sweden	93	No name	Mild to severe	Adults (≥16 years)
Meyer [49]	Germany	163	Deprexis	Moderately severe to severe	Adults without older adults (18-75 years)
Meyer [50]	Germany	396	Deprexis	NR	Adults (≥16 years)
Morgan [51]	UK Australia Canada Ireland New Zealand USA	1736	Mood Memos	Mild to severe	Adults (≥16 years)
Moritz [52]	Germany	210	Deprexis	NR	Adults without older adults (18-75 years)
Perini [53]	Australia	48	Sadness	Mild to severe	Adults (≥16 years)
Phillips [54]	UK	637	MoodGym	NR	Adults (≥16 years)
Preschl [55]	Switzerland	40	E-mental Health Butler System	Minimal to moderate	Older adults (≥50 years)
Richards [56]	Ireland	262	Space from Depression	Mild to moderate	Adults (≥16 years)
Richards [57]	Ireland	101	Beating the Blues	Mild to moderate	Adolescents (14-24 years)
Ruwaard [58]	Netherlands	54	No name	Minimal to moderate	Adults (≥16 years)
Sheeber [59]	USA	70	Mom-Net	NR	No age restrictions
Spek [60]	Netherlands	301	Lewinsohn's Coping With Depression Course	Subthreshold depression	Older adults (≥50 years)
Ström [61]	Sweden	48	No name	Mild to moderate	No age restrictions

XSL•FO
RenderX

Study	Location	N	Name	Severity	Age
Titov [62]	Australia	54	Managing Your Mood Course	Mild to moderate	Older adults (≥50 years)
Titov [63]	Australia	141	SADNESS	Mild to severe	Adults (≥16 years)
Vernmark [64]	Sweden	88	No name	mild to moderate	Adults (≥16 years)
Wagner [65]	Switzerland	62	No name	Minimal to severe	Adults (≥16)
Watts [66]	Australia	52	Get Happy (Mobile app of the sadness program)	Mild to moderate	Adults (≥16 years)
Depression and anxiety					
Johansson [70]	Sweden	57	No name	Moderate to severe	Adults (≥16 years)
Mullin [71]	Australia	31	UniWellbeing Course	Minimal to moderate	Adults (≥16 years)
Newby [72]	Australia	109	Worry and Sadness Program	Mild to severe	Adults (≥16 years)
Proudfoot [73]	UK	167	Beating the Blues	NR	Adults without older adults (18-75 years)
Titov [74]	Australia	290	Transdiagnostic Wellbeing Course (TD-CBT) or Disorder-Specific Mood Course (DS-CBT)	Mild to severe	Adults without older adults (18-75 years)
Titov [75]	Australia	93	Wellbeing Course	Moderate to severe	Adults (≥16 years)
Titov [76]	Australia	38	Wellbeing Course	Mild to severe	Adults (≥16 years)

XSL•FO
RenderX

Table 4. Summary of the study treatment and control groups and their relevant scores at baseline and follow-up (N=45).

Study	Treatment group[a]	Control group[a]	Baseline				Follow-up				
			Treatment		Control		Wks	Treatment		Control	
			n	Mean (SD)	n	Mean (SD)		n	Mean (SD)	n	Mean (SD)
Depression											
Agyapong [30]	Supportive text messages sent by a computer + TAU	Thank you text message + TAU	26	31.58 (7.70)	28	31.99 (9.50)	13	24	8.60 (7.90)	26	16.60 (9.80)
Andersson [31]	Guided Web-based CBT	Group CBT	33	24.00 (7.70)	36	25.30 (6.60)	9	32	13.60 (10.10)	33	17.90 (8.80)
Andersson [32]	Web-based CBT+ Web-based discussion group	Web-based discussion group only	36	20.50 (6.70)	49	20.90 (8.50)	10	36	12.20 (6.80)	49	19.50 (8.10)
Berger [33]	Low-intensity therapist-guided, computerized CBT	WLC	25	28.80 (8.20)	26	29.80 (8.60)	10	25	17.30 (10.20)	22	28.50 (9.40)
Burton [34]	Help4Mood (Self-report and biometric monitoring + elements of CBT) + TAU	TAU	14	19.60 (8.10)	13	21.80 (6.80)	4	12	13.90 (8.10)	9	17.60 (6.80)
Carlbring [35]	Web-based behavioral activation and acceptance-based treatment	WLC	40	26.32 (5.97)	40	25.13 (5.19)	8	40	16.65 (8.04)	38	23.43 (7.67)
Choi [36]	Web-based CBT	WLC	28	25.76 (8.53)	30	20.83 (7.58)	8	23	13.48 (9.28)	28	21.27 (7.86)
Clarke [37]	Web-based, pure self-help CBT	TAU	83	10.00 (0.80)	77	10.30 (0.80)	5	58	9.10 (0.70)	58	10.10 (0.70)
de Graaf [38]	Computerized CBT	TAU	100	28.20 (7.70)	103	27.90 (7.50)	9	95	20.60 (10.40)	97	22.10 (10.20)
Holländare [39]	Guided, Web-based CBT	Nonspecific support by an online therapist (email contact) + monthly rating of their depressive symptoms using the MADRS-S	42	17.00 (11.50)	42	17.70 (11.50)	10	38	9.30 (12.00)	39	13.40 (11.90)
Høifødt [40]	Guided, Web-based CBT (TAU)	WLC (TAU)	52	21.13 (6.85)	54	22.27 (6.74)	7	52	14.20 (8.15)	54	18.63 (8.64)
Johansson [41]	Web-based psychodynamic psychotherapy + online therapist contact	Web-based structured support intervention (psychoeducation and scheduled weekly contacts online)	46	26.54 (5.80)	46	26.33 (6.70)	10	42	11.48 (7.80)	46	20.22 (7.80)
Johansson [42]	Tailored Web-based CBT	Web-based discussion group with weekly discussion themes related to depression and the treatment of depression	36	26.44 (7.60)	42	26.24 (7.90)	10	36	13.78 (9.40)	39	21.67 (9.50)
Kay-Lambkin [43]	Computerized CBT therapy and motivational interviewing by a computer program	No further treatment	23	28.57 (9.89)	21	32.86 (9.59)	13	23	17.09 (12.14)	21	22.95 (10.46)
Kessler [44]	Web-based CBT + TAU	WLC (TAU)	149	32.80 (8.30)	148	33.50 (9.30)	17	113	14.50 (11.20)	97	22.00 (13.50)
Kivi [45]	Web-based CBT	TAU	45	25.50 (7.87)	47	26.09 (9.39)	13	30	13.23 (10.94)	35	14.46 (9.88)
Lappalainen [46]	Guided Web-based acceptance and commitment therapy without face-to-face contact	WLC	19	22.11 (7.79)	20	20.65 (6.80)	7	18	13.34 (6.75)	20	17.85 (7.34)

XSL•FO
RenderX

Study	Treatment group[a]	Control group[a]	Baseline				Follow-up			
			Treatment		Control		Wks	Treatment		Control	
			n	Mean (SD)	n	Mean (SD)		n	Mean (SD)	n	Mean (SD)
Lappalainen [47]	Guided Web-based acceptance and commitment therapy	Face-to-face Acceptance and Commitment therapy (ACT)	19	20.79 (9.34)	19	23.11 (6.38)	6	19	10.26 (8.20)	18	9.17 (5.24)
Ly [48]	Blended treatment (4 face-to-face sessions + a smartphone application used between sessions)	Full behavioral activation	46	28.96 (8.07)	47	27.32 (7.89)	9	44	15.17 (11.51)	46	13.43 (11.27)
Meyer [49]	Web-based CBT + TAU	WLC (TAU)	78	16.62 (3.44)	85	17.20 (3.86)	9	60	10.08 (6.37)	73	13.64 (6.14)
Meyer [50]	Web-based CBT + TAU	WLC (TAU)	320	26.72 (9.86)	76	27.11 (8.98)	9	159	19.87 (11.85)	57	27.15 (10.01)
Morgan [51]	Emails promoting the use of self-help strategies	Emails containing depression information	862	16.40 (5.98)	874	16.90 (5.76)	6	273	10.80 (6.84)	301	11.50 (6.72)
Moritz [52]	Web-based CBT	WLC	105	28.81 (11.11)	105	30.02 (10.18)	8	80	20.51 (12.22)	90	25.67 (11.65)
Perini [53]	Guided Web-based CBT	WLC	27	27.30 (7.30)	18	27.24 (6.18)	8	27	17.30 (9.86)	17	23.33 (9.29)
Phillips [54]	Computer-based CBT	Attention control (5 websites with general information about mental health)	311	14.60 (5.40)	318	14.60 (5.60)	6	164	9.90 (6.10)	176	10.20 (6.00)
Preschl [55]	Face-to-face life-review therapy including computer supplements	WLC	20	19.00 (6.60)	16	16.50 (5.60)	8	20	10.00 (6.30)	16	15.10 (7.80)
Richards [56]	Guided Web-based CBT	WLC	96	20.90 (3.83)	92	20.84 (4.17)	8	96	15.67 (7.68)	92	20.43 (6.97)
Richards [57]	Unguided Web-based CBT	Therapist-assisted email CBT	43	21.72 (5.30)	37	22.70 (4.70)	8	21	12.81 (6.90)	25	11.52 (4.90)
Ruwaard [58]	Guided Web-based CBT	WLC	36	19.70 (5.50)	18	21.30 (5.30)	11	36	9.80 (6.50)	18	15.60 (7.60)
Sheeber [59]	Guided Web-based CBT	WLC (TAU)	35	26.20 (9.80)	35	25.40 (9.00)	14	34	13.40 (10.40)	35	22.50 (11.00)
Spek [60]	Unguided Web-based CBT	Group face-to-face CBT	102	19.17 (7.21)	99	17.89 (9.95)	10	67	11.97 (8.05)	56	11.43 (9.41)
Ström [61]	Therapist-guided Web-based physical activity (guided self-help program)	WLC	24	26.92 (9.30)	24	28.25 (7.08)	9	24	17.88 (11.30)	24	24.04 (6.86)
Titov [62]	Guided Web-based CBT	WLC	27	11.04 (5.62)	25	12.04 (5.42)	8	23	3.96 (2.48)	22	12.68 (5.48)
Titov [63]	Clinician-guided Web-based CBT	WLC	41	27.15 (9.96)	40	26.33 (10.46)	8	41	15.29 (9.81) 7.59 (4.04)	40	26.15 (10.14)
Vernmark [64]	Web-based CBT, email supported	WLC	30	22.20 (5.30)	29	21.80 (6.60)	8	29	10.30 (5.20)	29	16.60 (7.90)
Wagner [65]	Guided Web-based CBT	Face-to-face CBT	32	22.96 (6.07)	30	23.41 (7.63)	8	25	12.41 (10.03)	28	12.33 (8.77)
Watts [66]	Smartphone-based CBT	Computer-based CBT	15	33.46 (2.95)	20	30.90 (2.55)	8	1010	12.53 (3.26)	15	13.68 (2.79)

XSL•FO
RenderX

Study	Treatment group[a]	Control group[a]	Baseline				Follow-up				
			Treatment		Control		Wks	Treatment		Control	
			n	Mean (SD)	n	Mean (SD)		n	Mean (SD)	n	Mean (SD)
Depression and anxiety											
Johansson [70]	Web-based psychodynamic, guided self-help treatment based on affect-phobia therapy (APT)	WLC	28	15.32 (3.30)	29	15.07 (4.40)	10	28	5.89 (2.80)	29	10.59 (6.40)
Mullin [71]	Web-based CBT (total sample: N=30); clinical subsample (PHQ-9 ≥10)	WLC	20	14.10 (3.62)	11	14.63 (3.35)	6	20	8.33 (4.86)	11	13.37 (7.42)
Newby [72]	Guided Web-based CBT (Worry and Sadness Program)	WLC	46	21.24 (6.98)	54	22.41 (9.17)	10	43	10.48 (8.30)	53	21.24 (10.56)
Proudfoot [73]	Computerized CBT	TAU	53	25.38 (11.05)	53	24.08 (9.78)	8	47	12.04 (10.45)	50	18.36 (12.65)
Titov [74]	Clinician-guided Web-based CBT	Self-guided Web-based CBT	112	15.07 (3.57)	105	15.23 (3.85)	8	112	7.36 (5.04)	105	8.44 (5.14)
Titov [75]	Transdiagnostic self-guided Web-based CBT with automated email	self-guided Web-based CBT without automated email	47	14.64 (3.34)	46	14.39 (3.33)	8	47	7.58 (4.60)	46	10.57 (6.16)
Titov [76]	Web-based CBT with email and phone support	WLC	18	14.39 (4.27)	20	13.35 (6.25)	10	18	7.67 (5.97)	20	12.15 (4.93)

[a] BDI: Beck Depression Index; CBT: cognitive behavioral therapy; PHQ: Patient Health Questionnaire; TAU: treatment as usual; WLC: waiting-list control.

Statistical Analyses

Each study was summarized in detail in the predeveloped data extraction table. The primary outcome of the BDI and PHQ was assessed as a continuous measure of effect in an additional table. Because moderate to substantial heterogeneity among the interventions was expected, mean effect sizes were calculated using a random-effects meta-analysis according to DerSimonian and Laird [69]. Review Manager 5 (RevMan 5) was used to conduct this systematic review [77]. In general, a *P* value of <.05 was considered statistically significant.

Calculation of Effect Sizes: Changes in Primary Outcome Measures Between Pre- and Posttreatment

For every technology-based intervention, to assess the within-group effect (uncontrolled effect size) of treatments, we calculated the standard mean difference (SMD) as effect size referring to the difference between baseline and postintervention, divided by the pooled standard deviation of each primary outcome measure, and the 95% confidence intervals around the effect sizes. According to the methodology described in Hedges [78], effect sizes were also adjusted to address small sample sizes. As a consequence of the interdependency of baseline and posttreatment values, the correlation between time points was required. However, because the majority of included studies did not provide the correlation between these values, a conservative value of .50 was used as suggested by Balk et al [79].

Calculation of Effect Sizes: Technology-Based Interventions Versus Control Conditions

We calculated the effect size (SMD or Hedges' *g* [78]) for each comparison between a technology-enabled intervention and a control condition. It indicates the difference between the two study groups at posttest (standardized mean difference) and the 95% confidence intervals around the effect sizes. Effect sizes resulted from the subtraction of the mean score of the intervention group at posttreatment from the mean score of the comparator group and dividing the result by the pooled standard deviation of the two groups. Values of 0.8 refer to large, 0.5 to moderate, and 0.2 to small effects [80]. To address small sample sizes, we also adjusted effect sizes according to the procedures described by Hedges [78]. Only those instruments that measured depressive symptoms were used in the calculation of effect sizes.

In cases in which more than one depression measure was provided, the BDI was preferred over the PHQ. If studies only assessed the PHQ score, the PHQ was used for calculations. In this analysis, only the effect sizes referring to the differences between the two study groups at posttreatment were used. Because the follow-up period varied considerably between studies, we decided not to examine the differential effects at these time points.

Assessment of Heterogeneity

As a consequence of the anticipated moderate-to-high level of diversity between study populations and interventions eligible for this systematic review, the Breslow-Day test was used to

XSL•FO
RenderX

test for heterogeneity [81]. To complement the common chi-square test for heterogeneity, the I^2 statistic proposed by Higgins et al [82] was used. Inconsistency (termed I^2) was calculated by the formula: I^2=max 0, 100%*(Q-df)/Q, where *Q* is the heterogeneity statistic and *df* its degrees of freedom. Because I^2 is not inherently dependent on the number of studies, this characteristic is of advantage in assessing the percentage of total variation across studies due to heterogeneity. An I^2 value greater than 50% was considered as strong inconsistency [83].

Subgroup and Sensitivity Analyses

Because a high degree of heterogeneity was to be expected, we tried to mitigate this issue by subgroup analyses. We tested prespecified hypotheses to assess the robustness of the findings and to explore sources of heterogeneity (relationships between study characteristics and intervention effects). The following hypotheses were proposed. It was assumed that the treatment effect was influenced by (1) duration of treatment, (2) severity of depression (baseline score), (3) age of participants, (4) methodological quality of studies, (5) type of control (eg, TAU, WLC), (6) inclusion of face-to-face therapist sessions, and (7) utilization of CBT techniques.

Assessment of Publication Bias

The data collection was based on the description of interventions in published literature. Thus, grey literature assessing the effectiveness of technology-based interventions was not taken into account. The potential presence of publication bias likely had a significant impact on the results of this study, not only with respect to differences in usage of online interventions in clinical settings, but also in more real-world settings [84]. To improve and standardize the description of technology-based interventions, it is suggested that future studies apply frameworks such as the Consolidated Standards of Reporting Trials (CONSORT) statement for eHealth [85], a protocol for systematic reviews [86] and guidelines for reporting online intervention research [73]. Language bias might be an issue in this review because only RCTs published in the English language were included. These limiting factors should be kept in mind when interpreting the findings of the current work. In order to identify cases of possible publication bias, a funnel plot was drawn for the main analysis [87]. Nonpublication of small trials would result in asymmetry of the plot. In addition, the funnel plot was evaluated for asymmetrical distributions. To confirm the visual interpretation, which can be subjective, the Begg and Mazumdar [88] adjusted rank correlation test for publication bias was used.

Results

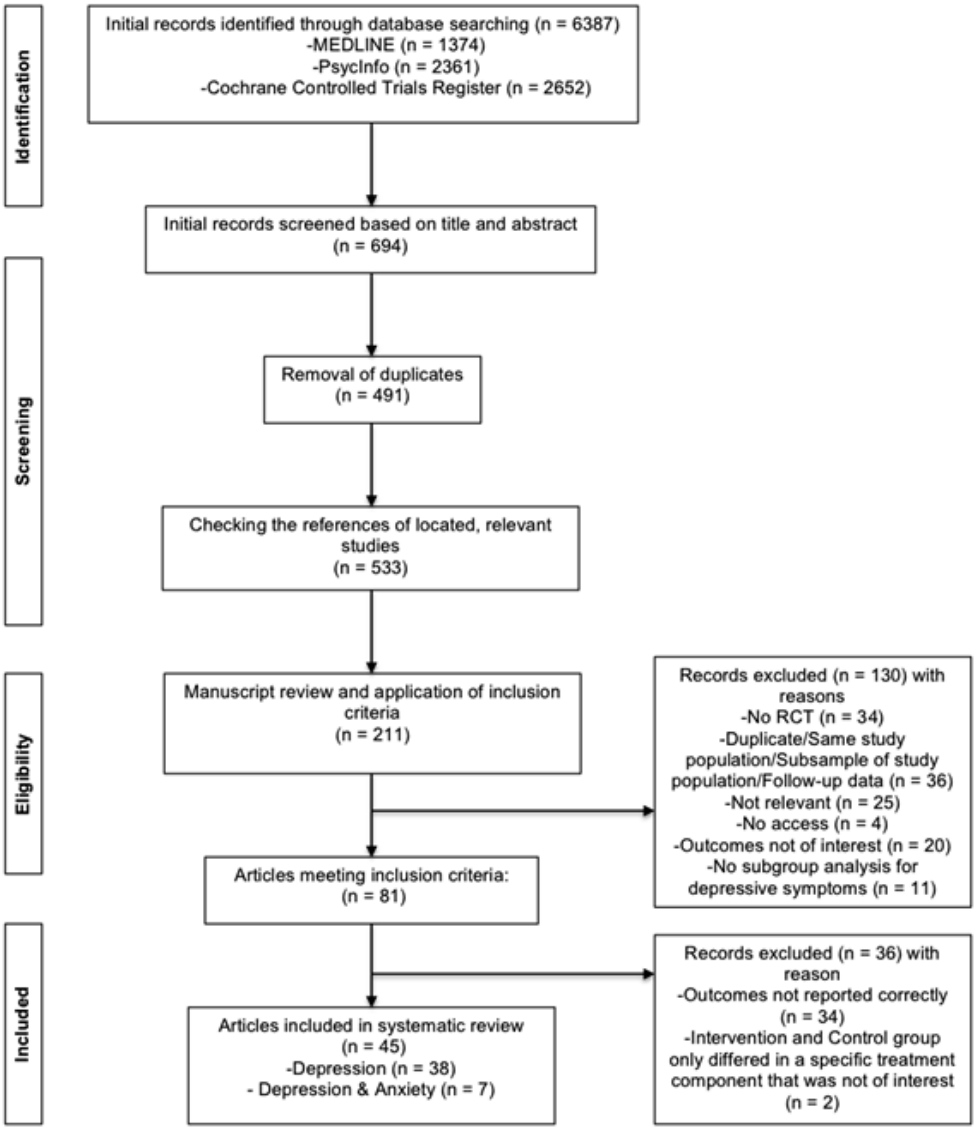
In this section, the findings of the different analyses that were carried out in this review are reported. Characteristics of studies are presented in tabular form.

Study Selection

The searches in MEDLINE, PsycInfo, and the Cochrane Controlled Trials Register identified a total of 6387 citations (articles and abstracts) published after 2000. After the adjustment for duplicates and the exclusion of noneligible trials based on titles and abstracts, 491 studies remained. Forty-two additional possibly eligible trials were identified by checking the reference lists of relevant articles already identified. A more detailed review of the full text of the remaining citations led to the detection and exclusion of 130 publications. Thirty-four trials were excluded because of the lack of appropriate reporting of outcomes. It was decided by consensus to exclude two additional studies that included an active control group that only differed from the study group in the use of a program component that was not relevant to this review. In total, 45 RCTs were included. Of these, seven trials analyzed patients with depressive symptoms comorbid with anxiety. Figure 2 summarizes the study selection process.

XSL•FO
RenderX

Figure 2. Flow diagram of study selection.



Characteristics of Studies and Risk of Bias Within Studies

Tables 3 and 4 provide information on the setting and outcomes of the 45 trials finally included in the analysis. These RCTs contributed a total of 7326 randomized and 4519 analyzable patients. The majority of trials studied CBT in adult patients with mild-to-moderate depressive symptoms.

The risk of bias at the level of the individual trials is addressed in Table 2 by reporting the modified Jadad score [19] and whether analysis was performed according to the intention-to-treat principle (see Assessment of Methodological Quality). None of the trials were described as "double-blind RCT" and we also noted a relatively high risk of bias due to the insufficient blinding of participants. Typically, the method of blinding was described insufficiently. This, however, is not uncommon for psychotherapy trials [67]. In general, achieving a successful blinding in psychotherapy trials is regarded to be more demanding than in pharmacological trials [89]. Study participants can easily identify the discrepancies in the contents of the treatment and control arms, and it is probably not very likely to successfully blind the participants or the therapists. Thus, the methods used to achieve proper blinding are rarely reported in psychotherapy trials [67]. The risk of bias introduced by selective reporting was small because all outcomes of interest were adequately described in the vast majority of the included RCTs.

System Components

In total, a set of 15 system components was identified based on occurrence in reviewed literature. These were either defined, hypotheses-driven, or derived from Oinas-Kukkonnen's model of persuasive systems design [18]. For each study included in the systematic review, we determined the presence of defined system components for later analysis. In the following, we present an overview of identified components, together with the underlying inclusion rationale.

XSL•FO
RenderX

Channel of Delivery

Technology-mediated MHIS can be administered using a range of available technologies. Although early interventions were based on offline programs, computerized programs and Internet-delivered Web interventions have become more popular in recent years [16]. Following the latest development, the mobile phone as a channel of delivery is getting more and more attention. This seems reasonable because almost half of the world's population has a mobile phone subscription, and it is expected that by 2020 the global penetration rate will reach approximately 60% [90]. In addition, it is suggested that differences in access to mobile technologies are diminishing at least for nonrural populations, thus offering an opportunity to reach underserved and marginalized populations [91]. The high global penetration and the rapid growth of mobile phone apps provide the opportunity to reach an increasing number of people who are in need of treatment for mental disorders [92].

Tailoring (Personalization)

Oinas-Kukkonen argued that "information provided by the system will be more persuasive if it is tailored to the potential needs, interests, personality, usage context, or other factors relevant to a user group" and that "a system that offers personalized content or services has a greater capability for persuasion" [18]. This was confirmed in the context of health behavior change by a systematic review [93] and presents a promising component for the treatment of depression in MHIS.

Supportive Text Messages (Tunneling/Praise)

Research by Agyapong et al [30] targeting the support for people with depression and comorbid alcohol use disorder presented promising results in the deployment of supportive text messages. This followed the Oinas-Kukkonen concept of tunneling [18] (ie, "using the system to guide users through a process or experience provides opportunities to persuade along the way") and the concept of praise, which is said to make users more open to persuasion [17].

Peer Support

Although no consensus with respect to effectiveness of online peer support was reached yet [94,95], anonymous online support groups and discussion forums might help users to overcome the feeling of being stigmatized by connecting patients with others. A further advantage of these social support components is that time and location are no longer obstacles for active participation [96]. Peer support follows the Oinas-Kukkonens [18] concept of social learning, social comparison, and social facilitation [96].

Case-Enhanced Learning

This form of learning uses educational stories that identify a problem and a solution with an example (ie, a case) the participant can potentially identify with [75]. These can be implemented via, for example, video vignettes of case-study patients.

Reminders

Stemming from Oinas-Kikkonens concept of reminders, Whitton et al [11] found that reminders play a decisive role in the engagement of users in mental health interventions and are a cost-effective approach for engaging users [11-13]. Furthermore, it is suggested that reminders not only enhance user engagement but also improve adherence [97-99] and counteract the high rates of nonusage attrition common to many online-based interventions [97,100].

Downloadable Material

Because a participant's preferred medium for reading might be paper [101], an option to download and print out summaries, lessons, or homework might influence treatment efficacy by providing a higher level of comfort.

Workbook/Homework Assignments

Homework assignments, as commonly used in standard care [102], are an important construct in CBT. In a recent study, LeBeau et al [103] concluded that "improvement of homework compliance has the potential to be a highly practical and effective way to improve clinical outcomes in CBT." Therefore, implemented in an appealing interactive way, this might represent an important component in MHIS.

Symptom Tracking

Tracking symptoms, either objectively using sensors [104] or by means of self-reports might be beneficial for the user following Oinas-Kikkonens self-monitoring concept [18]. It describes a "system that helps track one's own performance or status supports in achieving goals" [18].

Online Diary

Diaries form a way of self-monitoring and self-reflection and are frequently used in classical forms of CBT [105].

Summaries

We assessed whether included studies made use of summaries of content (eg, module or progress summaries), which represent another dimension drawing on the concept of self-monitoring and self-reflection [17].

Audio/Voiceover

A recent experimental study found that audience feedback is a valuable tool to enhance users' perceptions of health-related YouTube clips [106], which highlights the power of the participatory nature of the Web to increase the efficacy of Internet-based health interventions.

Illustrative Content/Video

Illustrative content in the form of graphics, photos, illustrations, comics, or video clips might increase the appeal of interactivity and visual attractiveness of Internet-based programs [107].

Gamification

The use of game-like strategies has demonstrated to produce positive outcomes in previous studies of technology-based health interventions [108,109]. Gamification of health apps might provide the option to set up goals and rules for personal health behavior and to track patient's actual behavior against these rules and goals [108]. The utilization of rewards in the context of health intervention might be promising because it was found

XSL•FO
RenderX

that rewards are able to stimulate positive thinking in users and are thus a powerful tool to drive long-term participation [110].

Animations/Virtual Assistant

Virtual agents or avatars could be used for persuasive purposes and to support self-management among patients [111]. A growing body of literature examines the relationship of virtual agents and their user, potentially holding vast opportunities for persuasive system design [112-114].

System Component Setting

Table 5 contains detailed information on the settings and system component configuration of the interventions.

Approximately 45% (20/45) of all included studies reported using email reminders and 10 of 45 studies (22%) reported providing SMS text message reminders, mostly as an alternative to reminders sent by email. In the included trials, reminders were typically intended to increase motivation and adherence to therapeutic interventions. As explained earlier, reminders play a decisive role in the engagement of users in mental health interventions. Most of the RCTs made use of the Internet for delivering mental health interventions for depression or depression comorbid with anxiety. The majority of RCTs included in this systematic review (91%, 41/45) did not make use of mobile phones or tablets. Typically, the interventions required interaction with the system, and many also included interaction with a therapist (face-to-face or online) and/or peers on the Web. In all, 80% (36/45) of included programs were based on CBT or used elements of CBT. Tunneling, which refers to the stepwise delivery of content, is typically found in technology-based interventions for depression [18], and was also used in the majority of studies (87%, 39/45) included in this work. Twenty-seven of 45 included studies used tailored content, tailored feedback, and/or tailored reminders.

Only a small number of included studies made use of self-monitoring components, such as symptom tracking and tracking reminders, yet they are seen as key features of psychotherapy in particular [40,96], and in behavior-change interventions in general [12,115-118].

Although social support is widely recognized as an important feature in behavior change [119,120], in this analysis, only seven studies (16%) used peer support. In the included studies, social support consisted of the use of Web-based discussion boards (asynchronous social support), which was intended to provide the ability to connect with other patients of the same intervention.

Regarding visual attractiveness, 44% of all included studies analyzed an intervention that included visually appealing content, such as graphs, illustrations, comics, or photos, which often serve a motivational purpose. Only 11 of 45 studies (24%) described the utilization of audio and/or voiceovers. Video footage, often containing case-enhanced learning, was also found in 11 RCTs (24%).

In addition, this analysis showed that game elements were only used in three RCTs and, if gaming was included, it was in the form of knowledge quizzes. See Figure 3 for a quantitative overview of components among included studies.

XSL•FO
RenderX

Table 5. MHIS system component configuration of included studies for health information system (HIS) channel and tailoring and (1) email/phone reminders, (2) supportive text messages, (3) peer support, (4) summaries of progress or content, (5) case-enhanced learning, (6) material to download/print, (7) homework assignments, (8) mood rating / symptom tracking, (9) online diary/journal, (10) audio/voiceovers, (11) illustrative content, (12) games/quizzes, and (13) animations (virtual agent).[a]

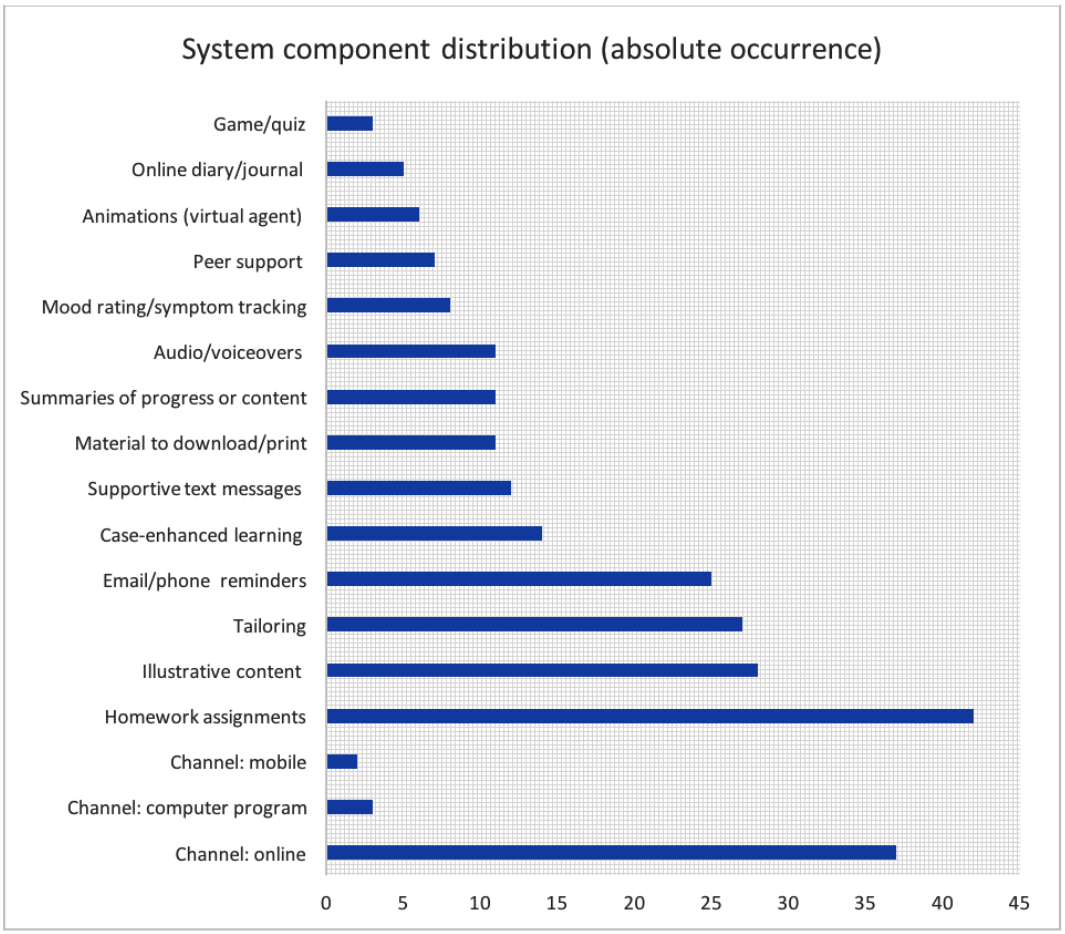
Study	HIS channel	Tailoring	1	2	3	4	5	6	7	8	9	10	11	12	13
Depression															
Agyapong [30]	Mobile phone	0/NR	1	1	0	0	0	0	1	0	0	0	0	0	0
Andersson [31]	Online	0/NR	0	0	0	0	0	0	1	0	0	0	0	0	0
Andersson [32]	Online	Partly	1	0	1	0	0	1	1	0	0	0	0	1	0
Berger [33]	Online	1	1	0	0	0	0	1	1	0	0	0	1	0	0
Burton [34]	Online	1	0	0	0	1	0	0	0	1	0	1	0	0	1
Carlbring [35]	Email	1	0	0	0	0	0	0	1	0	0	1	1	0	1
Choi [36]	Online	0/NR	1	0	0	1	1	0	1	0	0	0	1	0	0
Clarke [37]	Online	1	1	0	1	0	0	0	1	1	1	0	1	0	0
de Graaf [38]	Online	0/NR	0	1	0	0	0	0	1	1	1	0	1	0	0
Holländare [39]	Online	0/NR	1	0	0	0	0	0	1	1	0	0	1	0	0
Høifødt [40]	Online	Feedback/reminders	1	0	0	0	0	0	1	0	0	0	1	0	0
Johansson [41]	Online	Feedback/reminders	1	1	0	0	0	0	1	0	0	0	0	0	0
Johansson [42]	Online	1	1	1	1	0	0	1	1	0	0	0	0	0	0
Kay-Lambkin [43]	Computer program	0/NR	0	0	0	0	0	0	1	0	0	1	1	0	0
Kessler [44]	Online	1	0	1	0	0	0	0	0	0	0	0	0	0	0
Kivi [45]	Online	Feedback/reminders	0	0	0	0	0	1	1	0	1	1	1	0	0
Lappalainen [46]	Online	Feedback/reminders	1	0	0	0	0	0	1	0	1	1	1	0	0
Lappalainen [47]	Online	1	1	0	0	0	0	0	1	1	0	1	0	0	0
Ly [48]	Mobile phone	1	1	1	0	0	1	0	1	0	0	0	1	0	0
Meyer [49]	Online	1	1	1	0	0	0	1	1	1	0	1	1	0	0
Meyer [50]	Online	1	0	0	0	0	1	0	1	0	0	0	1	0	1
Morgan [51]	Email	0/NR	1	1	0	0	0	0	0	0	0	0	0	0	0
Moritz [52]	Online	1	0	0	0	0	1	0	1	0	0	0	1	0	1
Perini [53]	Online	0/NR	0	0	1	1	1	1	1	0	0	0	1	0	0
Phillips [54]	Online	Feedback/reminders	1	0	0	0	0	0	1	0	0	0	1	0	0
Preschl [55]	Computer program	1	0	0	0	0	0	1	1	0	0	1	1	0	0
Richards [56]	Online	1	1	0	0	1	1	0	1	0	0	0	1	1	0
Richards [57]	Online	Feedback /reminders	0	0	0	0	0	0	1	0	1	1	1	0	1
Ruwaard [58]	Online	1	0	1	0	0	0	0	1	1	0	0	0	0	0
Sheeber [59]	Online	Feedback/reminders	0	0	1	0	0	0	1	1	0	1	0	0	0
Spek [60]	Online	0/NR	0	0	0	0	0	0	1	0	0	0	1	0	0
Ström [61]	Online	1	0	1	0	0	0	0	1	0	0	0	0	0	0
Titov [62]	Online	0/NR	1	0	0	1	1	0	1	0	0	0	0	0	0
Titov [63]	Online	0/NR	1	0	1	1	1	1	1	0	0	0	1	0	0

XSL•FO
RenderX

Study	HIS channel	Tailoring	1	2	3	4	5	6	7	8	9	10	11	12	13
Vernmark [64]	Online	1	1	0	0	0	0	1	1	0	0	0	0	1	0
Wagner [65]	Online	Feedback/reminders	0	0	0	0	0	0	1	0	0	0	0	0	0
Watts [66]	Mobile phone vs PC	0/NR	0	0	0	0	1	0	1	0	0	0	1	0	0
Agyapong [30]	Online	0/NR	1	0	0	0	0	0	1	0	0	0	0	0	0
Depression and anxiety															
Johansson [70]	Online	Feedback/reminders	0	0	0	0	0	0	1	0	0	0	1	0	0
Mullin [71]	Online	0/NR	1	0	0	1	1	0	1	0	0	0	1	0	0
Newby [72]	Online	0/NR	1	0	0	1	1	1	1	0	0	0	0	0	0
Proudfoot [73]	Computer program	1	NR	0	0	1	0	1	1	0	0	1	1	0	1
Titov [74]	Online	0/NR	1	1	0	1	1	0	1	0	0	0	1	0	0
Titov [75]	Online	0/NR	1	0	0	0	1	0	1	0	0	0	1	0	0
Titov [76]	Online	0/NR	1	1	1	1	1	0	1	0	0	0	1	0	0

[a] Rating system is 1=component present and 0=component not present. CG: control group; NR: not reported; TG: treatment group

Figure 3. System component distribution of included studies (N=45).

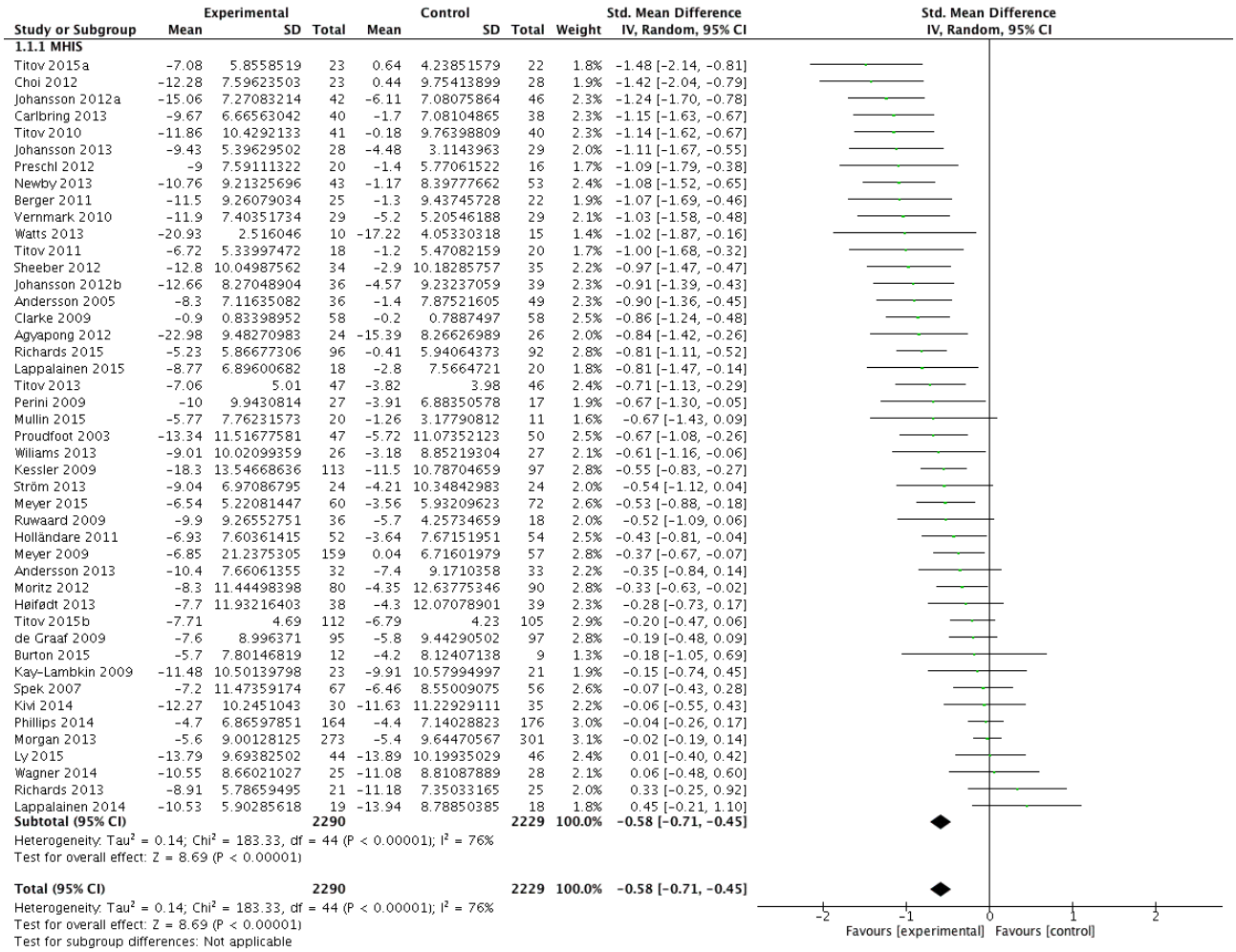


XSL•FO
RenderX

Synthesis of Results: Impact on Symptoms of Depression

Data from 45 trials (4519 patients) that reported on BDI or PHQ scores before and after the treatment were combined to estimate the overall effect of technology-based interventions on depressive symptomatology. Technology-supported treatments for depression showed a trend toward reduced depressive symptoms (SMD=–0.58, 95% CI –0.71 to –0.45; *P*<.001) (Figure 4). The chi-square test for heterogeneity (χ^2_{44}=183.3), across all patient types, was significant (*P*<.001) and total variation across trials due to heterogeneity (I^2) was 76%, a strong indication for inconsistency across included RCTs.

Figure 4. Effect of technology-based interventions on symptoms of depression in included studies (N=45).



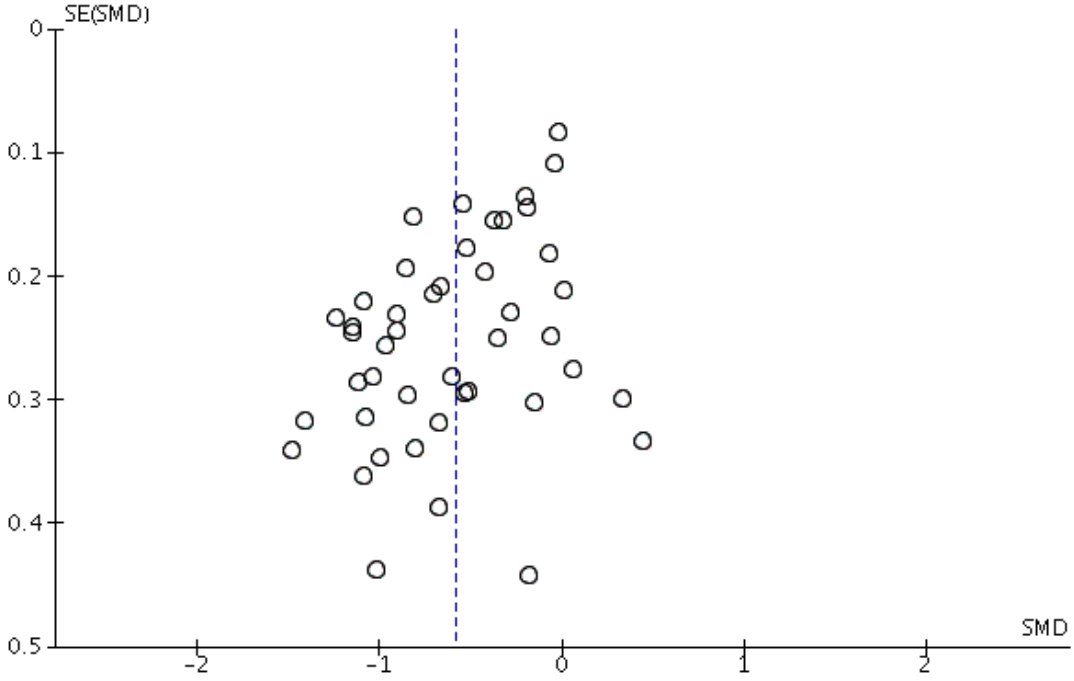
Risk of Bias Across Studies

The funnel plot drawn for the main analysis on the effectiveness of technology-delivered interventions for depression showed an asymmetry, which is evidence of missing studies suggesting publication bias (Figure 5). Additional testing utilizing the Begg and Mazumdar rank correlation method [88] confirmed that there is, in fact, evidence of publication bias (*P*=.03).

As a consequence of this finding, we also explored publication bias in the 22 RCTs that used a WLC group. The funnel plot of this subgroup of included trials showed no evidence of bias, which was confirmed by the Begg test (*P*=.08).

XSL•FO
RenderX

Figure 5. Risk of bias across all included studies (N=45).



Results of Subgroup Analyses

Subgroup analyses by study quality, treatment duration, provision of face-to-face contact, utilization of CBT techniques, severity of baseline depression, age of participants, and type of control (eg, TAU or WLC) were prespecified and served to test our hypotheses (see Subgroup and Sensitivity Analyses). The individual forest plots per subgroup analysis can be found in Multimedia Appendixes 1-7.

Study Quality

Differences in study findings could also be explained by biased results due to differences in quality of individual studies. Thus, a subgroup analysis based on the methodological quality of included trials was performed. It could be shown that the effect of technology-based interventions on symptoms of depression is consistent in trials of higher (quality score >3.5) and lower quality (quality score ≤3.5). In high-quality trials, technology-based interventions were associated with an SMD of –0.60 (95% CI –0.76 to –0.45, P<.001; I^2=77%) and a SMD of –0.53 (95% CI –0.77 to –0.29, P<.001; I^2=76%) in low-quality studies. It seems that methodological quality of trials only showed a small impact on effect sizes, but overall had a moderate effect (χ^2_1=0.3, P=.60; I^2=0%).

Treatment Duration

As statistical heterogeneity was found and to exclude the possibility of heterogeneity due to differences in the duration of the interventions, trials of different durations were compared to one another. Regarding depressive symptomatology, technology-based interventions showed to be effective, irrespective of treatment duration. In treatments with a duration of 10 weeks or less, the intervention was associated with an SMD of –0.60 (95% CI –0.76 to –0.44, P<.001; I^2=79%). Similarly, treatments with a duration of longer than 10 weeks resulted in an SMD of –0.52 (95% CI –0.70 to –0.33, P<.001; I^2=48%). There was no evidence for an association between duration of treatment and the effect of treatments on depressive symptomatology (χ^2_1=0.4, P=.50; I^2=0%).

Provision of Face-to-Face Contact

As statistical heterogeneity was found in our analysis of treatment effectiveness and in order to exclude the possibility of heterogeneity due to the provision of face-to-face contact, trials that incorporated face-to-face support were compared to interventions that did not use live contact with a therapist. We hypothesized interventions that included face-to-face sessions would show greater reductions in depressive symptoms than treatments that did not. Contrary to what was expected, effect size in treatments that did not use face-to-face support was larger (SMD –0.65, 95% CI –0.79 to –0.50, P<.001; I^2=78%) than in interventions that offered live support (SMD –0.28, 95% CI –0.52 to –0.03, P=.03; I^2=50%). An association between the effectiveness of treatments and the provision of face-to-face contact was found (χ^2_1=6.34, P=.01; I^2=84%).

Utilization of Cognitive Behavioral Therapy

We made the hypothesis that technology-based interventions using CBT techniques are more effective than treatments that do not use components of CBT, reflecting on the predominance of CBT in the literature. Effect size varied only slightly between trials that were based on CBT (SMD –0.58, 95%CI –0.72 to –0.45, P<.001; I^2=71%) and interventions that did not use a CBT protocol (SMD –0.56, 95% CI –0.96 to –0.16, P=.006; I^2=85%). Also, according to the test for subgroup differences,

XSL•FO
RenderX

there was no association between the effect of technology-supported treatments on depressive symptoms and utilization of CBT methods (χ^2_1=0, *P*=.90; I^2=0%).

Severity of Baseline Depression

Comparing trials in which patients showed higher depression scores (BDI ≥25 or PHQ ≥15) with trials in which patients had lower scores (BDI <25 or PHQ <15), technology-based treatment showed to be effective in both groups (higher severity: SMD –0.61, 95% CI –0.79 to –0.44, *P*<.001; I^2=77%; lower severity: SMD –0.54, 95% CI 0.75 to –0.33, *P*<.001; I^2=75%). The test for subgroup differences resulted in a chi-square value of 0.3 (df=1, *P*=.59; I^2=0%).

Age of Participants

Age did not show a significant impact on effectiveness of technology-based interventions for the treatment of depression. Although effect sizes varied between adults (SMD –0.63, 95% CI –0.79 to –0.46, *P*<.001; I^2=79%), adults excluding older adults (SMD–0.48, 95% CI –0.69 to –0.27, *P*<.001; I^2=55%), older adults (SMD –0.53, 95% CI –1.52 to 0.46, *P*=.29; I^2=84%), and adolescents (SMD –0.28, 95% CI –1.45 to 0.88, *P*=.63; I^2=91%), subgroup analyses pointed in the same direction (see Table 1 for a description of the different age categories). The effect of technology-based interventions on symptoms of depression was not associated with patient age (χ^2_3=1.4, *P*=.71; I^2=0%).

Type of Control Condition

Comparing technology-based treatments to TAU resulted in a moderate effect (SMD –0.48, 95% CI –0.78 to –0.18, *P*=.002; I^2=62%) and the comparison of technology-supported interventions for depression to WLC showed a large effect (SMD –0.79, 95% CI –0.93 to –0.64, *P*<.001; I^2=51%). The test for subgroup differences resulted in a chi-square value of 3.3 (df=1, *P*=.07; I^2=69.3%).

Discussion

This systematic literature review had the following objectives: (1) to collect all relevant clinical studies of technology-based interventions that analyzed the effectiveness for the treatment of depression in order to accurately depict the body of literature and (2) to identify a set of system components of technology- and Internet-based interventions for depression. In general, the results are in line with previous analyses and showed that technology-supported interventions, in fact, reduce depressive symptoms [5,6,121]. This study is one of the first that provides an overview of technical components used in the current set of RCT trials that made use of computerized and online interventions on the treatment of depression.

Principal Results

Forty-five publications with a total number of 7326 randomized and 4519 analyzable participants were included in this systematic review, and most of the interventions analyzed were able to reduce symptoms of depression. The majority of included studies were of fair (60%) to good (33%) quality, and almost every study included analyzed an intervention that was modular in setup and typically lasted for approximately 10 weeks. Thirty-six studies (80%) deployed a CBT approach. This extends the systematic review by Saddichha et al [8], which consisted of 29 RCT studies using CBT. Usually, the programs were aimed to be used about once a week. This is in line with traditional CBT, which is seen as a step-by-step, short-term treatment with weekly or biweekly therapist sessions for 10 to 20 weeks [17].

Subgroup analysis showed that study quality, treatment duration, provision of face-to-face contact, utilization of CBT, and age of participants had relatively small impact on the outcome of the interventions. For study quality, it is plausible because the intervention quality is not inevitably reflected by the study quality. Likewise, there are reasonable explanations that the age of the participants and treatment durations did not have a decisive impact on treatment outcome. We probably overestimated the influence of technology literacy in older participants because our results confirm findings of literature showing comparable treatment results for all age groups. The small difference in effect for treatment durations can also be explained taking into account that traditional therapy has a similar range of time spans to deliver the same amount of structured information and is chosen depending on, for example, severity of illness and level of support [122]. We were surprised that provision of face-to-face therapy did not show significant effects on the treatment outcome. In fact, this confirmed a recent study comparing traditional therapy with a computerized intervention [65].

Interestingly, the use of CBT components also did not show significant effects on treatment outcome compared to interventions that were not based on CBT. Although there exists evidence that other types of intervention can be equally effective, predominance of CBT in traditional therapy as well as in technology-mediated MHIS urged us to test its superiority in our analysis.

As explained earlier, reminders play a decisive role in the engagement of users in mental health interventions. In this systematic review, approximately 45% of all included studies reported to use email reminders and 10 studies (22%) reported providing phone reminders, mostly as an alternative to reminders sent by email. In the included trials, reminders were typically intended to increase motivation and adherence to therapeutic interventions.

With respect to the design and effective use of reminders, research indicates that a variety of factors show an impact on the efficiency of these alerts. Firstly, it was found that there is a high risk that motivational emails provided within a workplace setting are easily ignored as a consequence of a full email inbox [123,124]. Thus, it might be of importance to offer the possibility to choose between mobile phone or email reminders and to customize time points at which users will be reminded to complete program modules. Secondly, regarding the content of expert-initiated contact, it is postulated that contacts delivering behavior-change techniques might be of greater effectiveness than simple messages that prompt users to access

XSL•FO
RenderX

the intervention [10]. Thirdly, it is suggested that reminders containing short motivational messages, quotes, or facts might counteract negative feedback cycles that maintain perceptions of low self-worth and associated depressive symptomatology [11,125]. Lastly, it could also be shown that the personalization of reminders and sending them out frequently enhances the effectiveness of treatment [98]. Nevertheless, nonspecific factors, such as encouragement, empathy, and hopefulness of improvement, may also independently enhance therapeutic gains. To conclude, it seems evident that creating a sense of being continuously supported and encouraged is crucial for user engagement, treatment adherence, and buffering against the development of negative feedback cycles. The regular and consistent receipt of well-designed reminders, motivational messages, and tips may thus be a very powerful means of reminding patients that they are actively working on gaining control over their symptoms [11].

Most of the RCTs included made use of the Internet for delivering mental health interventions for depression or depression comorbid with anxiety. The majority of RCTs included in this systematic review (93%, 42/45) did not make use of mobile phones or tablets, which is surprising because the benefits with respect to user engagement and adherence seem apparent. Typically, the interventions required the interaction with the system and many also included the interaction with a therapist (face-to-face or online) and/or peers on the Web. In all, 80% (36/45) of included programs were based on CBT or used elements of CBT. Given the fact that therapeutic interventions for depression are commonly based on CBT techniques and psychoeducation, which follow a stepwise approach and are usually delivered in person by a therapist, these findings support the authors' premise. Twenty-seven of 45 included studies used tailored content, tailored feedback, and/or tailored reminders. In the opinion of researchers, the adaptation of information to factors that are relevant to one individual or a group of individuals is an important feature in effective health communication [93,126,127]. In fact, van Genugten et al [128] found that interventions that are more flexible in use (easy to handle for both advanced as well as novice users), provide structure (which is comprehensible to the user and that the user knows at what point he or she is in the process), and use default settings are more likely to be effective. With respect to treatment exposure, it is suggested that personally tailored feedback and goal setting are among the important factors related to the use and exposure to Web-based interventions. In the study by Brouwer et al [129], exposure was regarded as the time spent on the website, page views, and the number of times the user logged on. Although there is a relationship between exposure, adherence, and treatment outcomes, focusing on exposure only gives a limited insight into the pattern of usage and adherence [97].

Doherty et al [130] showed that user engagement also depends on depression severity and that users with minimal symptoms engage much less than other groups. To serve the needs of this patient group, users might benefit from more flexibility (eg, in the elements of the intervention they wish to focus on). Therefore, new approaches such as tailoring of interventions that are more lightweight are needed. However, due to the severity of the disorder, patients with more pronounced symptoms face specific difficulties in engaging with the program. Because these patients require more intensive support than patients with less severe symptoms, they should also be given the choice to change the means of support throughout the treatment. Thus, mechanisms are needed that allow for requesting or being offered face-to-face contact even if the user has initially commenced online treatment. We assume that customization of programs is necessary to enhance long-term adherence and outcomes of interventions.

Only a small number of included studies made use of self-monitoring components, such as symptom tracking and tracking reminders, yet they are seen as key features of psychotherapy [11,96] and in behavior-change interventions in general [99,115-118].

Although social support is widely recognized as an important feature in behavior change [119,120], in this analysis only seven studies (16%) used peer support. In the included studies, social support consisted of the use of Web-based discussion boards (asynchronous social support), which intended to provide the ability to connect with other patients using the same intervention. However, the literature shows that there is disagreement with regard to use and benefit of discussion forums and chat rooms. Although many studies support the use of these components [123,131,132], others do not [133-135]. The effectiveness of peer support depends on individual factors, such as perception of the credibility of Internet-based peer advice and perceived quality of interaction [10]. Furthermore, effectiveness might also rely on user involvement. It could be shown that users that actively post and respond to messages are more likely to benefit than users that participate only passively [132].

Regarding visual attractiveness, 44% of all included studies analyzed an intervention that included visually appealing content, such as graphs, illustrations, comics, or photos, which often serve a motivational purpose. Only 11 studies (24%) described the utilization of audio and/or voiceovers. Video footage, often containing case-enhanced learning, was also found in 11 RCTs (24%).

In addition, this analysis showed that game elements were only used in three RCTs and if gaming was included, it was in the form of knowledge quizzes. Relatively few studies have incorporated games as part of their persuasive design.

Although virtual reality has shown to be effective in the treatment of anxiety and pediatric disorders [136], so far there is no study utilizing this technology for the treatment of depression and this might open a promising direction [137]. With respect to virtual agents and synthesized speech, Morrison et al [10] were also not able to show an association between digitized speech and improved outcomes in depression. To date, technological advances and improved design of animations and avatar-based systems are likely to permit the development of sufficiently sophisticated services to simulate real interaction [10] and, therefore, might show more impact on treatment outcomes. Future research is needed that concerns the effectiveness of serious gaming and virtual reality in

XSL•FO
RenderX

technology-supported mental health interventions currently underrepresented in literature.

Limitations and Future Directions

The list of factors that influence user friendliness as well as the different platforms for delivery included in this analysis is not exhaustive. In fact, the majority of RCTs included in this systematic review did not make use of mobile phones or tablets. It is expected that especially newer studies could use different channels of service delivery (eg, mobile phone or tablet delivery). Consequently, studying future interventions that make use of these delivery channels would give an interesting insight into the influence of different modes of delivery on treatment effectiveness.

Further limitations are related to the strict process of study selection applied in this systematic review. Many trials were excluded because (1) they were not described as being randomized, (2) participants showed no symptoms of depression at baseline, (3) they included other mental health disorders, and (4) they did not assess one of the outcomes of interest. In fact, the decision to only include RCTs might lead to potential limitations of this systematic review. Even though RCTs are regarded as the "gold standard" of reliable evidence, the criteria to only include RCTs might lead to the exclusion of relevant articles that examined the effectiveness of MHIS, but used a different study design. Primarily, the exclusion of non-RCTs in this review lead to a facilitated analysis of studies because differences in methodological quality are, although not completely removed, limited. A possible consequence of this decision might be that we missed studies of newer interventions, which might not yet be evaluated in an RCT study format because they are undergoing their piloting phase at the current time [15,104].

In addition, as a consequence of considering a wide range of MHIS, included trials differed considerably in the type of therapeutic programs they used, baseline depression severity, age of participants, duration of treatment, type of control condition, methodological quality, and the various system components they utilized to enhance user engagement, motivation, and effectiveness of the intervention. As a consequence of this moderate-to-high level of heterogeneity across included trials, comparability is restricted and results should be handled with care. Subgroup analyses demonstrated that there is, in fact, a significant association between the effectiveness of interventions and the provision of face-to-face contact as well as the type of control they used. As previously noted, the inclusion criteria stated in the protocol also included other mental disorders such as anxiety disorders. However, the vast number of articles identified in the electronic search posed an additional challenge. Consequently, it was decided to focus on depression and depression comorbid with anxiety only. Considering that many mental health problems often co-occur [138], study findings might be constrained as a result of this relatively strict inclusion/exclusion of certain mental disorders.

Because technology-based psychological interventions adapt established methods of treatment and only the means of delivery are altered, the issue of noninferiority plays a major role in this review. With respect to the overall effectiveness of technology-based interventions, it is of utmost importance to review the literature from the perspective of noninferiority trials that compare an established evidence-based treatment (eg, CBT) with a new one (eg, technology-based CBT). It also needs to be clarified that the absence of a significant difference between two interventions in an RCT cannot be equated with noninferiority and that the comparison of treatment effects between studies are only appropriate if the new and existing treatments are compared against a reference that does not substantially differ in methods and population [2].

In addition, identifying the points of disengagement and gaining deeper insight into the patterns of program usage is crucial for the refinement of system components that are most strongly associated with user engagement and symptom improvement. Data on patterns of use further offer an opportunity to refine content, means of delivery and to adapt both to the needs and preferences of specific groups of users [130]. Additional research is needed to overcome these shortcomings by assessing the association of patterns of component use and the improvement of symptoms by means of advanced statistical analyses. Furthermore, it is suggested that RCTs assess symptom reduction more frequently to obtain improved information about nonlinear relationships between patterns of usage and therapeutic gains. To enhance the understanding of when and how to choose and use different system components, it also is important to comprehend the dose-effect relationship of different components [128]. Future studies should be aimed to clarify the causal relationship between patterns of program usage and symptom improvement by assigning participants to different system components. In addition, it is likely that other predictors such as age, gender, and education affect the relationship between usage of system components, engagement, and outcomes. Therefore, additional studies are needed to assess whether differences exist between those populations.

In regard to designing mental health interventions, it is of utmost importance to understand what users need and expect from computerized or Web-based mental health interventions and how individuals rate different system components with respect to usefulness, practicality, connectivity, time demands, professional support, social interaction, convenience, novelty, reliability, confidentiality, trustworthiness, motivation, and engagement. Qualitative feedback offers a solution to find answers to the proposed questions. Likewise, user feedback might disclose disparities between user expectations and actual results. It can be assumed that user preferences vary greatly from individual to individual. This, in turn, supports the rationale of customization and tailoring of programs to create unique user experiences based on client's preferences without losing the effectiveness of interventions.

Moreover, to clarify the clinical feasibility of computer- and Web-supported mental health interventions, it also is important and worthwhile to repeatedly listen to the opinion of therapists. Apart from the time and cost savings, there is a need for a thorough understanding of the program, which could be achieved by the provision of a protocol, printed manual, and an overview of the program to the therapists involved. In addition, more detailed information on their practices and how to deal with

XSL•FO
RenderX

clients who are not engaging with the program should be provided [130].

Although clinicians tend to be very self-protective about their time commitments and skeptical about technology [139], value in health care should always be defined around the customer. Nonetheless, because value in health care is measured by the outcomes achieved, value depends on results not input [140]. First, computer-supported interventions should be designed and modified to optimize clients' benefits. Although value in health care is measured by the outcomes achieved, only assessing the effect on symptom reduction fails to include the value of mediators that might explain behavior change (eg, factors such as skills, attitudes, and self-efficacy) [128]. Thus, a further shortcoming of the current meta-analysis is the lack of analyses of potential mediators of the effect.

Implications and Recommendations

From a system component perspective, there is a strong need to counteract the decrease of program usage over the course of the intervention that is typically found in unguided technology-mediated interventions for mental health [118,141,142]. Thus, it is of great importance to implement the program components that are associated with improved and regular program engagement. One of the factors that have been linked to a decrease in module completion rate is obliging users to complete module sessions in a predetermined sequence [143]. On the one hand, the delivery of module sessions in a tunneled format bears many advantages, such as a greater number of website pages accessed, greater time spent on the website, and greater knowledge gained from the website [143]. On the other hand, tunneling might lead to a decrease in module completion because users are required to complete the between-session homework assignments before being able to start the next module session [11]. Therefore, developers have to enable a high level of flexibility in the choice of relevant modules as well as the speed with which users proceed through the modules to keep users engaged. One approach to allow for a greater flexibility, if not violating the underlying psychological theory, is to make homework assignments optional [11] or to leave it to the user at what stage they prefer to complete the homework tasks. Future studies are needed to confirm whether the provision of more user-driven programs might offer an advantage over the frequently used linear delivery of interventions in mental health interventions [139].

In addition, little is known about the synergistic effects of behavior-change components, modes of delivery, and user friendliness. Van Genugten et al [128] found synergistic effects in interventions that made use of a specific CBT components in combination with the provision of rewards. In general, little synergistic effects were found. Thus, there is a need to further assess the cumulative effect of different system components on treatment effectiveness and to analyze how specific combinations affect behavior change. To fully determine the most optimal delivery mode, further studies are needed that randomly assign participants to different platforms for delivery [128].

Although no consensus with respect to effectiveness of online peer support was reached yet [94,95], anonymous online support groups and discussion forums might help users to overcome the feeling of being stigmatized by connecting patients with others. A further advantage of these social support components is that time and location are no longer obstacles for active participation [96]. Nevertheless, further research is needed concerning the type of online support, such as expert-led or user-driven, moderated or nonmoderated, and synchronous (eg, chat rooms) or asynchronous (eg, discussion forums) [96]. From a design and engagement perspective, a nuanced view on target groups suggests that complementarity between content of interventions that target different mental disorders is crucial when designing computerized and Web-based mental health interventions. Patients with multiple disorders present a considerable challenge in the design of technology-supported interventions and, as a consequence, are often excluded from studies even though they might profit from certain components. One reason for the exclusion of comorbid and multimorbid patients is the pressure toward relatively stringent and precisely defined interventions, which are amenable to RCTs [130]. Regarding the high comorbidity of mental disorders, it is of greatest importance that this topic is given more attention in the design of technology-amenable interventions.

Conclusions

The development of MHIS targeting the change in health behavior requires great expertise and a thorough understanding of the problem area, underlying therapeutic strategies, and the design of persuasive systems. The findings of this systematic review contribute to the body of knowledge on the effectiveness of technology-supported therapeutic interventions for the treatment of depressive symptoms. This work is intended to provide a basis for the assessment of the impact of specific system components on treatment effects in RCTs of technology- and Web-based interventions for depression. Thus, the overall goal of this review was to identify such components and to enhance the understanding of the mechanisms through which technology-enabled interventions exert their therapeutic benefits by means of such.

Further quantitative studies are needed to assess the impact of identified components and to identify other system components that are relevant for the design of future technology-mediated MHIS for the treatment of depression and other mental disorders.

Because of the high relevance of the anatomy of MHIS, attention should be paid to design issues when developing new eHealth services in the future. To enhance dissemination and utilization of technology-based MHIS, the focus needs to be not only on how the interventions affect users, but also on how patients use and interact with technology and one another through them. Therefore, future studies are needed that add to the body of knowledge of technology-supported interventions for the treatment of depression by assessing patterns of program usage and user engagement.

To conclude, health information technology is a fast-growing field of research, which has the potential to effectively treat people suffering from mental disorders. Despite that, there is still room for improvement in the design of technology-based interventions for the treatment of depression. The delivery of

XSL•FO
RenderX

interventions via technology is a promising and cost-effective approach to diminish the significant treatment gap and the various barriers associated with the disorder.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Subgroup analyses by study quality. Q>3.5: high quality studies; Q<= 3.5: low quality studies.

[PNG File, 333KB - jmir_v19i5e191_app1.png]

Multimedia Appendix 2

Subgroup analyses by duration of treatment. Comparison of studies with a duration of equal or less than 10 weeks with trials of more than 10 weeks duration.

[PNG File, 334KB - jmir_v19i5e191_app2.png]

Multimedia Appendix 3

Subgroup analyses by face-to-face contact. Comparison of studies that made use of face-to-face therapy (F2F_Y) with those that did not use live therapist contact (F2F_N).

[PNG File, 332KB - jmir_v19i5e191_app3.png]

Multimedia Appendix 4

Subgroup analyses by the use of a CBT protocol. Comparison of interventions that were based on CBT (CBT_Y) with treatments that did not make use of CBT techniques (CBT_N).

[PNG File, 333KB - jmir_v19i5e191_app4.png]

Multimedia Appendix 5

Subgroup analyses by severity of baseline depression. Comparison of interventions that studied patients with a high level of baseline depression (SEV_H) with trials that included patients with a low level of baseline severity (SEV_L).

[PNG File, 332KB - jmir_v19i5e191_app5.png]

Multimedia Appendix 6

Subgroup analyses by age of included participants. Comparison of trials in A) adult patients (AGE_1) with B) trials that studied effects in adults excluding older adults (AGE_2), C) trials in older adults only (AGE_3), and D) trials that studied the effect on symptoms of depression in adolescents only (AGE_4).

[PNG File, 342KB - jmir_v19i5e191_app6.png]

Multimedia Appendix 7

Subgroup analyses by type of comparator. Comparison of trials with a control group that received treatment as usual (TAU) with RCTs that compared the intervention to a waiting list control (WLC) group.

[PNG File, 241KB - jmir_v19i5e191_app7.png]

References

1. World Health Organization. The WHO mental health policy and service guidance package URL: http://www.who.int/mental_health/policy/essentialpackage1/en/ [WebCite Cache ID 6q3vn0kDc]
2. 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;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
3. Griffiths K, Christensen H. Review of randomised controlled trials of Internet interventions for mental disorders and related conditions. Clinical Psychologist 2007:16-29. [doi: 10.1080/13284200500378696]

XSL•FO
RenderX

4. 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]
5. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
6. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
7. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
8. Saddichha S, Al-Desouki M, Lamia A, Linden I, Krausz M. Online interventions for depression and anxiety-a systematic review. Health Psychol Behav Med 2014 Jan 01;2(1):841-881. [doi: 10.1080/21642850.2014.945934] [Medline: 25750823]
9. Crutzen R, de Nooijer NJ, Brouwer W, Oenema A, Brug J, de Vries NK. A conceptual framework for understanding and improving adolescents' exposure to Internet-delivered interventions. Health Promot Int 2009 Sep;24(3):277-284. [doi: 10.1093/heapro/dap018] [Medline: 19515716]
10. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
11. Whitton AE, Proudfoot J, Clarke J, Birch M, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]
12. Branson CE, Clemmey P, Mukherjee P. Text message reminders to improve outpatient therapy attendance among adolescents: a pilot study. Psychol Serv 2013 Aug;10(3):298-303. [doi: 10.1037/a0026693] [Medline: 23937089]
13. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
14. Landenberger NA, Lipsey MW. The positive effects of cognitive–behavioral programs for offenders: a meta-analysis of factors associated with effective treatment. J Exp Criminol 2005 Dec 10;1(4):451-476. [doi: 10.1007/s11292-005-3541-7]
15. Topolovec-Vranic J, Mansoor Y, Ennis N, Lightfoot D. Technology-adaptable interventions for treating depression in adults with cognitive impairments: protocol for a systematic review. Syst Rev 2015 Apr 07;4:42 [FREE Full text] [doi: 10.1186/s13643-015-0032-4] [Medline: 25875655]
16. Wahle F, Kowatsch T. Towards the design of evidence-based mental health information systems: a preliminary literature review. In: Proceedings of the Thirty Fifth International Conference on Information Systems, Auckland 2014. 2014 Presented at: Thirty Fifth International Conference on Information Systems; Dec 14-17, 2014; Auckland, New Zealand p. 12-14.
17. Mayo Clinic. 2016. Cognitive behavioral therapy: what you can expect URL: http://www.mayoclinic.org/tests-procedures/cognitive-behavioral-therapy/details/what-you-can-expect/rec-20188674 [accessed 2017-01-22] [WebCite Cache ID 6nho3NPL1]
18. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(28):500 [FREE Full text]
19. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996 Feb;17(1):1-12. [Medline: 8721797]
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-e34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
21. American Psychiatric Association. American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders URL: http://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596 [accessed 2017-01-22] [WebCite Cache ID 6nhn8iuw4]
22. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
23. de Mello MF, de Jesus MJ, Bacaltchuk J, Verdeli H, Neugebauer R. A systematic review of research findings on the efficacy of interpersonal therapy for depressive disorders. Eur Arch Psychiatry Clin Neurosci 2005 Apr;255(2):75-82. [doi: 10.1007/s00406-004-0542-x] [Medline: 15812600]
24. Bell AC, D'Zurilla TJ. Problem-solving therapy for depression: a meta-analysis. Clin Psychol Rev 2009 Jun;29(4):348-353. [doi: 10.1016/j.cpr.2009.02.003] [Medline: 19299058]
25. Brent DA. A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Arch Gen Psychiatry 1997 Sep 01;54(9):877. [doi: 10.1001/archpsyc.1997.01830210125017]

XSL•FO
RenderX

26. Cuijpers P. A psychoeducational approach to the treatment of depression: a meta-analysis of Lewinsohn's "coping with depression" course. Behav Ther 1998;29(3):521-533. [doi: 10.1016/S0005-7894(98)80047-6]
27. Ströhle A. Physical activity, exercise, depression and anxiety disorders. J Neural Transm (Vienna) 2009 Jun;116(6):777-784. [doi: 10.1007/s00702-008-0092-x] [Medline: 18726137]
28. Beck AT. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [Medline: 13688369]
29. Uebelacker LA, German NM, Gaudiano BA, Miller IW. Patient health questionnaire depression scale as a suicide screening instrument in depressed primary care patients: a cross-sectional study. Prim Care Companion CNS Disord 2011;13(1):1 [FREE Full text] [doi: 10.4088/PCC.10m01027] [Medline: 21731830]
30. Agyapong VI, Ahern S, McLoughlin DM, Farren CK. Supportive text messaging for depression and comorbid alcohol use disorder: single-blind randomised trial. J Affect Disord 2012 Dec 10;141(2-3):168-176. [doi: 10.1016/j.jad.2012.02.040] [Medline: 22464008]
31. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
32. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry Internet 2005 Nov 1:456-461 [FREE Full text] [doi: 10.1192/bjp.187.5.456]
33. 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]
34. Burton C, Szentagotai TA, McKinstry B, Matheson C, Matu S, Moldovan R, et al. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
35. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki A, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
36. Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
37. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
38. de Graaf LE, Gerhards SA, Arntz A, Riper H, Metsemakers JF, Evers SM, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2009 Jul;195(1):73-80 [FREE Full text] [doi: 10.1192/bjp.bp.108.054429] [Medline: 19567900]
39. Holländare F, Johnsson S, Randestad M, Tillfors M, Carlbring P, Andersson G, et al. Randomized trial of Internet-based relapse prevention for partially remitted depression. Acta Psychiatr Scand 2011 Oct;124(4):285-294. [doi: 10.1111/j.1600-0447.2011.01698.x] [Medline: 21401534]
40. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsgaard T, Eisemann M, Waterloo K, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial. J Med Internet Res 2013;15(8):e153 [FREE Full text] [doi: 10.2196/jmir.2714] [Medline: 23916965]
41. Johansson R, Ekbladh S, Hebert A, Lindström M, Möller S, Petitt E, et al. Psychodynamic guided self-help for adult depression through the internet: a randomised controlled trial. PLoS One 2012;7(5):e38021 [FREE Full text] [doi: 10.1371/journal.pone.0038021] [Medline: 22741027]
42. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS One 2012;7(5):e36905 [FREE Full text] [doi: 10.1371/journal.pone.0036905] [Medline: 22615841]
43. Kay-Lambkin FJ, Baker AL, Lewin TJ, Carr VJ. Computer-based psychological treatment for comorbid depression and problematic alcohol and/or cannabis use: a randomized controlled trial of clinical efficacy. Addiction 2009 Mar;104(3):378-388. [doi: 10.1111/j.1360-0443.2008.02444.x] [Medline: 19207345]
44. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered Internet psychotherapy for depression in primary care: a randomised controlled trial. Lancet 2009 Aug 22;374(9690):628-634. [doi: 10.1016/S0140-6736(09)61257-5] [Medline: 19700005]
45. Kivi M, Eriksson MC, Hange D, Petersson E, Vernmark K, Johansson B, et al. Internet-based therapy for mild to moderate depression in Swedish primary care: short term results from the PRIM-NET randomized controlled trial. Cogn Behav Ther 2014;43(4):289-298 [FREE Full text] [doi: 10.1080/16506073.2014.921834] [Medline: 24911260]
46. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-Based Acceptance and Commitment Therapy for Depressive Symptoms With Minimal Support: A Randomized Controlled Trial. Behav Modif 2015 Aug 6;39(6):834. [doi: 10.1177/0145445515598142] [Medline: 26253644]

XSL•FO
RenderX

47. Lappalainen P, Granlund A, Siltanen S, Ahonen S, Vitikainen M, Tolvanen A, et al. ACT Internet-based vs face-to-face? A randomized controlled trial of two ways to deliver Acceptance and Commitment Therapy for depressive symptoms: an 18-month follow-up. Behav Res Ther 2014 Oct;61:43-54. [doi: 10.1016/j.brat.2014.07.006] [Medline: 25127179]
48. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
49. 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 [FREE Full text] [doi: 10.1016/j.invent.2014.12.003] [Medline: 19632969]
50. 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]
51. Morgan AJ, Jorm AF, Mackinnon AJ. Self-help for depression via e-mail: a randomised controlled trial of effects on depression and self-help behaviour. PLoS One 2013;8(6):e66537 [FREE Full text] [doi: 10.1371/journal.pone.0066537] [Medline: 23805231]
52. Moritz S, Schilling L, Hauschildt M, Schröder J, Treszl A. A randomized controlled trial of internet-based therapy in depression. Behav Res Ther 2012 Aug;50(7-8):513-521. [doi: 10.1016/j.brat.2012.04.006] [Medline: 22677231]
53. Wims E, Titov N, Andrews G, Choi I. Clinician-assisted Internet-based treatment is effective for panic: A randomized controlled trial. Aust N Z J Psychiatry 2010 Jul;44(7):599-607. [doi: 10.3109/00048671003614171] [Medline: 20560847]
54. Phillips R, Schneider J, Molosankwe I, Leese M, Foroushani PS, Grime P, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. Psychol Med 2014 Mar;44(4):741-752 [FREE Full text] [doi: 10.1017/S0033291713001323] [Medline: 23795621]
55. Preschl B, Maercker A, Wagner B, Forstmeier S, Baños RM, Alcañiz M, et al. Life-review therapy with computer supplements for depression in the elderly: a randomized controlled trial. Aging Ment Health 2012;16(8):964-974 [FREE Full text] [doi: 10.1080/13607863.2012.702726] [Medline: 22788983]
56. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31 [FREE Full text] [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
57. Richards D, Timulak L, Hevey D. A comparison of two online cognitive-behavioural interventions for symptoms of depression in a student population: The role of therapist responsiveness. Couns Psychother Res 2013 Sep;13(3):184-193. [doi: 10.1080/14733145.2012.733715]
58. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009 Dec;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]
59. Sheeber LB, Seeley JR, Feil EG, Davis B, Sorensen E, Kosty DB, et al. Development and pilot evaluation of an Internet-facilitated cognitive-behavioral intervention for maternal depression. J Consult Clin Psychol 2012 Oct;80(5):739-749 [FREE Full text] [doi: 10.1037/a0028820] [Medline: 22663903]
60. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
61. Ström M, Uckelstam C, Andersson G, Hassmén P, Umefjord G, Carlbring P. Internet-delivered therapist-guided physical activity for mild to moderate depression: a randomized controlled trial. PeerJ 2013;1:e178 [FREE Full text] [doi: 10.7717/peerj.178] [Medline: 24109561]
62. Titov N, Dear B, Ali S, Zou J, Lorian C, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. Behav Ther 2015 Mar;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008] [Medline: 25645168]
63. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
64. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
65. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-154:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
66. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]

XSL•FO
RenderX

67. Watanabe N, Furukawa TA, Shimodera S, Katsuki F, Fujita H, Sasaki M, et al. Can assessors in a psychotherapy trial be successfully blinded? Analysis of a randomized controlled trial on psychotherapy for refractory insomnia in residual depression. Psychother Psychosom 2013;82(6):401-403 [FREE Full text] [doi: 10.1159/000353174] [Medline: 24080757]
68. Effective Practice and Organisation of Care. Synthesising results when it does not make sense to do a meta-analysis. EPOC Resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services; 2013. URL: http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/20 [WebCite Cache ID 6nhnFGhQp]
69. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [Medline: 3802833]
70. Johansson R, Björklund M, Hornborg C, Karlsson S, Hesser H, Ljótsson B, et al. Affect-focused psychodynamic psychotherapy for depression and anxiety through the Internet: a randomized controlled trial. PeerJ 2013;1:e102 [FREE Full text] [doi: 10.7717/peerj.102] [Medline: 23862104]
71. Mullin A, Dear BF, Karin E, Wootton BM, Staples LG, Johnston L, et al. The UniWellbeing course: a randomised controlled trial of a transdiagnostic internet-delivered cognitive behavioural therapy (CBT) programme for university students with symptoms of anxiety and depression. Internet Interven 2015 May;2(2):128-136 [FREE Full text] [doi: 10.1016/j.invent.2015.02.002]
72. Newby JM, Mackenzie A, Williams AD, McIntyre K, Watts S, Wong N, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. Psychol Med 2013 Dec;43(12):2635-2648. [doi: 10.1017/S0033291713000111] [Medline: 23419552]
73. Proudfoot J, Swain S, Widmer S, Watkins E, Goldberg D, Marks I, et al. The development and beta-test of a computer-therapy program for anxiety and depression: hurdles and lessons. Comput Human Behav 2003 May;19(3):277-289. [doi: 10.1016/S0747-5632(02)00062-6]
74. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: A randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
75. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS One 2013;8(7):e62873 [FREE Full text] [doi: 10.1371/journal.pone.0062873] [Medline: 23843932]
76. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
77. Cochrane Community. About RevMan 5 URL: http://community.cochrane.org/tools/review-production-tools/revman-5/about [accessed 2017-01-22] [WebCite Cache ID 6nhna0xmV]
78. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Orlando, FL: Academic Press; 1985.
79. Balk E, Earley A, Patel K, Trikalinos T, Dahabreh I. Empirical Assessment of Within-Arm Correlation Imputation in Trials of Continuous Outcomes [Internet]. Report No.: 12(13)-EHC141-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012.
80. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press; 1977.
81. Watson FR. Statistical methods in cancer research. Major Probl Clin Surg 1979;5:613-684. [Medline: 459557]
82. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
83. Higgins JP, Green S, editors. Front Matter, in Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Chichester, UK: John Wiley & Sons; 2008.
84. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
85. van Gemert-Pijnen JE, Nijland N, van LM, 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]
86. Higgins J, Green S. Selecting studies and collecting data. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: Wiley-Blackwell; 2008:151-185.
87. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 2001 Oct;54(10):1046-1055. [Medline: 11576817]
88. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994 Dec;50(4):1088-1101. [Medline: 7786990]
89. Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. PLoS Med 2007 Feb;4(2):e61 [FREE Full text] [doi: 10.1371/journal.pmed.0040061] [Medline: 17311468]
90. Mobile Economy 2014. 2014. GSMA mobile economy 2014 URL: http://www.gsma.com/mobileeconomy/ [accessed 2017-01-22] [WebCite Cache ID 6nhp0vssi]

XSL•FO
RenderX

91. Chou W, Prestin A, Lyons C, Wen K. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013;103(1):18. [Medline: 1312689039]
92. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
93. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
94. Griffiths KM, Calear AL, Banfield M, Tam A. Systematic review on Internet Support Groups (ISGs) and depression (2): What is known about depression ISGs? J Med Internet Res 2009;11(3):e41 [FREE Full text] [doi: 10.2196/jmir.1303] [Medline: 19793718]
95. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
96. Lehto T, Oinas-Kukkonen H. Examining the persuasive potential of web-based health behavior change support systems. Trans Human-Computer Interact 2015;7(3):140.
97. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
98. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
99. 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]
100. 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]
101. Myrberg C, Wiberg N. Screen vs paper: what is the difference for reading and learning? Insights 2015 Jul 7;28(2):49-54 [FREE Full text] [doi: 10.1629/uksg.236]
102. Burns DD, Spangler DL. Does psychotherapy homework lead to improvements in depression in cognitive-behavioral therapy or does improvement lead to increased homework compliance? J Consult Clin Psychol 2000 Feb;68(1):46-56 [FREE Full text] [Medline: 10710839]
103. Lebeau RT, Davies C, Culver N, Craske M. Homework compliance counts in cognitive-behavioral therapy. Cogn Behav Ther 2013;42(3):171-179. [doi: 10.1080/16506073.2013.763286] [Medline: 23419077]
104. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
105. Prasko J, Mozny P, Novotny M, Slepecky M, Vyskocilova J. Self-reflection in cognitive behavioural therapy and supervision. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 2012 Dec;156(4):377-384 [FREE Full text] [doi: 10.5507/bp.2012.027] [Medline: 23073517]
106. Walther J, Deandrea D, Kim J, Anthony J. The influence of online comments on perceptions of antimarijuana public service announcements on YouTube. Hum Commun Res 2010;36(4):492. [doi: 10.1111/j.1468-2958.2010.01384.x]
107. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015 Mar;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
108. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
109. Wang H. Game reward systems: gaming experiences and social meanings. In: Proceedings of DiGRA conference: Think Design Play. 2011 Presented at: DiGRA; 14–17, 2011; Hilversum, The Netherlands p. 1-15 URL: http://gamelearninglab.nctu.edu.tw/ctsun/10.1.1.221.4931.pdf
110. Deterding S. Gamification. interactions 2012 Jul 01;19(4):14 [FREE Full text] [doi: 10.1145/2212877.2212883] [Medline: 317369900033]
111. Oinas-Kukkonen H, Harjumaa M. A systematic framework for designing and evaluating persuasive systems. In: Persuasive Technology. PERSUASIVE 2008. Lecture Notes in Computer Science, vol 5033. Berlin: Springer; 2008:164-169.
112. Gulz A, Haake M, Silvervarg A, Sjödén B, Veletsianos G. Building a social conversational pedagogical agent. In: Perez-Marin D, editor. Conversational Agents and Natural Language Interaction: Techniques and Effective Practices. Spain: Information Science Reference; 2011:128-155.
113. Baylor AL. The design of motivational agents and avatars. Education Tech Research Dev 2011 Feb 27;59(2):291-300. [doi: 10.1007/s11423-011-9196-3]
114. Baylor AL. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3559-3565 [FREE Full text] [doi: 10.1098/rstb.2009.0148] [Medline: 19884150]

XSL•FO
RenderX

115. de Bruin M, Sheeran P, Kok G, Hiemstra A, Prins JM, Hospers HJ, et al. Self-regulatory processes mediate the intention-behavior relation for adherence and exercise behaviors. Health Psychol 2012 Nov;31(6):695-703. [doi: 10.1037/a0027425] [Medline: 22390738]
116. Shaw K, O'Rourke P, Del MC, Kenardy J. Psychological interventions for overweight or obesity. Cochrane Database Syst Rev 2005 Apr 18(2):CD003818. [doi: 10.1002/14651858.CD003818.pub2] [Medline: 15846683]
117. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
118. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
119. van Dam HA, van der Horst FG, Knoops L, Ryckman RM, Crebolder HF, van den Borne BH. Social support in diabetes: a systematic review of controlled intervention studies. Patient Educ Couns 2005 Oct;59(1):1-12 [FREE Full text] [doi: 10.1016/j.pec.2004.11.001] [Medline: 16198213]
120. White M, Dorman SM. Receiving social support online: implications for health education. Health Educ Res 2001 Dec;16(6):693-707 [FREE Full text] [Medline: 11780708]
121. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
122. Beck Institute. 2014. Determining treatment length in CBT URL: https://www.beckinstitute.org/determining-treatment-length-in-cbt/ [WebCite Cache ID 6q3v77R8l]
123. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
124. Spittaels H, De Bourdeaudhuij BI, Brug J, Vandelanotte C. Effectiveness of an online computer-tailored physical activity intervention in a real-life setting. Health Educ Res 2007 Jun;22(3):385-396 [FREE Full text] [doi: 10.1093/her/cyl096] [Medline: 16971674]
125. Leahy R. Cognitive models of depression. In: Clinical Advances in Cognitive Psychotherapy: Theory and Application. New York: Springer Publishing Company; 2002:29-61.
126. Jones RB, Bental D, Cawsey A, Pearson J, Marsden J. Book review: Tailoring Health Messages: Customizing Communication with Computer Technology. Edited by Matthew Kreuter, David Farrell, Laura Olevitch and Laura Brennan. Lawrence Erlbaum Assoc, 1999. 270pp, ISBN: 0805833870. Psycho-oncol 2002 Nov;11(6):542-543 [FREE Full text] [doi: 10.1002/pon.609]
127. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
128. van Genugten L, Dusseldorp E, Webb TL, van Empelen P. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis. J Med Internet Res 2016;18(6):e155 [FREE Full text] [doi: 10.2196/jmir.4218] [Medline: 27268104]
129. Brouwer W, Kroeze W, Crutzen R, de Nooijer NJ, 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]
130. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press; 2012 Presented at: Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX p. 1421-1430 URL: http://dl.acm.org/citation.cfm?id=2208602
131. Patten CA, Rock E, Meis TM, Decker PA, Colligan RC, Pingree S, et al. Frequency and type of use of a home-based, Internet intervention for adolescent smoking cessation. J Adolesc Health 2007 Nov;41(5):437-443 [FREE Full text] [doi: 10.1016/j.jadohealth.2007.05.016] [Medline: 17950163]
132. Severson HH, Gordon JS, Danaher BG, Akers L. ChewFree.com: evaluation of a Web-based cessation program for smokeless tobacco users. Nicotine Tob Res 2008 Feb;10(2):381-391. [doi: 10.1080/14622200701824984] [Medline: 18236303]
133. Ferney SL, Marshall AL, Eakin EG, Owen N. Randomized trial of a neighborhood environment-focused physical activity website intervention. Prev Med 2009 Feb;48(2):144-150. [doi: 10.1016/j.ypmed.2008.10.022] [Medline: 19028519]
134. McConnon A, Kirk SF, Ransley JK. Process evaluation of an internet-based resource for weight control: use and views of an obese sample. J Nutr Educ Behav 2009;41(4):261-267. [doi: 10.1016/j.jneb.2008.07.008] [Medline: 19508931]
135. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]

XSL•FO
RenderX

136. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
137. Falconer CJ, Rovira A, King JA, Gilbert P, Antley A, Fearon P, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016 Jan;2(1):74-80 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002147] [Medline: 27703757]
138. Andrews G, Slade T, Issakidis C. Deconstructing current comorbidity: data from the Australian National Survey of Mental Health and Well-Being. Br J Psychiatry 2002 Oct;181:306-314 [FREE Full text] [Medline: 12356657]
139. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. Interact Comput 2010 Jul;22(4):243-252. [doi: 10.1016/j.intcom.2010.02.006]
140. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
141. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595 [FREE Full text] [doi: 10.1093/her/cyp063] [Medline: 19897515]
142. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
143. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]

Abbreviations

BDI: Beck Depression Inventory
CBT: cognitive behavior therapy
HIS: health information system
MHIS: mental health information system
PHQ: Patient Health Questionnaire
RCT: randomized controlled trial
SMD: standard mean difference
SMS: short message service
TAU: treatment as usual
WLC: waiting-list control

Edited by G Eysenbach; submitted 22.01.17; peer-reviewed by L Bijker, S Hermsen, T Aledavood, G Constantinescu; comments to author 26.02.17; revised version received 10.03.17; accepted 06.04.17; published 31.05.17.

Please cite as:
Wahle F, Bollhalder L, Kowatsch T, Fleisch E
Toward the Design of Evidence-Based Mental Health Information Systems for People With Depression: A Systematic Literature Review and Meta-Analysis
J Med Internet Res 2017;19(5):e191
URL: http://www.jmir.org/2017/5/e191/
doi:10.2196/jmir.7381
PMID:28566267

©Fabian Wahle, Lea Bollhalder, Tobias Kowatsch, Elgar Fleisch. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Response to Obamacare on Twitter

Matthew A Davis[1], MPH, PhD; Kai Zheng[2], PhD; Yang Liu[3], PhD; Helen Levy[4], PhD

[1]University of Michigan School of Nursing, Ann Arbor, MI, United States

[2]Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA, United States

[3]University of Michigan School of Information, Ann Arbor, MI, United States

[4]University of Michigan Institute for Social Research, Ann Arbor, MI, United States

Corresponding Author:
Matthew A Davis, MPH, PhD
University of Michigan School of Nursing
400 North Ingalls
Ann Arbor, MI, 48109
United States
Phone: 1 734 764 2814
Fax: 1 734 764 2416
Email: mattadav@umich.edu

Abstract

Background: The Affordable Care Act (ACA), often called "Obamacare," is a controversial law that has been implemented gradually since its enactment in 2010. Polls have consistently shown that public opinion of the ACA is quite negative.

Objective: The aim of our study was to examine the extent to which Twitter data can be used to measure public opinion of the ACA over time.

Methods: We prospectively collected a 10% random sample of daily tweets (approximately 52 million since July 2011) using Twitter's streaming application programming interface (API) from July 10, 2011 to July 31, 2015. Using a list of key terms and ACA-specific hashtags, we identified tweets about the ACA and examined the overall volume of tweets about the ACA in relation to key ACA events. We applied standard text sentiment analysis to assign each ACA tweet a measure of positivity or negativity and compared overall sentiment from Twitter with results from the Kaiser Family Foundation health tracking poll.

Results: Public opinion on Twitter (measured via sentiment analysis) was slightly more favorable than public opinion measured by the Kaiser poll (approximately 50% vs 40%, respectively) but trends over time in both favorable and unfavorable views were similar in both sources. The Twitter-based measures of opinion as well as the Kaiser poll changed very little over time: correlation coefficients for favorable and unfavorable public opinion were .43 and .37, respectively. However, we found substantial spikes in the volume of ACA-related tweets in response to key events in the law's implementation, such as the first open enrollment period in October 2013 and the Supreme Court decision in June 2012.

Conclusions: Twitter may be useful for tracking public opinion of health care reform as it appears to be comparable with conventional polling results. Moreover, in contrast with conventional polling, the overall amount of tweets also provides a potential indication of public interest of a particular issue at any point in time.

(J Med Internet Res 2017;19(5):e167) doi:10.2196/jmir.6946

KEYWORDS

Patient Protection and Affordable Care Act; health care reform; social media; data collection

Introduction

Americans have strong opinions about health care reform. Polls of the general public consistently indicate that less than half of Americans support the Affordable Care Act (ACA) [1]. The ACA (or "Obamacare," as it is more often called) is a federal statute that was enacted under President Barack Obama in 2010. Among the most significant health care reform efforts in US history, the ACA contains a series of provisions that have been implemented since it was signed into law. The overarching goal of the ACA is to expand access to affordable insurance coverage. However, the ACA remains controversial among policymakers and the general public. Although individual elements of the law, such as subsidies for low-income families to purchase health insurance or the individual mandate, may be more or less popular than the law as a whole [2], the public opinion regarding

XSL•FO
RenderX

the law has been surprisingly stable in the 5-plus years since the it was enacted. This stability has endured despite numerous ups and downs in the ACA's fortunes, including three major Supreme Court decisions (one mostly affirming and the other entirely affirming the law, with a third dealing its supporters a blow) and a botched rollout of the law's signature initiative, private health insurance exchanges. Table 1 presents a timeline of key events in the implementation of the ACA's coverage provisions [3-7].

Table 1. Timeline of key events related to the implementation of the Affordable Care Act.

Date	Event
March 23, 2010	ACA[a] signed into law by President Obama. Key coverage provisions—Medicaid expansion and health insurance exchanges—are scheduled to take effect in January 2014. Multiple lawsuits challenging different provisions of the law are filed shortly after its enactment.
2010-2011	Early ACA provisions are implemented, including consumer protections (eg, prohibitions on annual and lifetime caps on coverage) and the requirement that employer-sponsored plans must offer coverage for dependent children up to age 26 years. Most of these take effect as private plans were renewed; as a result, they do not have a single "headline" date for implementation.
December 19, 2011	The SCOTUS[b] announces it will hear oral arguments in NFIB[c] versus Sebelius, challenging the constitutionality of two key ACA provisions: the requirement that all individuals have coverage (the "individual mandate") and the expansion of Medicaid to all individuals with incomes below 138% of poverty.
March 26-28, 2012	The SCOTUS hears oral arguments in NFIB versus Sebelius, generating tremendous speculation.
June 28, 2012	The SCOTUS rules in NFIB versus Sebelius. The individual mandate is affirmed whereas the Medicaid expansion is effectively rendered optional for states: a mixed decision, but on balance regarded as a win for the ACA.
November 6, 2012	President Barack Obama reelected.
October 1, 2013	The first open enrollment period begins for private health insurance exchanges; the federal exchange website healthcare.gov fails to work properly, generating negative publicity.
November 26, 2013	The SCOTUS announces it will hear oral arguments in Burwell versus Hobby Lobby, challenging a private employer's refusal on religious grounds to provide full insurance coverage for contraception.
January 1, 2014	Expanded coverage through health insurance exchanges starts.
March 31, 2014	The first open enrollment period ends.
March 25, 2014	The SCOTUS announces it will hear oral arguments in Burwell versus Hobby Lobby, which is about whether corporations owned by religious families can refuse to comply with an ACA requirement that their health insurance must fully cover contraception for female workers.
June 30, 2014	The SCOTUS rules in favor of the corporations in Burwell versus Hobby Lobby (a blow to the ACA).
November 8, 2014	The SCOTUS announces it will hear oral arguments in King versus Burwell, which challenges the payment of federal subsidies for health insurance in states that rely on healthcare.gov (a majority of states).
November 15, 2014	The second open enrollment period begins for private health insurance exchanges; healthcare.gov works as intended.
February 15, 2015	The second open enrollment period ends for private health insurance exchanges.
March 4, 2015	The SCOTUS hears oral arguments in King versus Burwell.
June 25, 2015	The SCOTUS Court rules in favor of the Obama administration in King versus Burwell.

[a]ACA: Affordable Care Act.

[b]SCOTUS: Supreme Court of the United States.

[c]NFIM: National Federation of Independent Business.

Public opinion may have briefly dipped or risen immediately after these key events [8,9], but at the time of this writing, the law's favorable and unfavorable ratings in the Kaiser health tracking poll both stand at 42% —statistically indistinguishable, given the poll's ±3% point margin of error, from the 46% favorable or 40% unfavorable ratings the law had in April 2010, weeks after it was first enacted [10].

Monitoring public response to new laws and regulations, such as those included in the ACA, is of considerable interest to health policymakers, government agencies, and the media. Traditionally, measuring public response has relied on expensive and time-consuming surveys administered by polling agencies including the Pew Research Center and the Kaiser Family Foundation. Changes in technology introduce new opportunities for tracking public response. One particularly rich source of data is Twitter. Twitter has been used to study natural disasters [11,12], infectious disease outbreaks [13,14], drug and alcohol use [15,16], and public responses to health policies [17-19]. Whereas use of social media data has some limitations [20], it is inexpensive, immediate, and can offer contextual insights not captured by traditional survey questionnaires.

The aim of this paper was to examine the extent to which Twitter data can be used to measure public response to the rollout of the ACA. Our specific research questions were: (1) To what

XSL•FO
RenderX

extent can ACA-related tweets be accurately identified? (2) Does the overall volume of ACA-related tweets respond to key events in the implementation of ACA? (3) Is there an association between public opinion (ie, favorable vs unfavorable) measured using ACA-related tweets and conventional polling data from the Kaiser Family Foundation health tracking poll? and (4) What are common words used in favorable versus unfavorable ACA-related tweets?

Methods

Twitter Data

We examined the extent to which Twitter data can be used to measure public response to the ACA over time. To do so, we identified relevant tweets over a 6-year time period, examined them, and compared the ACA-related tweet sentiment with conventional polling data of public opinion. This study used publicly available data for all analyses and was deemed to be exempt from institutional board review.

We prospectively collected a 10% random sample of daily tweets (approximately 52 million since July 2011) using Twitter's streaming API (ie, the "Twitter Gardenhose") from July 10, 2011 to July 31, 2015. All analyses were restricted to English-language tweets.

To identify tweets about the ACA in this sample, we developed a list of key search terms. From the ACA Wikipedia page [21] and an arbitrary sample of comments to ACA lay media articles, we examined word frequencies in order to identify common text expressions related the ACA. Next, we used Google Trends to identify other words associated with Google searches about the ACA [22]. Based on the collection of terminology used across these data sources, we developed a list of common words and phrases to be used in our search (Textbox 1). Using regular expressions for our search terms (which account for differences in capitalization and spelling), we identified a tweet as being ACA-related if it included any of these words or phrases [23]. This resulted in a sample of 3,300,648 ACA-related tweets.

We also used Twitter hashtags to expand our identification of tweets about the ACA (Textbox 2). A Twitter hashtag is a short phrase that assigns a tweet to a specific topic. The inclusion of tweets identified exclusively based on ACA hashtags resulted in an addition of 75,133 tweets. Thus our final sample comprised 3,375,781 tweets potentially related to the ACA.

To check the validity of this method for identifying tweets related to the ACA we pulled a random sample of 1000 tweets. Two separate members of the research teach reviewed each tweet to determine if it was indeed relevant to the ACA. Thirty-seven of these tweets were not in fact ACA-related (in several the tweet in question used the term ACA as an abbreviation for "acapella" and the tweet was related to singing; Table 2). Therefore, we conclude that our identification strategy had a positive predictive value of 96.3%.

Textbox 1. Search terms used to identify tweets about the Affordable Care Act.

Terms used in tweets
• Affordable Care Cct or ACA
• Healthcare insurance exchanges
• Healthcare reform act or bill
• Healthcare insurance act or bill
• Obamacare
• Patient Protection and Affordable Care Act or PPACA

Textbox 2. Hashtags used to identify tweets about the Affordable Care Act.

Hashtags
• #ACA
• #aca
• #Obamacare
• #ObamaCare
• #obamacare

XSL•FO
RenderX

Table 2. Selected examples of relevant and nonrelevant Affordable Care Act-related tweets.

Type	Examples
Relevant favorable ACA[a] tweets	"Finally, my two favorite things come together: online shopping and buying health insurance."
	"In response to Obamacare, nearly 1 in 3 health facilities are adding doctors."
	"The GOP Is Terrified Obamacare Could Be a Success."
	"Thanks to the ACA, Over 5800 Californians with Pre-Existing Conditions Now Getting Care."
	"Obamacare winning one step at a time, sometimes take double steps. Today, good news for people with heart disease."
Relevant unfavorable ACA tweets	"Dems Throwing Granny Off the Cliff: Obamacare Cuts Medicare, Seniors Losing Doctors."
	"4 Years Later ObamaCare Still a Crime Against Democracy That The American People Will Never Accept."
	"Obamacare: Biggest Job-Killing TAX in US History!"
	"The people of America have no concept at this point as to just how miserable Obamacare is going to make individual lives."
	"Weird new error screen for Obamacare."
Nonrelevant ACA tweets	"You're one of those acapella girls, I'm one of those acapella boys, and we're gonna have aca-children."
	"Who's watching the ACA's?!"
	"Thank you guys so much for last night. The aca awards were a blast. Thanks for making 2013 unbelievable."

[a]ACA: Affordable Care Act or acapella.

Sentiment of Affordable Care Act (ACA) Tweets

We used standard text sentiment analysis to assign each ACA tweet a measure of positive to negative sentiment. Text sentiment analysis uses a lexicon of words each with previously assigned numeric measures of emotion (ranging from negative to positive, ie, −1.0 to +1.0). In this study, tweet sentiment was measured using labMT, a lexicon developed by Dodds et al based on human review of terms from language used in Twitter, Google Books, music lyrics, and the New York Times) [24]. This lexicon has been widely used in studies of Web-based product reviews and temporal patterns of happiness.

After tweets were processed to remove words that do not convey specific content (such as "a" or "the"), the assigned scores for words in a given text were summed up to arrive at an overall score of the sentiment. Tweets with a sentiment score greater than zero were coded favorable while those with a score less than zero were coded unfavorable.

Kaiser Family Foundation Health Tracking Poll

Since the enactment of the ACA in March 2010, the Kaiser Family Foundation's health tracking poll has been conducted monthly to evaluate the public views of the ACA [10]. Briefly, the Kaiser poll is a random telephone dial sample (both landline and cell) of approximately 1000 to 1500 persons annually aged 18 years and older residing in the United States. The poll collects basic information on sociodemographic characteristics, health, and, relevant to our study, specifically asks respondents: "As you may know, a health reform bill was signed into law in 2010. Given what you know about the health reform law, do you have a generally favorable or unfavorable opinion of it?" The response set for this questions consists of: (1) favorable, (2) unfavorable, and (3) don't know/refused.

From the Kaiser poll data, we determined the percent of respondents who reported being favorable versus unfavorable toward the ACA by month. Kaiser data were not available for 5 months (December 2012, January 2013, May 2013, July 2013, and August 2014).

Analyses

Descriptively, we sought to examine the influence of key events regarding the ACA implementation on public response. Therefore, we identified the following historical events that took place during the study period (see Table 1 for details): three Supreme Court cases regarding the ACA; the reelection of President Obama (November 6, 2012), the first exchange open enrollment period (from October 1, 2013 to March 31, 2014), the start date for major expansions of health insurance coverage through the ACA (January 1, 2014), and the second exchange open enrollment period (from November 15, 2014 to February 15, 2015).

Across calendar months, we used Spearman correlation to evaluate for associations between public response measured using ACA tweets and the Kaiser poll. For instance, we compared the percentage of unfavorable ACA tweets per month with the percentage of Kaiser poll respondents who were unfavorable toward the ACA. As young adults tend to use Twitter more than older adults, we also examined correlations stratified by the age of Kaiser poll respondents [25]. Finally, we determined how the overall volume of ACA tweets throughout the study period varied over time.

To determine whether Americans were referring to the ACA as "Obamacare" more or less over time, we show the volume of ACA-related tweets that do and do not contain this term. For all analyses, we used R statistical software (R Foundation for Statistical Computing, Vienna, Austria). A 2-sided *P* value of less than .05 was considered statistically significant.

XSL•FO
RenderX

We also performed a subanalysis to test the robustness of the associations we observed to determine whether tweets from political and special interest groups impacted our results. To do so, we reanalyzed associations after excluding the 310,862 clearly political ACA tweets that included hashtags such as #gop (Grand Old Party), #teaparty, #p2 (Progressive 2.0), #PJNET (Patriot Journalist Network), #tlot (Top Libertarians on Twitter), #ccot (Christian Conservatives on Twitter), and #tcot (Top Conservatives on Twitter).

Finally, to provide some insight into the content of favorable versus unfavorable ACA tweets, we calculated the most frequently used other words (ie, not used to identify the tweets as ACA-related) and displayed these using word clouds.

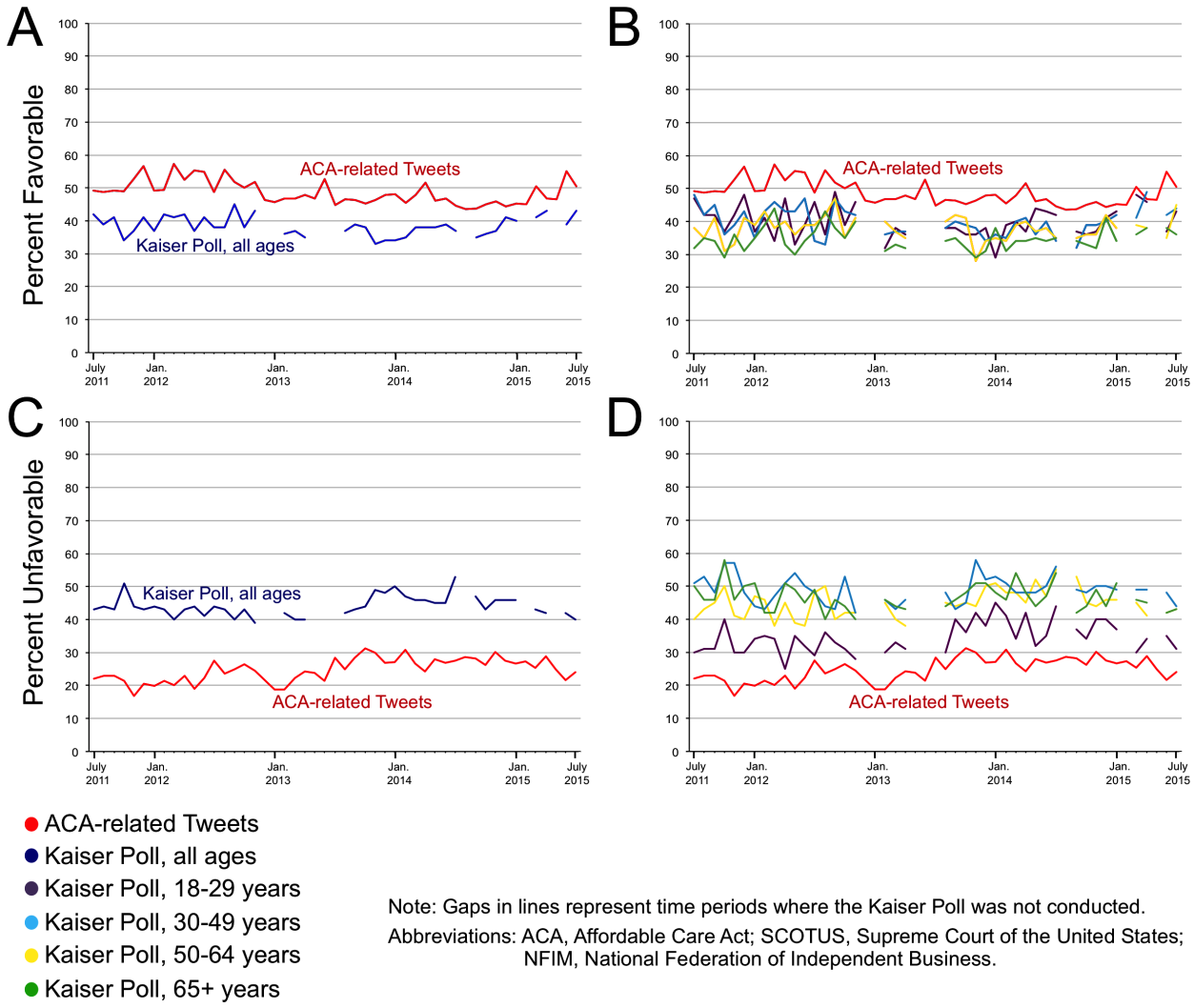
Results

Comparison of Public Response Using Tweets to the Kaiser Poll

Figure 1 displays the percent favorable and unfavorable ACA-related tweets over time compared with public response measured by the Kaiser poll. Gaps in the Kaiser lines are time periods where the poll was not conducted. Approximately 50% of tweets were favorable compared with approximately 40% of Kaiser respondents being favorable toward the ACA throughout the time period (Figure 1). According to specific age categories, the percent of favorable Kaiser poll respondents differed little over time. Over the 5-year time period, approximately 20% of ACA tweets were unfavorable compared with 40% of Kaiser respondents being unfavorable toward the ACA . Across age categories, older Kaiser respondents were more likely to be unfavorable toward the ACA (eg, approximately 50% of adults aged 65 years and older reported being unfavorable toward the ACA).

In spite of these differences our Twitter-based measure of public opinion track results of the Kaiser poll quite well over time. The correlation coefficient between percentage of unfavorable ACA tweets and Kaiser respondents was .43, *P* value=.01 over the study time period. Likewise, the correlation coefficient between percentage of favorable ACA tweets and Kaiser respondents was .37, *P* value=.02 (Table 3).

Figure 1. Favorable (A and B) versus unfavorable (C and D) public response to the Affordable Care Act using Tweets compared to results from the Kaiser Poll.



Note: Gaps in lines represent time periods where the Kaiser Poll was not conducted.

Abbreviations: ACA, Affordable Care Act; SCOTUS, Supreme Court of the United States; NFIM, National Federation of Independent Business.

XSL•FO
RenderX

Because Twitter users are likely to be younger than average, we also compared the percentage of favorable and unfavorable tweets according to Kaiser respondent age group. The strongest correlation was for unfavorable public response among Kaiser respondents between 18 and 29 years of age—a correlation coefficient of .47, *P* value=.01 (Table 3). The correlation coefficients for both favorable and unfavorable response between the two approaches were approximately .4, *P* value=.01 among Kaiser respondents between 30 and 49 years of age. Correlations were weak and statistically insignificant among older Kaiser respondents. These correlations persisted in our subanalysis where we remove tweets with political and special interest hashtags.

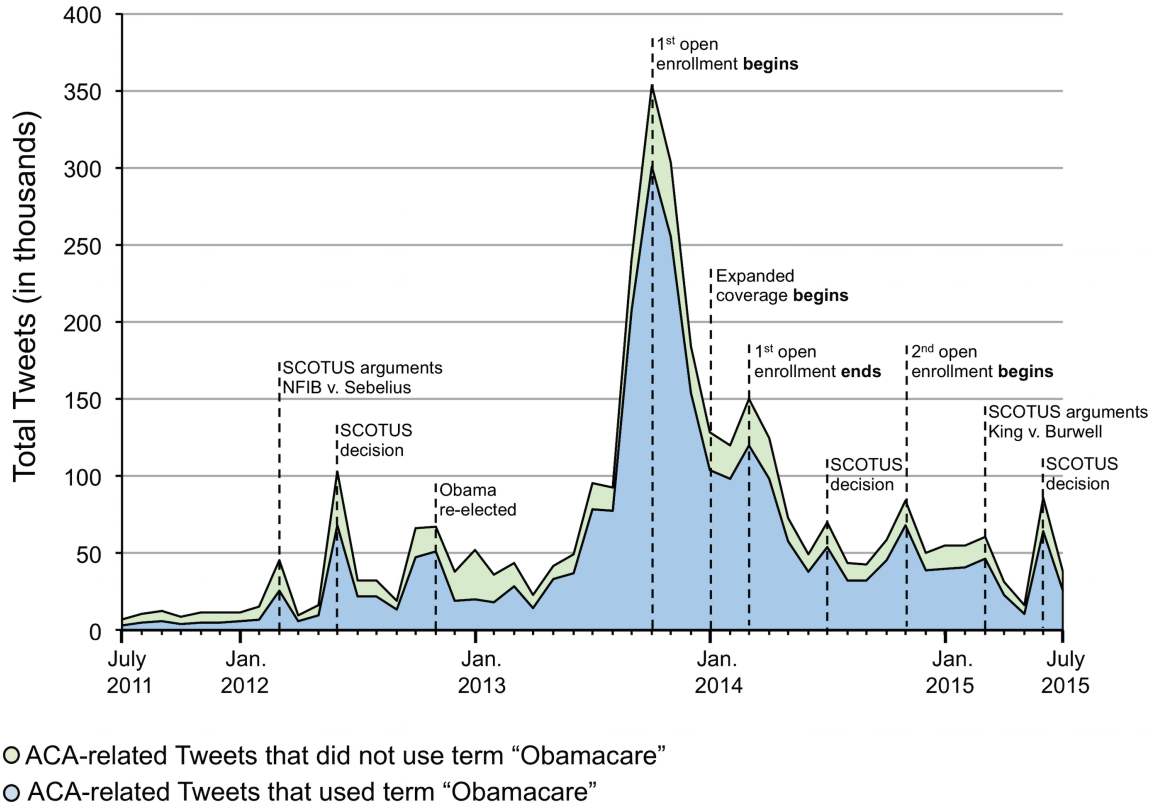
The Volume of ACA Tweets Over Time

Whereas public opinion may not change (much) in response to significant events in the ACA's history, the volume of ACA-related tweets certainly does (Figure 2). Overall, the number of ACA-related tweets peaked during the first open enrollment period. In particular, the single month with the largest amount of ACA-related tweets was October 2013 (a total of 353,890 ACA tweets) —the beginning of the first open enrollment period. Other notable events such as the Supreme Court of the United States decision in June 2012 and beginning of the second open enrollment period in November 2014 also led to sharp spikes in ACA tweets. The term "Obamacare" was used in the great majority of ACA-related tweets throughout this period, with no evidence that this term became more or less common over time.

Table 3. Correlation between percentage of favorable (or unfavorable) tweets and percentage of favorable (or unfavorable) Kaiser poll respondents about the Affordable Care Act.

Kaiser respondents	Spearman correlation coefficient (*P* value)	
	Favorable	Unfavorable
All	.37 (.02)	.43 (.01)
By age category		
18-29 years	.14 (.36)	.47 (.01)
30-49 years	.41 (.01)	.40 (.01)
50-64 years	.21 (.21)	.12 (.43)
65+ years	.22 (.17)	.08 (.59)

Figure 2. Total number of Affordable Care Act-related Tweets per month from July 2011 to January 2015.

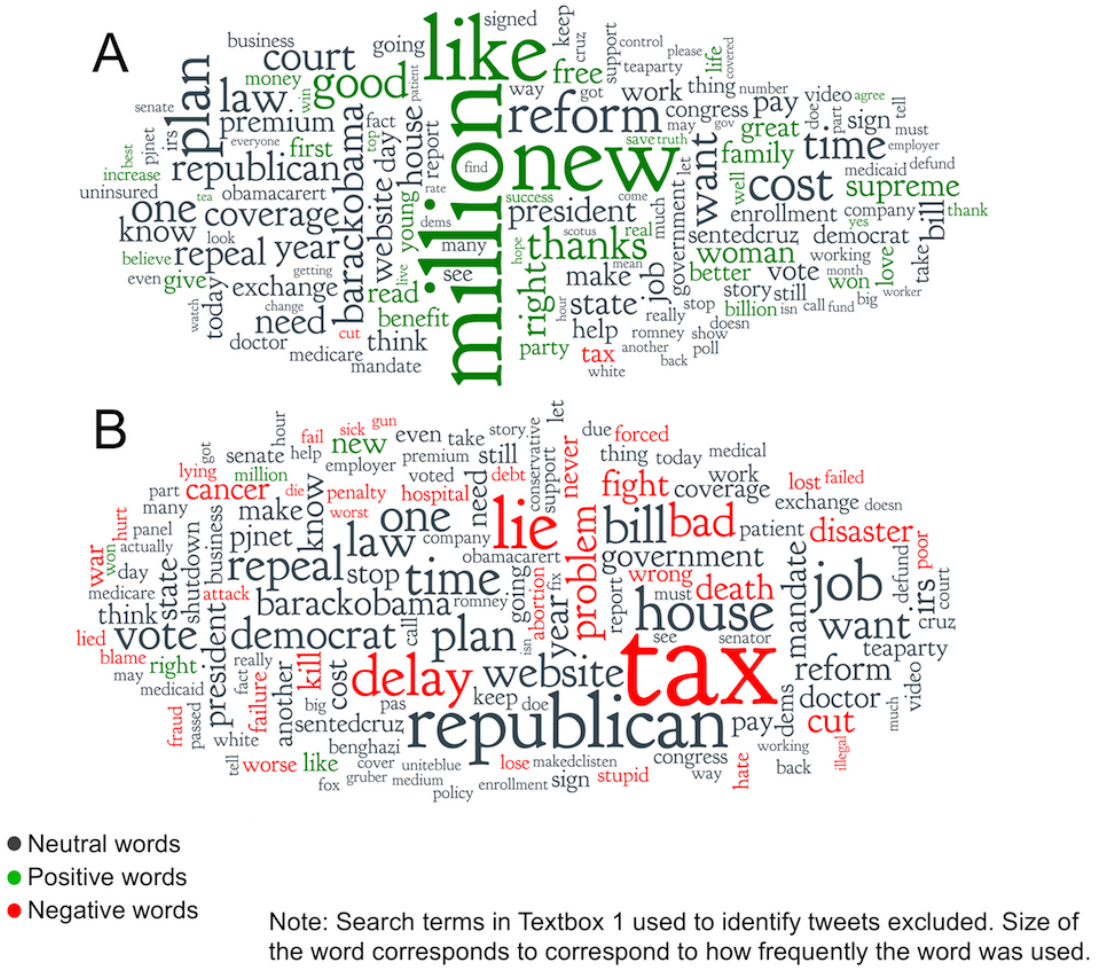


XSL•FO
RenderX

The Language of Favorable Versus Unfavorable ACA Tweets

Finally, in order to shed some light on the content of favorable and unfavorable ACA-related tweets, we tabulated the words most commonly used in each type of tweet, excluding the search terms in Textbox 1 that were used to identify ACA-related tweets initially. The results are presented as word clouds, an admittedly unscientific but nonetheless entertaining graphic tool in which size of the word corresponds to correspond to how frequently the word was used (Figure 3). A few of the most common words used in favorable ACA-related tweets included "like," "million," and "new." Unfavorable ACA-related tweets included a more eclectic mix of words such as "tax," "lie," and "delay."

Figure 3. Common words used in favorable (A) versus unfavorable (B) Affordable Care Act-related Tweets.



Discussion

Principal Findings

To our knowledge this is the first study to compare Twitter-based measures of public opinion regarding the ACA to traditional polling results. Overall, we found evidence that Twitter data can be effectively leveraged to estimate public opinion, including the response (or lack of response) of opinion to specific events such as health care reform. Trends in the overall public response measured by sentiment of tweets paralleled the results of Kaiser poll, and the levels of favorable and unfavorable response were quite similar over time. Not surprisingly, public response to the ACA on Twitter correlated most highly with polling data for younger adults —the age group most likely to use social media platforms [25]. For policymakers interested in tracking public response, our results suggest that Twitter data can provide a less costly and more immediate alternative to traditional opinion polling, particularly for younger Americans. Examining the text of tweets themselves can offer insight into public opinion (or perhaps, the language used by those wishing to influence public opinion) beyond that of traditional polls. For instance, unfavorable tweets used language regarding taxation, dishonesty, and other negative terminology. Words in tweets also hinted at political affiliations associated with tweeting about the ACA, for example, "republican," "democrat," and "teaparty" were commonly used in ACA tweets.

The most striking finding may be that our Twitter-based measures of ACA public response exhibit the same remarkable stability over time that characterizes results from the Kaiser poll. One of the central puzzles about public opinion toward the ACA—why are opinions changing so little over time, even as major components of the law have been implemented and provided health insurance coverage to millions of Americans?—is just as pronounced in the immediate-response

XSL•FO
RenderX

world of social media as in the more staid world of traditional opinion polling. Public opinion on Twitter toward the ACA could be more volatile, or more malleable, than opinions measured by the Kaiser poll—but they aren't. The lack of significant change in favorable or unfavorable views toward the ACA over time does not mean people aren't paying attention. On the contrary, they are not only paying attention, they are also expressing opinions in response to key events as we identified large changes in the volume of ACA tweets over time. The most striking spike in ACA tweets was in response to the first open enrollment period in October 2013 (during which the exchange enrollment website healthcare.gov failed to function properly). These large changes in volume, coupled with the lack of concomitant change in the favorable or unfavorable nature of the sentiments being addressed, echoes the thesis first advanced by Iyengar and Kinder that news may not so much change opinions as change how they are expressed [26]. In the age of social media, significant changes in the number of Americans expressing themselves in the absence of lack of change in public opinion might be reframed by saying that events in the real world do not change opinions, but they do change how often they are expressed.

There is growing use of Twitter to quantify public response. For instance, the sentiment expressed in tweets detected using either automated or manual annotation has been used to measure public response to vaccinations [27,28], the "Internet of things" [29], issues regarding electronic cigarettes [30], and climate change [31]. Collectively, this growing body of evidence points to encouraging findings regarding the use of Twitter as real-time barometer of collective public attitude. Our study contributes to several other studies that specifically used Twitter data to measure public response to health care reform [17-19,32]. Most similar to our study, King and colleagues used over 120,000 tweets related to the health and social care bill passed through parliament in England to examine public response [17]. They too found large spikes in the volume of tweets related to key events as the legislation was passed and moderate evidence of sentiment of tweets being correlated with public response. In a more recent report by Wong and colleagues, sentiment of ACA-related tweets were examined in relation to state-level health insurance marketplace enrollment [19]. After geocoding nearly 450,000 ACA-related tweets, the authors found a moderate association between ACA-related tweet sentiment and state-level enrollment. Our findings offer further support to the use of Twitter to quantify public response to health care reform as it correlates to some degree with a large national, ongoing poll. Furthermore, our findings of large spikes of tweets in relation to key events parallels findings observed in other studies (eg, [17,27,31,33]); in particular, we found significant fluxes in the amount of tweeting in the absence of swings in public opinion [17,26].

Study Limitations

The chief limitation of Twitter data is that Twitter users are, by definition, not representative of the general population. Any ways in which Twitter users are different from the typical American—for example, being younger, more tech-savvy, or having a different political orientation—could bias our Twitter-based estimate of ACA sentiment, if these underlying differences also affect attitudes toward the ACA. This is the reason why our analysis begins by comparing our estimates with estimates from the nationally-representative Kaiser poll. Other limitations include the fact that our algorithms for identifying ACA-related tweets and for encoding the sentiments they contain could introduce systematic bias. Whereas these methods represent the current state of the art in social media analysis, this relatively young field is evolving rapidly and subsequent methodological refinements may improve on the approach we use here.

Conclusions

In this study we found some evidence that Twitter may be useful for tracking public opinion regarding US health care reform as it appears to be comparable with conventional polling results. Similar to previous studies that used Twitter to measure public response, we found large changes in the amount of tweets in relation to key events; yet, during these time periods public opinion appeared to changed very little. Thus, the overall amount of tweets may also provide a potential indication of general public interest of a particular issue at any point in time. Whereas use of social media data for tracking public opinion is not without limitations, it is inexpensive, immediate, and can offer contextual insights beyond that of conventional polling.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) [K01 AT006162 (Davis) and K01 AG034232 (Levy)]. The NIH had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript. The views expressed are those of the authors and do not necessarily reflect those of the NIH.

Conflicts of Interest

None declared.

References

1. Blendon RJ, Benson JM. Voters and the affordable care act in the 2014 election. N Engl J Med 2014 Nov 13;371(20):e31. [doi: 10.1056/NEJMsr1412118] [Medline: 25353968]
2. Dalen JE, Waterbrook K, Alpert JS. Why do so many Americans oppose the affordable care act? Am J Med 2015 Aug;128(8):807-810. [doi: 10.1016/j.amjmed.2015.01.032] [Medline: 25731135]

XSL•FO
RenderX

3. Executive Office of the President, Council of Economic Advisers. Chapter 5: Health Care Reform. In: 2011 Economic Report of the President. Washington, DC: U.S. Government Printing Office; 2011.
4. Supreme Court of the United States, SCOTUSblog. National Federation of Independent Business v. Sebelius URL: http://www.scotusblog.com/case-files/cases/national-federation-of-independent-business-v-sebelius/ [accessed 2017-05-01] [WebCite Cache ID 6q95A5het]
5. Supreme Court of the United States, SCOTUSblog. Burwell v. Hobby Lobby Stores, Inc URL: http://www.scotusblog.com/case-files/cases/sebelius-v-hobby-lobby-stores-inc/ [accessed 2017-05-01] [WebCite Cache ID 6q95J6uL3]
6. Supreme Court of the United States, SCOTUSblog. King v. Burwell URL: http://www.scotusblog.com/case-files/cases/king-v-burwell/ [accessed 2017-05-01] [WebCite Cache ID 6q95UBrKi]
7. The Hill. Timeline of botched obamacare rollout URL: http://thehill.com/blogs/blog-briefing-room/news/190485-timeline-of-botched-implementation-of-obamacare [accessed 2017-05-01] [WebCite Cache ID 6q95XQs30]
8. Campbell AL, Persily N. The health care case in the public mind. In: Persily N, Metzger GE, Morrison TW, editors. The Health Care Case: The Supreme Court's Decision and Its Implications. New York, NY: Oxford Scholarship Online; 2013.
9. Kaiser Family Foundation Health Tracking Poll, 2013 URL: http://kff.org/health-reform/poll-finding/kaiser-health-tracking-poll-november-2013/ [WebCite Cache ID 6q96L3gH3]
10. Kaiser Family Foundation Health Tracking Poll, 2015: Tthe public's views on the ACA URL: http://kff.org/interactive/kaiser-health-tracking-poll-the-publics-views-on-the-aca/ [WebCite Cache ID 6q96QtqjW]
11. Black DR, Dietz JE, Stirratt AA, Coster DC. Do social media have a place in public health emergency response? J Emerg Manag 2015 May;13(3):217-226. [doi: 10.5055/jem.2015.0235] [Medline: 26150365]
12. Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, Cebrian M. Performance of social network sensors during Hurricane Sandy. PLoS One 2015 Feb;10(2):e0117288 [FREE Full text] [doi: 10.1371/journal.pone.0117288] [Medline: 25692690]
13. 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 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
14. 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 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
15. 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 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
16. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. "Hey Everyone, I'm Drunk." An evaluation of drinking-related twitter chatter. J Stud Alcohol Drugs 2015 Jul;76(4):635-643 [FREE Full text] [Medline: 26098041]
17. King D, Ramirez-Cano D, Greaves F, Vlaev I, Beales S, Darzi A. Twitter and the health reforms in the English National Health Service. Health Policy 2013 May;110(2-3):291-297. [doi: 10.1016/j.healthpol.2013.02.005] [Medline: 23489388]
18. Prabhu V, Lee T, Loeb S, Holmes JH, Gold HT, Lepor H, et al. Twitter response to the United States Preventive Services Task Force recommendations against screening with prostate-specific antigen. BJU Int 2015 Jul;116(1):65-71 [FREE Full text] [doi: 10.1111/bju.12748] [Medline: 24661474]
19. Wong CA, Sap M, Schwartz A, Town R, Baker T, Ungar L, et al. Twitter sentiment predicts affordable care act marketplace enrollment. J Med Internet Res 2015 Feb 23;17(2):e51 [FREE Full text] [doi: 10.2196/jmir.3812] [Medline: 25707038]
20. Pew Research Center. Twitter reaction to events often at odds with overall public opinion URL: http://www.pewresearch.org/2013/03/04/twitter-reaction-to-events-often-at-odds-with-overall-public-opinion/ [accessed 2017-05-01] [WebCite Cache ID 6q97L8j94]
21. Wikipedia. 2016. Patient protection and affordable care act URL: https://en.wikipedia.org/wiki/Patient_Protection_and_Affordable_Care_Act [accessed 2017-05-01] [WebCite Cache ID 6q97RkZli]
22. Google. 2016. Google Trends URL: https://trends.google.com/trends/ [accessed 2017-05-01] [WebCite Cache ID 6q97XhrP5]
23. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011 Sep;18(5):544-551 [FREE Full text] [doi: 10.1136/amiajnl-2011-000464] [Medline: 21846786]
24. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 2011 Dec;6(12):e26752 [FREE Full text] [doi: 10.1371/journal.pone.0026752] [Medline: 22163266]
25. Pew Research Center. 2015. Demographics of key social networking platforms URL: http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/ [accessed 2017-05-01] [WebCite Cache ID 6q97g5qzj]
26. Iyengar S, Kinder DR. News That Matters: Television and American Opinion. Chicago, IL: University of Chicago Press; 2010.
27. Becker BF, Larson HJ, Bonhoeffer J, van Mulligen EM, Kors JA, Sturkenboom MC. Evaluation of a multinational, multilingual vaccine debate on Twitter. Vaccine 2016 Dec 07;34(50):6166-6171. [doi: 10.1016/j.vaccine.2016.11.007] [Medline: 27840012]
28. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]

XSL•FO
RenderX

29. Bian J, Yoshigoe K, Hicks A, Yuan J, He Z, Xie M, et al. Mining twitter to Assess the public perception of the "Internet of Things". PLoS One 2016 Jul;11(7):e0158450 [FREE Full text] [doi: 10.1371/journal.pone.0158450] [Medline: 27391760]
30. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on twitter. J Med Internet Res 2015 Oct 27;17(10):e243 [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]
31. Cody EM, Reagan AJ, Mitchell L, Dodds PS, Danforth CM. Climate change sentiment on twitter: an unsolicited public opinion poll. PLoS One 2015 Aug;10(8):e0136092 [FREE Full text] [doi: 10.1371/journal.pone.0136092] [Medline: 26291877]
32. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette twitter campaign. J Med Internet Res 2014;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
33. Kendra RL, Karki S, Eickholt JL, Gandy L. Characterizing the discussion of antibiotics in the twittersphere: what is the bigger picture? J Med Internet Res 2015;17(6):e154 [FREE Full text] [doi: 10.2196/jmir.4220] [Medline: 26091775]

Abbreviations

ACA: Affordable Care Act
API: application programming interface

Edited by G Eysenbach; submitted 05.11.16; peer-reviewed by F Rothenfluh, SH Hong; comments to author 05.01.17; revised version received 01.03.17; accepted 06.03.17; published 26.05.17.

Please cite as:
Davis MA, Zheng K, Liu Y, Levy H
Public Response to Obamacare on Twitter
J Med Internet Res 2017;19(5):e167
URL: http://www.jmir.org/2017/5/e167/
doi:10.2196/jmir.6946
PMID:28550002

©Matthew A Davis, Kai Zheng, Yang Liu, Helen Levy. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 26.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prioritizing Measures of Digital Patient Engagement: A Delphi Expert Panel Study

Lynn A Garvin[1*], MBA, PhD; Steven R Simon[2,3,4*], MPH, MD

[1]Department of Health Law, Policy and Management, School of Public Health, Boston University, Boston, MA, United States

[2]Geriatrics and Extended Care Service, Veterans Affairs Boston Healthcare System, Boston, MA, United States

[3]The Center for Healthcare Organization and Implementation Research, Veterans Affairs Boston Healthcare System, Boston, MA, United States

[4]Division of General Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Lynn A Garvin, MBA, PhD
Department of Health Law, Policy and Management
School of Public Health
Boston University
Talbot 342 W
715 Albany Street
Boston, MA, 02118
United States
Phone: 1 617 414 1327
Email: lagarvin@bu.edu

Abstract

Background: Establishing a validated scale of patient engagement through use of information technology (ie, digital patient engagement) is the first step to understanding its role in health and health care quality, outcomes, and efficient implementation by health care providers and systems.

Objective: The aim of this study was to develop and prioritize measures of digital patient engagement based on patients' use of the US Department of Veterans Affairs (VA)'s MyHealtheVet (MHV) portal, focusing on the MHV/Blue Button and Secure Messaging functions.

Methods: We aligned two models from the information systems and organizational behavior literatures to create a theory-based model of digital patient engagement. On the basis of this model, we conducted ten key informant interviews to identify potential measures from existing VA studies and consolidated the measures. We then conducted three rounds of modified Delphi rating by 12 national eHealth experts via Web-based surveys to prioritize the measures.

Results: All 12 experts completed the study's three rounds of modified Delphi ratings, resulting in two sets of final candidate measures representing digital patient engagement for Secure Messaging (58 measures) and MHV/Blue Button (71 measures). These measure sets map to Donabedian's three types of quality measures: (1) antecedents (eg, patient demographics); (2) processes (eg, a novel measure of Web-based care quality); and (3) outcomes (eg, patient engagement).

Conclusions: This national expert panel study using a modified Delphi technique prioritized candidate measures to assess digital patient engagement through patients' use of VA's My HealtheVet portal. The process yielded two robust measures sets prepared for future piloting and validation in surveys among Veterans.

(J Med Internet Res 2017;19(5):e182) doi:10.2196/jmir.4778

KEYWORDS

patient reported outcome measures; patient engagement; patient-centered care; personal health record; health information technology; veterans health

XSL•FO
RenderX

Introduction

Patient portals are Web-based platforms that provide patients with access to health information and elements of their medical record and equip them with tools to interact with their clinical teams [1]. Numerous studies have assessed the adoption of patient portals, including the factors that predispose patients to adoption as well as the barriers and challenges they face [2-8]. Moving beyond adoption, measuring the nature and extent of patients' use of portal tools is a priority for the US Department of Veterans Affairs (VA) and other health systems. This emerging focus on measuring *digital patient engagement* [9] follows organically from—and is inextricably linked with—measuring *clinician meaningful use* of health information technology (HIT) as supported by the US Centers for Medicare & Medicaid Services's "Meaningful Use" rules [10]. However, to date, neither VA nor any other US health system has established nationally validated measures of digital patient engagement to provide an indication of the extent to which patients are genuinely using portal-based tools and the degree to which those tools are engaging patients with their health care. The main objective of this study was therefore to develop and prioritize measures of patients' experience using functions of a patient portal, with specific attention to the degree to which their experience using those functions promotes engagement with their health care team. To achieve this objective, we used a modified Delphi panel. This systematic approach aggregated experts' opinions and perceptions of which measures would be most valuable and appropriate for assessing Veterans' use of Blue Button and Secure Messaging, two salient features of My HealtheVet, VA's patient portal. [11]. Our study focused on the Blue Button function, which allows Veterans to download their personal health record, and the Secure Messaging function, which enables Veterans to communicate via secure email with their health care team.

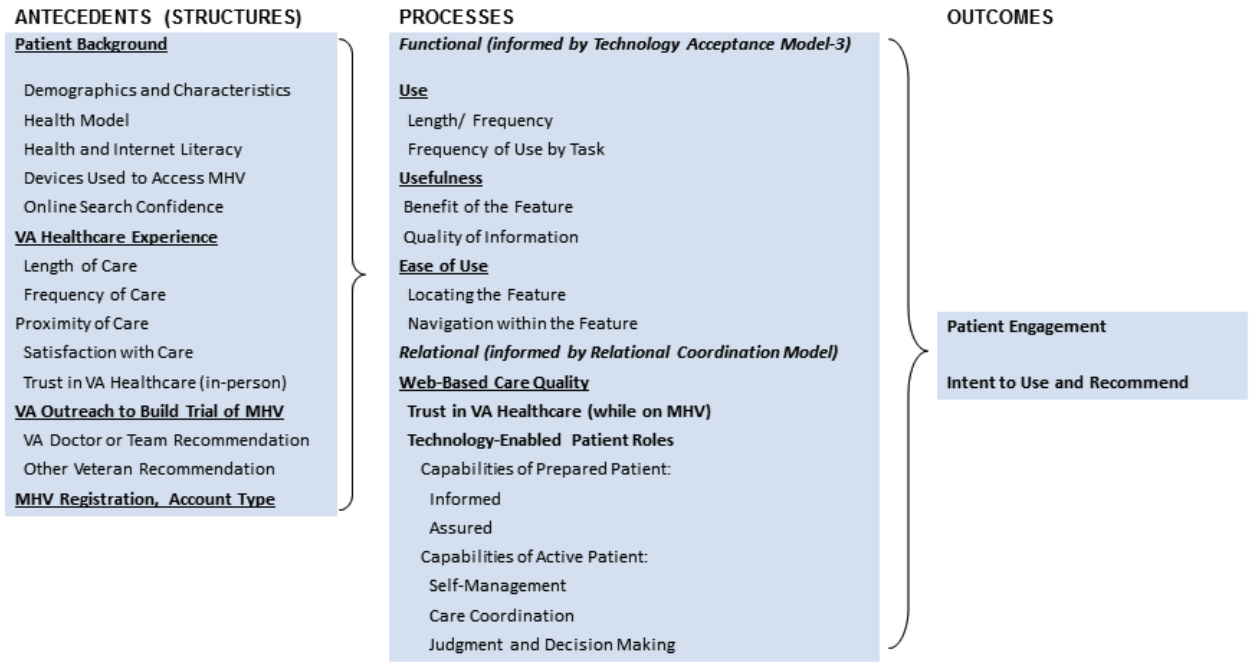
Methods

This study involved a sequence of three phases: (1) literature review; (2) key informant interviews; and (3) Delphi panel process. For the literature review, we sought to identify prior work that would enable us to design a theory-based model for the study; as such, we researched three literature streams of established frameworks and validated scales of patients': (1) health and health care, (2) use of IT or HIT, and (3) relations with providers and health care systems. This literature review indicated that existing models of patient engagement and of technology adoption did not sufficiently overlap or integrate with each other to provide a framework for measuring patients' engagement in their health and health care through technology.

We therefore defined *digital patient engagement* as the value that a patient (or family member or caregiver) assigns to the accrued experience with and results of using a system feature or service plus the expectation of similar future experience and results. On the basis of our theory-based definition, we aligned two established models: (1) Technology Acceptance Model [12] to reflect the functional dimension and (2) Relational Coordination Model to reflect the patient-provider dimension of digital patient engagement [13]. The resulting model included both patient engagement and intent to use and recommend the technology as outcomes reflecting digital patient engagement.

Our resulting theory-based model is shown in Figure 1. The "account type" mentioned in the figure refers to the type of My HealtheVet account the user possesses. Account types included basic, which provided online access to general health information; advanced, which included access to the Blue Button and the ability to view other elements of the personal health record; and premium, which included all of the advanced features but added secure messaging capability.

Figure 1. Digital patient engagement model.



XSL•FO
RenderX

For the second phase of the study, in late 2012, we conducted semistructured key informant telephone interviews with principal investigators of all current VA-funded MyHealtheVet (MHV) studies (N=10) to identify existing measures of MHV adoption and use. The ten participants were identified through communication with the MHV Program Office and with researchers in the field. One or both of the authors conducted telephone interviews with each participant.

Following interviews, we conducted the third phase of the study, a modified Delphi process. We followed the methods of previously published studies [14-21] and tailored them to the objectives herein. The modified Delphi technique is an intensive, iterative approach to elicit and refine experts' opinions on novel conceptual fields with the goals of gaining consensus on candidate measures and evolving the framework. From the literature review and in-depth interviews, we generated a set of candidate measures to be considered by the Delphi panel, with the goal of identifying measures of use of the Blue Button and Secure Messaging that represented digital patient engagement. On the basis of Delphi theory, our in-depth interviews, and the theoretical model, we established success criteria to guide inclusion in the preliminary measures, process goals, and consensus criteria for each round of the Delphi process.

For the Delphi panel, we convened 12 national (US) eHealth experts, who were principally physicians. Our Delphi protocol, conducted in March-October 2013, involved three rounds of panelists' independent rating of the measures; panelists submitted their ratings through a secure online questionnaire, enabling the research team to score and analyze the results while maintaining panelist anonymity to all but the researchers. To ensure that the process would ultimately yield measures of digital patient engagement, and following procedures established in prior Delphi panel studies, we asked panelists to rate the importance of each proposed measure on an 11-point integer scale, ranging from −5 (strongly disagree) to +5 (strongly agree). Our objective criteria enabled us to accept or reject a measure after each round, or to revise it for retesting in the next round. For acceptance, a measure was required to meet all three of the following conditions: (1) median score ≥+3; (2) interquartile range (IQR) ≤2; and (3) ≤1 outliers (defined as a score of >1.5×IQR from the 25th or 75th percentile). For example, consider the following 12 panelists' scores for one measure: 0, 0, 1, 2, 3, 3, 3, 3, 4, 4, 4, 5. The median score is 3, satisfying condition (1). The IQR is 2, satisfying condition (2). There are no outliers, that is, no scores lower than −1 and no scores greater than 7 (the latter not being a possible value, given the −5 to +5 scale), thus satisfying condition (3). Therefore, this measure would be accepted and not considered further in subsequent panel iterations. For revision and retesting (in a subsequent round), a measure was required to meet two of the three conditions. If it failed to meet at least two conditions, it was rejected.

Results

All 12 Delphi expert panel members completed the study's three rounds of measures rating. The final candidate measures of digital patient engagement comprised two similar but separate sets: 58 measures for Secure Messaging and 71 measures for MHV/Blue Button, where Antecedents represented 20 comparable measures for both functions, Processes represented 32 Secure Messaging and 45 MHV/Blue Button measures, and Outcomes represented six comparable measures for Secure Messaging and MHV/Blue Button.

As an example of how the Delphi panel results were used to include or exclude measures, Table 1 shows the four digital patient engagement outcome measures. For each of these four outcome measures, acceptance was achieved in the third and final round of the Panel's deliberations. Among the four measures, there was only one outlier panel member rating, reflecting high level of agreement on the value of these items in measuring digital patient engagement.

Table 1. Digital patient engagement outcome measures—Delphi panel statistics. Secure Messaging = first statistic and MyHealtheVet /Blue Button = second statistic, reported (in #/# format).

Measure: patient engagement	Accept round	IQR[a]	Outlier	Median	Mean	SD[b]
I have all the information I need to manage my health and health care.	3	1.5/2.0	1/0	3.5/4.0	3.1/4.0	2.3/0.9
I am confident in working with my VA[c]health care team to manage my health and health care.	3	2.0/1.5	0/0	3.0/3.0	3.7/3.6	1.0/0.8
I feel in control of my health and health care (such as taking part in decisions or following through on any medication, treatment, or health routine).	3	1.5/2.0	0/0	4.0/4.0	3.5/3.6	1.4/1.5
I am able to achieve my long-term health and health care goals (such as being self-reliant, living longer and better, or knowing that my family and friends can depend on me).	3	2.0/2.0	0/0	3.5/3.5	3.5/3.6	1.5/1.6

[a]IQR: interquartile range.

[b]SD: standard deviation.

[c]VA: US Department of Veterans Affairs.

XSL•FO
RenderX

Discussion

Principal Findings

Measuring how patients use HIT is a high priority for US health care. Nevertheless, existing meaningful use measurements have focused on clinicians' use of technology with few guidelines for patient adoption and use. While various scales have emerged to assess patient engagement and satisfaction with health care, none has combined patients' affinity for the technology with patients' trust in their relationship with clinicians, in person and online, to demonstrate how these variables influence digital patient engagement with their health and health care.

In this national expert panel study using a modified Delphi technique, we consolidated and refined two complementary versions of candidate measures to assess patients' use of VA's My HealtheVet patient portal, one set of measures for its Secure Messaging feature and another for its Blue Button personal health record and other MHV tools, with the potential for gauging digital patient engagement.

This study offers a number of strengths and innovations. Guided by a theory-based framework, we first developed and refined a new four-item digital patient engagement outcome measure based on Rogers' "Diffusion of Innovation" model [22]. That is, the four new measures we developed (ie, awareness and understanding, skills and confidence, trial and regular use, and use loyalty and recommendation) map roughly to Rogers' four stages of diffusion of innovation, namely knowledge, persuasion, decision, and confirmation. Grounding our measures in Rogers' framework was suggested by and affirmed by the Delphi panel. The four-item digital patient engagement measures also align with Hibbard et al's patient activation measures [23], paraphrased as (1) belief in having an active role in care, (2) confidence and knowledge to take action, (3) taking action, and (4) staying the course under stress.

As an innovation, the Delphi panel led us to introduce the novel process dimension of patient online care quality. This measure reflects the quality of the interaction of users with the technology.

We strengthened the content validity of the measures, a principal goal of using the Delphi technique, by assuring 100% participation of our content experts across the three rating rounds. To mitigate threats to external validity, such as selection bias, we selected our national experts to reflect a broad perspective of various health systems, diverse patient user groups, and an array of patient portal architectures and features. Generalizability of results also benefited from the inclusion of the expert views of the key informants. We reduced Delphi panel process threats by (1) making study goals and procedural guidelines clear at the start, (2) presenting a fair and transparent rating process with timely survey administration and response to panelist questions, and (3) extending full consideration and discussion on any dissenting opinions by panelists.

Limitations

A potential limitation of the study is that we employed a panel of experts, rather than patients themselves, to refine and prioritize the measures for digital patient engagement. We chose our approach because we considered the tasks of measure selection to require not only familiarity with the patient portal tools and their role in health care delivery but also a comfort level with the process of questionnaire item development and measurement scales. To ensure that the measures developed in this study truly reflect digital patient engagement, they must be validated among a population of patients who are users of the portal and its functions.

Conclusions

Establishing a valid and reliable scale is the first step to measuring digital patient engagement and its role in health and health care quality, outcomes, and effective, efficient implementation by health care providers and health care systems. This study yielded a robust set of candidate measures of what Veterans value in Blue Button and Secure Messaging. These measures and the scales they constitute can thus be tested empirically to examine their psychometric properties and may ultimately be used in measuring the extent to which patient portals and other patient-facing technologies can engage patients in their health care.

Conflicts of Interest

None declared.

References

1. HealthIT. What is a patient portal? URL: https://www.healthit.gov/providers-professionals/faqs/what-patient-portal [accessed 2017-05-03] [WebCite Cache ID 6qBwdW0JN]
2. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
3. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
4. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]

XSL•FO
RenderX

5. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and using a patient portal: a qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016 Jul 04;23(2):848 [FREE Full text] [Medline: 27869582]
6. Fix GM, Hogan TP, Amante DJ, McInnes DK, Nazi KM, Simon SR. Encouraging patient portal use in the patient-centered medical home: three stakeholder perspectives. J Med Internet Res 2016 Nov 22;18(11):e308 [FREE Full text] [doi: 10.2196/jmir.6488] [Medline: 27876686]
7. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 02;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
8. Mishuris RG, Stewart M, Fix GM, Marcello T, McInnes DK, Hogan TP, et al. Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study. Health Expect 2015 Dec;18(6):2296-2305. [doi: 10.1111/hex.12199] [Medline: 24816246]
9. Dolan B. Mobihealthnews. In-Depth: provider adoption of digital patient engagement tools URL: http://www.mobihealthnews.com/41542/in-depth-provider-adoption-of-digital-patient-engagement-tools/ [accessed 2017-05-03] [WebCite Cache ID 6qBwkw4Cs]
10. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
11. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011;6(6):e20476 [FREE Full text] [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
12. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
13. Gittell JH. New Directions for Relational Coordination Theory. Oxford: Oxford University Press; 2011.
14. Dalkey N, Helmer O. An experimental application of the DELPHI method to the use of experts. Manag Sci 1963 Apr;9(3):458-467. [doi: 10.1287/mnsc.9.3.458]
15. Sackman H. Delphi Technique: Expert Opinion, Forecasting and Group Process. Lexington, MA: Lexington Books; 1975.
16. Lindsay P, Schull M, Bronskill S, Anderson G. The development of indicators to measure the quality of clinical care in emergency departments following a modified-delphi approach. Acad Emerg Med 2002 Nov;9(11):1131-1139 [FREE Full text] [Medline: 12414461]
17. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Information & Management 2004 Dec;42(1):15-29. [doi: 10.1016/j.im.2003.11.002]
18. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Med Res Methodol 2005 Dec 01;5:37 [FREE Full text] [doi: 10.1186/1471-2288-5-37] [Medline: 16321161]
19. Deshpande AM, Shiffman RN, Nadkarni PM. Metadata-driven Delphi rating on the Internet. Comput Methods Programs Biomed 2005 Jan;77(1):49-56. [doi: 10.1016/j.cmpb.2004.05.006] [Medline: 15639709]
20. Landeta J. Current validity of the Delphi method in social sciences. Technol Forecast Soc Change 2006;73(5):467-482.
21. Schneider F, van Osch L, de Vries H. Identifying factors for optimal development of health-related websites: a Delphi study among experts and potential future users. J Med Internet Res 2012 Feb 14;14(1):e18 [FREE Full text] [doi: 10.2196/jmir.1863] [Medline: 22357411]
22. Rogers EM. Diffusion of Innovation Theory. Boston, MA: The Free Press; 1995.
23. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]

Abbreviations

HIT: health information technology
IQR: Interquartile range
MHV: MyHealtheVet
VA: US Department of Veterans Affairs

XSL•FO
RenderX

Edited by G Eysenbach; submitted 30.05.15; peer-reviewed by D Sands, L Laranjo, S Baker; comments to author 19.10.15; revised version received 01.03.17; accepted 18.03.17; published 26.05.17.

Please cite as:
Garvin LA, Simon SR
Prioritizing Measures of Digital Patient Engagement: A Delphi Expert Panel Study
J Med Internet Res 2017;19(5):e182
URL: http://www.jmir.org/2017/5/e182/
doi:10.2196/jmir.4778
PMID:28550008

©Lynn A Garvin, Steven R Simon. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 26.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer

Peter Hanlon[1], MSc, MBChB; Luke Daines[2], MBChB; Christine Campbell[1], PhD; Brian McKinstry[3], MD; David Weller[1], PhD; Hilary Pinnock[2], MD

[1]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

[2]Allergy and Respiratory Research Group, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

[3]E-Health Group, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Hilary Pinnock, MD
Allergy and Respiratory Research Group
Usher Institute of Population Health Sciences and Informatics
University of Edinburgh
Medical School
Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 131 650 9474
Fax: 44 131 650 9119
Email: hilary.pinnock@ed.ac.uk

Abstract

Background: Self-management support is one mechanism by which telehealth interventions have been proposed to facilitate management of long-term conditions.

Objective: The objectives of this metareview were to (1) assess the impact of telehealth interventions to support self-management on disease control and health care utilization, and (2) identify components of telehealth support and their impact on disease control and the process of self-management. Our goal was to synthesise evidence for telehealth-supported self-management of diabetes (types 1 and 2), heart failure, asthma, chronic obstructive pulmonary disease (COPD) and cancer to identify components of effective self-management support.

Methods: We performed a metareview (a systematic review of systematic reviews) of randomized controlled trials (RCTs) of telehealth interventions to support self-management in 6 exemplar long-term conditions. We searched 7 databases for reviews published from January 2000 to May 2016 and screened identified studies against eligibility criteria. We weighted reviews by quality (revised A Measurement Tool to Assess Systematic Reviews), size, and relevance. We then combined our results in a narrative synthesis and using harvest plots.

Results: We included 53 systematic reviews, comprising 232 unique RCTs. Reviews concerned diabetes (type 1: n=6; type 2, n=11; mixed, n=19), heart failure (n=9), asthma (n=8), COPD (n=8), and cancer (n=3). Findings varied between and within disease areas. The highest-weighted reviews showed that blood glucose telemonitoring with feedback and some educational and lifestyle interventions improved glycemic control in type 2, but not type 1, diabetes, and that telemonitoring and telephone interventions reduced mortality and hospital admissions in heart failure, but these findings were not consistent in all reviews. Results for the other conditions were mixed, although no reviews showed evidence of harm. Analysis of the mediating role of self-management, and of components of successful interventions, was limited and inconclusive. More intensive and multifaceted interventions were associated with greater improvements in diabetes, heart failure, and asthma.

Conclusions: While telehealth-mediated self-management was not consistently superior to usual care, none of the reviews reported any negative effects, suggesting that telehealth is a safe option for delivery of self-management support, particularly in

XSL•FO
RenderX

conditions such as heart failure and type 2 diabetes, where the evidence base is more developed. Larger-scale trials of telehealth-supported self-management, based on explicit self-management theory, are needed before the extent to which telehealth technologies may be harnessed to support self-management can be established.

(J Med Internet Res 2017;19(5):e172) doi:10.2196/jmir.6688

KEYWORDS

telehealth; telemonitoring; self-management; chronic disease; diabetes; heart failure; asthma; COPD; pulmonary disease, chronic obstructive; cancer

Introduction

The rising prevalence of long-term conditions is a major clinical and public health challenge [1]. Telehealth has attracted considerable interest as a means of delivering care to those with long-term conditions [2]. Definitions of telehealth are many and varied [3], and the technologies involved are novel and evolving [4]: in this paper we use the term to refer to *any use of information and communication technology to facilitate communication or transfer of information between patient and health care provider over a distance* [5]. Self-management—whereby individuals take on tasks to deal with medical management, role management, or emotional aspects of their condition [6]—is also increasingly recognized as important for effective management of long-term conditions [7-9]. Telehealth has been proposed as one mechanism by which self-management may be promoted and enabled [10], although, in contrast, it has also been suggested that telehealth may sometimes increase dependence on health professionals rather than promoting self-management [11]. The extent to which telehealth effectively promotes self-management, and the components of telehealth interventions that contribute to this goal, remain unclear.

Self-management and its support comprise a wide range of potential activities and interventions [12]. While self-management support is an important aspect of the wider management of a range of long-term conditions, its nature and the approach to supporting successful self-management varies depending on the condition, as well as the individual patient. A systematic overview of self-management interventions (Practical Reviews in Self-Management Support [PRISMS]) demonstrated that self-management support interventions across a range of 14 long-term conditions are complex and multifaceted, involve both the patient and health care professional, and need to be tailored to the individual and their specific condition and context [12]. The review found that no single component of self-management interventions could be identified as being more important than others, but also that the detail and quality of reporting of these complex interventions was a barrier to their wider implementation and the understanding of their effective components [12,13]. This observation, along with the lack of a suitable tool to analyze the important components of self-management interventions, led to the development of the PRISMS taxonomy of self-management support [14]. This taxonomy identified 14 separate components that might be adapted and used to support self-management across a range of long-term conditions. Several of these components could be potentially delivered via telehealth, and may be grouped and considered under the following headings:

- Patient education and information provision
- Remote monitoring with feedback and action plans (eg, peak expiratory flow or blood glucose monitoring with action plans)
- Telehealth-facilitated clinical review
- Adherence support (eg, medication or lifestyle intervention adherence)
- Psychological support
- Lifestyle interventions (eg, smoking cessation, exercise, weight loss)

Given the wide variety of technologies and interventions that telehealth encompasses [2], and the varied nature of self-management interventions [12], we aimed to gain a broad overview of the evidence for telehealth-mediated self-management support using metareview methodology [15]. We focused on 6 specific conditions in which telehealth has been widely used and evaluated as a method of delivering care and in which the principles of self-management are considered an important component of disease management [12,16]. With respect to telehealth-supported self-management, we aimed to (1) assess the impact on disease control and health care utilization, and (2) identify components of self-management support delivered by the telehealth interventions and assess the impact of these components on disease control and the process of self-management.

Methods

This metareview aimed to synthesize systematic review evidence on telehealth interventions to support self-management in diabetes mellitus (types 1 and 2), heart failure, asthma, chronic obstructive pulmonary disease (COPD), and cancer. We carried it out according to a prespecified protocol. We were unable to register our protocol as PROSPERO (an international prospective register of systematic reviews) does not accept protocols for metareviews.

XSL•FO
RenderX

Table 1. Inclusion criteria and database search for systematic reviews of randomized controlled trials (RCTs) of telehealth interventions incorporating components of supported self-management.

Criteria	Description
Population	Adults or children with 1 or more specified long-term conditions (diabetes mellitus type 1 or 2, heart failure, asthma, chronic obstructive pulmonary disease, and cancer).
	Reviews of multiple conditions included if disease-specific findings reported separately.
Intervention	Telehealth[a] interventions to support self-management[b].
Comparator	"Usual care" or alternative means of delivering the intervention (eg, face-to-face, paper-based).
Outcomes	Markers of disease control (see Table 2).
	Unscheduled use of health care services (see Table 2).
	Self-management process outcomes (see Table 2).
Settings	Any health care setting.
Study design	Systematic reviews of RCTs.
	Systematic reviews of multiple study designs included if RCT data reported separately.
Databases	MEDLINE, Embase, CINAHL, PsycINFO, AMED, Web of Science, and Cochrane Database of Systematic Reviews.
Manual searching	Reference lists of all eligible reviews searched.
Forward citations	Performed for all included systematic reviews (using Web of Science).
In-progress studies	Abstract used to identify recently published reviews.
Restrictions	No language restriction applied.
Dates	Initial search: January 2000 to November 2014 (limited to studies later than 2000 due to the relatively recent introduction of the technological solutions and the rapid rate of development of the field. Few studies prior to 2000 were identified in scoping search).
	Update search: May 2016.
	Forward citation search: completed June 2016.
Other exclusions	Less detailed versions of Cochrane reviews published (data taken from the Cochrane review).
	Previous versions of reviews that had been subsequently updated.
	Reviews lacking analyses of quantitative RCT data (narrative or meta-analysis).
	Interventions in which there was no transfer of clinical information between patient and health care provider (eg, peer-to-peer online forums), or where evidence of this was not clear (eg, computer- or Internet-based interventions that gave or recorded information without transfer).

[a]Telehealth was defined as any intervention in which clinical information is transferred remotely between patient and health care provider, regardless of the technology used to record or transmit the information.

[b]Self-management was defined as any intervention that aimed to empower patients to be active decision makers who deal with emotional, social, or medical management of their illness with the aim of improving their independence and quality of life.

Search Strategy

Following an initial scoping, we searched for systematic reviews of randomized controlled trials (RCTs) of telehealth interventions incorporating components of supported self-management. The basic search strategy combined "telehealth terms" AND "self-management terms" AND "long term conditions terms" AND "systematic review terms." The search used a combination of keyword searches and Medical Subject Headings (MeSH). We searched 7 databases for reviews published from January 2000 to May 2016. The search was limited to studies published later than 2000 because telehealth is a relatively recent innovation and rapid advances in technology mean that any earlier work is unlikely to be relevant to contemporary health care. Table 1 summarizes the search strategy and sources, and Table 2 details the outcomes. Multimedia Appendix 1 shows full search terms for the MEDLINE database; we adjusted these for the other databases.

XSL•FO
RenderX

Table 2. Outcomes and definitions.

Outcome	Definition
Markers of disease control	
Nondisease specific	Mortality
	Symptoms and exacerbations: reported symptoms or symptom scores
	Measured rate or frequency of exacerbations
	Other (not disease-specific) biological markers such as blood pressure, lipids
Disease specific	Diabetes: hemoglobin A_{1c}
	Heart failure: body weight, exercise tolerance
	Asthma: PEF[a], FEV_1[b], etc
	Chronic obstructive pulmonary disease: FEV_1
	Cancer: recurrence
Health care utilization	
Health service utilization	Use of health care services (eg, admissions, length of stay, use of unscheduled services or emergency department)
Self-management process outcomes	
Self-efficacy	The confidence that an individual has in their own ability to perform a specific task or behavior
Self-management behaviors	Measures of self-management adoption behavior (eg, use of or adherence to action plan, medication adherence, frequency of monitoring [PEF, blood glucose, etc], avoidance of triggers, use of environmental resources)
Quality of Life	
Quality of life	As assessed by validated tool
Disease-specific quality of life	As assessed by validated quality-of-life assessment tool (eg, asthma quality-of-life score, St George's respiratory questionnaire)

[a]PEF: peak expiratory flow rate.

[b]FEV_1: forced expiratory volume in the first second of expiration.

Inclusion and exclusion criteria are summarized using the population, intervention, comparison, outcomes, setting, and study design headings in Table 1 and Table 2. Initial scoping revealed that few reviews of telehealth interventions identified self-management support explicitly as an aim of the telehealth intervention under consideration (eg, self-management was a specified inclusion criterion). We therefore also included reviews in which self-management support was an implied component or mechanism of the telehealth intervention under consideration and in which outcomes relevant to self-management were assessed as part of the review. For the purpose of inclusion, we defined self-management as "any intervention which aimed to empower patients to be active decision makers who deal with emotional, social or medical management of their illness with the aim of improving their independence and quality of life" [6]. We excluded reviews when the telehealth aspect simply involved remote physiological monitoring (eg, of oxygen saturation or blood sugar levels) without an explicit decision-making role on the part of the patient. Reviews in which patients were educated and supported to interpret and act on the clinical information they were recording we considered self-management and therefore included. We included reviews in which self-management was hypothesized as the mechanism by which the telehealth intervention had an impact, but we considered these to be implied self-management unless self-management support was also a specified aim of the intervention.

Screening of Titles, Abstracts, and Full Texts

Inclusion criteria were piloted by 2 authors (PH and HP) and disagreements were resolved by discussion with all authors. PH then assessed all titles and abstracts against the inclusion criteria. Where no abstract was available, articles were retained for full-text assessment. A random sample of 250 abstracts was screened by 2 reviewers (PH and LD). A kappa statistic of agreement was calculated using IBM SPSS version 22 (IBM Corporation) and was high (0.96). Full texts of all potentially eligible articles were assessed independently by 2 reviewers (PH and LD).

Weighting and Quality Assessment

We assessed the quality of the included reviews using the revised A Measurement Tool to Assess Systematic Reviews (R-AMSTAR) quality assessment tool. Each included review was assessed independently by 2 reviewers (PH and LD) and disagreements were resolved by discussion. We combined the R-AMSTAR score with the size of the review and an assessment of self-management focus to assign a star-based weighting to the evidence from each review. We awarded 1 star for each of the following:

XSL•FO
RenderX

- R-AMSTAR score >30
- >1000 participants (or >10 RCTs if information on participants was not available)
- Explicit self-management focus (ie, self-management support was specified in the inclusion criteria of the systematic review)

We then used the weighting of each review to inform the synthesis. Any disagreements in the full-text screening, quality assessment, or data extraction were resolved by discussion, involving a third author when agreement could not be reached.

Outcomes

We grouped outcomes of interest into disease control outcomes (clinical and physiological markers of disease control, unscheduled health care utilization, and validated measures of symptoms and quality of life) and self-management process outcomes (eg, self-efficacy, medication adherence). These are defined in Table 2.

Data Extraction and Interpretation

Data were extracted using a piloted data extraction template for included studies (PH), and each study was checked for accuracy (LD). We extracted the description of intervention(s) and component(s) of self-management support; inclusion and exclusion criteria; population of interest; duration and intensity of intervention; outcomes measured; and results as presented in the review. The synthesis and conclusions of each review were collated: we did not analyze results from individual RCTs.

The overlap in included RCTs (ie, reviews with similar inclusion criteria may include the same RCTs) precluded meta-analysis of the review findings; we therefore undertook a narrative synthesis. We used harvest plots to illustrate the disease control outcomes related to telehealth-supported self-management components [17,18]. Harvest plots use bars representing individual reviews placed on a plot matrix to indicate whether the review intervention showed an overall positive, negative, or no consistent effect for the outcome in question. To construct the harvest plots, we needed to judge each review as to whether it showed an overall positive effect for each outcome or group of outcomes. Given the heterogeneity in outcomes and methods of data synthesis among the included reviews, and to ensure objective and consistent assessment of each review, we devised a set of rules to underpin decisions about whether a review showed a positive, negative, or no effect. These were devised and refined by discussion between the authors and are described in Table 3. Interpretation of the findings was aided by regular discussion within the research team.

Results

Search Results

The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flowchart shown in Figure 1 illustrates the search results and review selection.

A total of 53 systematic reviews met the inclusion criteria [16,19-70]. These presented data from 231 unique RCTs (119 diabetes, 58 heart failure, 28 asthma, 23 COPD, and 3 cancer). The year of publication ranged from 2000 to 2016. Information on the geographical spread of the RCTs included within the SRs was incomplete, but included studies from North and South America, Europe, Asia, and Oceania. Multimedia Appendix 2 lists the RCTs included in each systematic review.

Study Characteristics

Multimedia Appendix 3 shows details of participant demographics, interventions and comparison, setting, and content and intensity of interventions for each of the included systematic reviews.

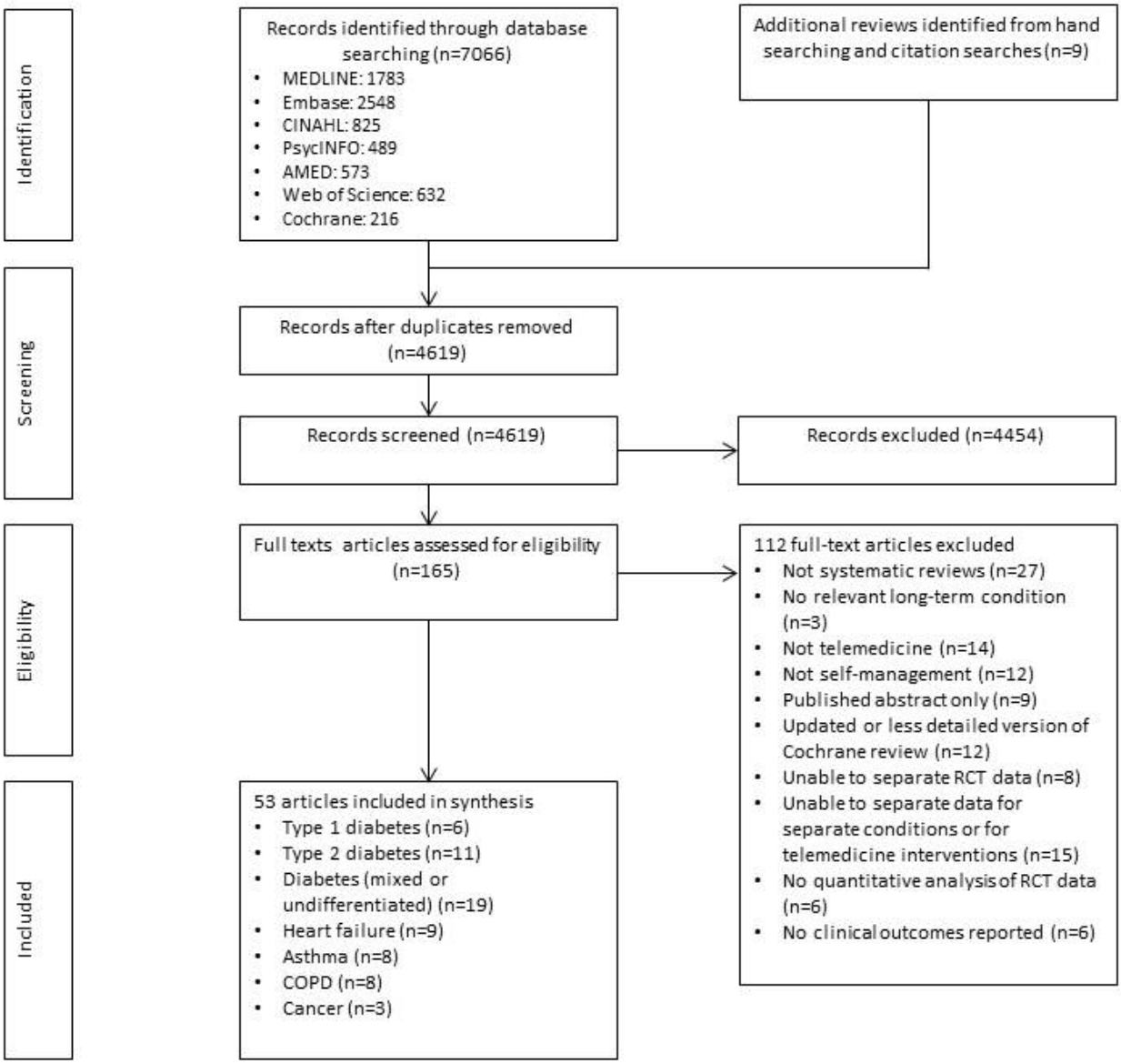
Table 3. Rules for assessment of systematic reviews (SRs) for analysis in harvest plots.

Rule no.	Rules as applied
1	If a review contains a meta-analysis, this result will be used in the harvest plot prioritized over the results of a narrative synthesis.
2	Where a review reports multiple meta-analyses of related outcomes (eg, mean HbA_{1c}[a] concentration; proportion of participants with a normal HbA_{1c}; HbA_{1c} at different time points), and when these outcomes show conflicting results, the result of the SR's primary outcome takes priority as the overall result of the review. Where reviews have no, or more than 1, primary outcome, then the review will be considered as having an overall positive effect if >50% of reported outcomes (or of primary outcomes if multiple) show a positive effect.
3	Where no meta-analysis is available and the review contains a narrative synthesis, overall effect will be judged by the proportion of studies reporting statistically significant positive effects in relevant outcomes. Between 0% and 50% of studies showing positive results will be shown as no consistent effect. Those with >50% of studies showing a positive effect will be shown as positive overall, with those between 50% and 75% hatched to indicate inconsistency. As for meta-analysis in the event of multiple analyses of related outcomes, the result of the SR's defined primary outcome takes priority.
4	If a review reports positive results for an outcome, but it is not clear from the review how many studies in total measured that outcome (ie, no denominator is available), then this outcome will not be included on the harvest plot on grounds of incomplete data. These results will be displayed as reported in the table of review results.

[a]HbA_{1c}: hemoglobin A_{1c}.

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) diagram of literature search. COPD: chronic obstructive pulmonary disease; RCT: randomized controlled trial.



Quality and Weighting of Included Reviews

R-AMSTAR scores for the included systematic reviews ranged from 19 to 43 out of a possible 44. Multimedia Appendix 4 gives scores for the individual components of the R-AMSTAR score. Taking into account quality assessment, explicit self-management focus, and total population size, 8 reviews received an evidence weighting of 3 stars [16,21,25,26,38,45,53,56], 23 received 2 stars [22,24,27-31, 33,37,40,42,44,46,49,50,52,54,55,58,62,63,67,68], 21 received 1 star [19,20,23,32,34-36,39,43,47,48,51,57, 59-61,64-66,69,70], and 1 received no stars [41]. The first column of the table in Multimedia Appendix 5 displays these criteria.

Overview of Presentation of Results

Multimedia Appendix 5 provides an overview of the focus, quality, findings, and conclusions of each of the included reviews. It also displays how the interventions described map to Pearce et al's taxonomy of self-management support [14]. Additional detail is shown in Multimedia Appendix 6. The text that follows synthesizes the findings against the 2 aims of the metareview.

Impact of Telehealth Interventions on Disease Control and Health Care Utilization

Diabetes

A total of 5 reviews focused on type 1 diabetes [23,24,41,48,49] and 1 analyzed type 1 diabetes within a mixed review [19]. The reviews of de Jongh et al [24] and Viana et al [49] were both weighted 2 stars. The former focused on mobile messaging interventions, including medication reminders, and found no improvement in hemoglobin A_{1c} (HbA_{1c}) compared with usual care in a meta-analysis of 2 RCTs [24]. Viana et al meta-analyzed adherence support interventions using various technologies and also found no improvement in HbA_{1c} [49]. Three 1-star narrative syntheses showed improvement in HbA_{1c} limited to only a few RCTs, and all concluded that further evaluation was required [19,23,48]. A meta-analysis scoring no

XSL•FO
RenderX

stars in the weighting system also showed no improvement in glycemic control [41].

A total of 10 reviews [21,29,30,33,42,43,46,52-54], and 1 analysis in a mixed review [19], analyzed HbA_{1c} in type 2 diabetes. Of these, 2 were awarded a 3-star weighting [21,53]. Wu et al meta-analyzed 7 RCTs of telephone follow-up and found no significant improvement in HbA_{1c} compared with usual care in the overall pooled analysis [53]. However, a prespecified subgroup analysis of more intensive targeted interventions (n=3 RCTs) showed a statistically significant benefit on glycemic control. Beatty et al reviewed Web-based lifestyle interventions and found no impact on glycemic control [21]. Of seven 2-star reviews [29,30,33,42,46,52,54], 4 were meta-analyses, all of which showed significant reductions in HbA_{1c} compared with controls, although the effect sizes were small. Interventions included telemonitoring of blood glucose [33,42,54] and mobile messaging [46]. Narrative syntheses found that telehealth-delivered educational interventions [52] and telephone interventions [29] did not improve glycemic control. Greenwood et al analyzed the components of remotely monitored blood glucose and found that multifaceted interventions carried greater benefit (discussed under self-management components) [30]. Two 1-star reviews (1 of mobile telemonitoring and 1 of Web-based telemonitoring) showed improved HbA_{1c} compared with control [19,43].

In total, 19 reviews included both type 1 and 2 diabetes in their analyses, or included RCTs that did not differentiate [16,22,25,27,28,31,32,34-40,44,45,47,50,51]. Of these, 4 were awarded a 3-star weighting [16,25,38,45], of which Farmer et al [25] , Small et al [45], and Liang et al [38] analyzed control or health care utilization outcomes, or both. Farmer et al assessed telemonitoring of self-monitored blood glucose and found no significant improvement in HbA_{1c} in a meta-analysis of 9 RCTs, and either no difference or an increase in health care utilization in 6 RCTs [25]. Liang et al evaluated mobile phone interventions targeting glycemic control and found a significant improvement over usual care in a meta-analysis of 11 RCTs. This effect was more marked for type 2 than for type 1 diabetes [38]. Small et al reviewed telephone-only interventions incorporating “lay health workers” and showed a small but significant improvement in HbA_{1c} over usual care [45]. A total of 8 reviews were weighted 2 stars [22,27,28,31,37,40,44,50]; 4 of these included meta-analyses of impact on HbA_{1c}, with 3 showing significant improvements over controls. Interventions included telemonitoring of blood sugar [40,44] and interactive telehealth excluding telephone support [27]. A meta-analysis of teleconsultations showed no benefit in terms of HbA_{1c} [50]. Narrative syntheses showed mixed results, with modest benefit in telehealth interventions both in place of and supplementing usual care [37], little evidence of benefit from telehealth interventions that excluded telephone support [28], and positive results for telehealth-delivered behavioral interventions (n=13 RCTs) [22]. Three 1-star narrative syntheses suggested improved outcomes with mobile phone interventions [32,35,36], but others showed no overall benefit from teleconsultations [47,51] or telemonitoring [34,39].

Overall, the evidence for diabetes suggests that telehealth-supported self-management interventions for type 2 diabetes may be effective, with evidence for type 1 suggesting no overall benefit in glycemic control. Active self-monitoring of blood glucose data appeared to be most consistently associated with improved outcomes, although this was not consistent across reviews. Evidence for telephone support was more limited, although it may be effective as part of intensive interventions.

Heart Failure

A total of 9 reviews analyzed telehealth interventions for heart failure [16,20,28,55-60]. The highest weighted of these (Ciere et al, 3 stars) did not include any disease control outcomes [56], nor did 3 others [16,59,60]. The highest-weighted meta-analysis was Inglis et al [58]. This review separately analyzed structured telephone support and telemonitoring of physiological parameters. Self-management support was explicitly hypothesized as a mechanism by which these interventions might exert their effect, and both interventions reduced all-cause mortality. A sensitivity analysis of the telephone-based interventions showed no difference between symptom monitoring and education-focused telephone calls. Heart failure hospitalizations, but not all-cause hospital admissions, were also significantly lower with either telemonitoring or telephone support.

In contrast, a 2-star narrative review showed no mortality benefit from telephone-only interventions and a variable effect on hospital admissions [55]. The use of telehealth, without telephone support, showed no impact on health care utilization in a 2-star weighted narrative synthesis of 6 RCTs [28]. A 1-star weighted meta-analysis of telemonitoring for heart failure (excluding telephone interventions) showed a significant reduction in mortality and heart failure admissions, but not all-cause admissions or emergency department visits [57].

Asthma

A total of 8 reviews assessed the impact on asthma control through symptom scores, quality of life, or unscheduled health care [20,24,27,28,36,61-63]. The highest weighted of these each scored 2 stars [24,27,62,63], including 3 narrative syntheses and 1 meta-analysis. McLean et al analyzed a wide range of telehealth interventions, including structured telephone support, education support, telemonitoring, and action plan components [63]. Meta-analyses showed no significant improvement in emergency department use or hospitalization at 3 months, but a significant reduction in hospitalizations compared with usual care at 12 months. Some studies reported improvements in symptom scores; however, most showed no benefit. The authors concluded that benefits were unlikely in mild asthma but that those at higher risk of hospitalization may benefit. de Jongh et al [24] and Marcano Belisario et al [62] reviewed mobile interventions and smartphone apps, respectively. The number of included RCTs was small for both (1 and 2, respectively). Both highlighted the potential benefit from some positive findings in the studies but acknowledged that these findings were not consistent and that the evidence base required development. A 2-star weighted review of multiple conditions included a narrative synthesis of 5 RCTs of interactive telehealth

XSL•FO
RenderX

interventions, excluding telephone-only support, and concluded no overall evidence of benefit in asthma [27]. Three 1-star reviews, all with narrative syntheses, concluded that further evaluation was needed before conclusions could be reached [28,36,61], while 1 other, focused on mobile interventions in developing countries, found evidence for improved symptom scores in a single RCT [20].

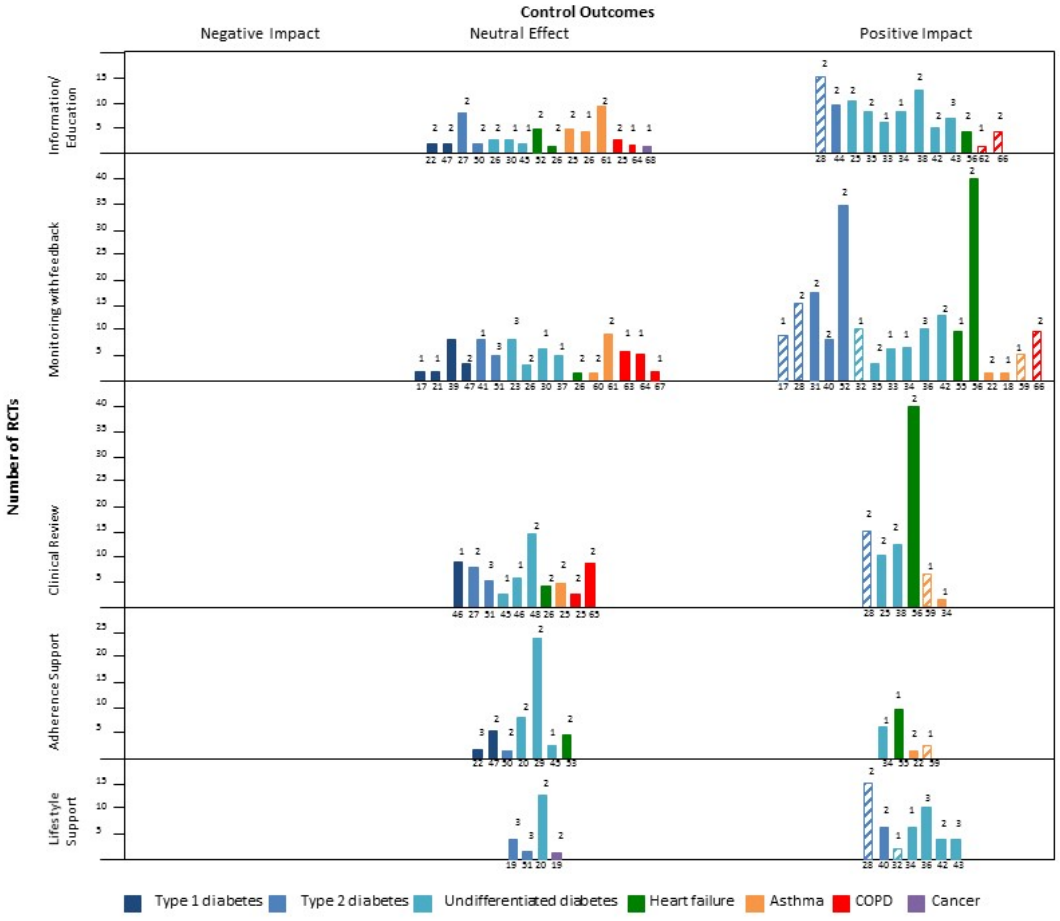
Chronic Obstructive Pulmonary Disease

Of the 8 COPD reviews, 7 analyzed the impact on disease control [27,64-69]. These included 2 meta-analyses, both weighted 2 stars [67,68]. A total of 4 reviews (3 meta-analyses and 1 narrative syntheses) analyzed the impact of telehealth on all-cause mortality, and none showed a significant difference versus usual care [65,66,68,69,71,72]. Both 2-star meta-analyses concerned home-based telehealth using a variety of technologies, incorporating information transmission with personalized feedback. Lundell et al showed a significant improvement in physical activity with telehealth, but no impact on dyspnea [67]. McLean et al found significantly fewer emergency department visits and hospital admissions than with usual care. Findings for quality of life assessed by St George's Respiratory Questionnaire were inconsistent [68]. The authors emphasized that, despite some positive findings, the telehealth interventions were components in complex interventions and further evaluation would be required to clarify their role in COPD [68]. Another meta-analysis, weighted 1 star, showed no impact on mortality but evidence of fewer hospitalizations in a narrative synthesis [69]. Four 1-star weighted narrative syntheses emphasized either a lack [27,64,65] or a low quality [66] of evidence for improved health outcomes.

Cancer

Of the 3 reviews that included cancer RCTs, 2 contained analyses of physical outcomes [21,70]. Both reviews were 2-star weighted. One showed no evidence of improved quality of life or emotional or physical wellbeing in an RCT of moderated Internet-based self-help for breast cancer patients [21]. The other review analyzed Internet-based education programs linking patients with clinicians and found no improvement in quality of life in 2 RCTs and an improvement in symptom scores in 1 of the RCTs [70].

Figure 2. Harvest plot of overall findings of reviews. Number below bar: review reference number. Number above bar: star weighting of review (based on size, revised A Measurement Tool to Assess Systematic Reviews score, and explicit self-management focus). Height of bar: number of randomized controlled trials (RCTs) concerning that self-management component. Block color: consistent effect. Hatched: inconsistent effect (see Table 3). Outcomes assessed were diabetes (hemoglobin A_{1c}), heart failure (mortality, hospital admission), asthma and chronic obstructive pulmonary disease (COPD) (validated symptom or quality of life, scores, physiological measurements), and cancer (validated symptom or quality of life).



XSL•FO
RenderX

Telehealth-Supported Components of Self-Management and Their Impact on Disease Control and the Process of Self-Management

The overall findings of the reviews for impact on disease control outcomes of telehealth-supported components of self-management are illustrated in the harvest plot shown in Figure 2.

All of the interventions in the included systematic reviews were complex interventions with multiple components. Reporting of the details of the components of interventions was highly variable, and we therefore limited analyses of which specific components were associated with improvements in disease outcomes to a subset of the included systematic reviews. No single self-management component was found to be consistently effective, or consistently ineffective.

Education and Information

Supported education and information interventions were particularly effective in the context of diabetes, with 9 larger systematic reviews showing evidence of improvement in HbA_{1c} [27,30,35-37,40,44-46]. Similar interventions in other conditions showed either no evidence of benefit or inconsistent positive effects. The highest weighted of the heart failure reviews (Ciere et al) specifically analyzed whether the beneficial effects of telehealth interventions on clinical outcomes were mediated by increases in knowledge, self-care, or self-efficacy [56]. The authors tested the 2 hypotheses that, first, increased monitoring by health care providers and, second, improved knowledge or self-efficacy leading to improved self-management by patients were mechanisms by which telehealth interventions may be effective. They found that evidence linking telehealth interventions with increased knowledge or self-care behaviors, or linking self-efficacy with self-care was “ambiguous,” and concluded that, on the basis of their findings and the poor methodological quality of the included studies, the evidence neither supports nor refutes their models [56].

Monitoring and Feedback

This was the most commonly described telehealth component in the included systematic reviews and was associated with improved clinical outcomes in diabetes and heart failure [33,35,36,38,42,44,54,58]; however, findings for asthma and COPD were mostly neutral or inconsistent [61-63,65,66,68,69].

Facilitation of Remote Clinical Review

Teleconsultations, videoconferencing, and telephone follow-up designed to review symptoms or clinical course were important aspects of telehealth interventions described by Inglis et al for heart failure, whose meta-analysis showed reduced mortality and heart failure hospital admission with telemonitoring and telephone review [58]. Several diabetes reviews included elements of remote clinical review, with some reporting positive outcomes [27,30,40]. In asthma and COPD, the findings were typically neutral [27,67].

Adherence Support and Lifestyle Interventions

Fewer reviews focused on adherence support or lifestyle interventions, and, of those that did, several did not report disease control outcomes. Findings were more mixed than for other components. Achieving improved clinical outcomes with such interventions may be challenging, as they involve significant behavior change. A 3-star weighted review by Farmer et al analyzed the impact on medication adherence of interventions using remote monitoring, messaging, or a combination [26]. Meta-analysis of 8 interventions showed a nonsignificant effect size; and only 6 of the 15 interventions reported some improvement in medication adherence. The impact of this on glycemic control or other outcomes was not reported. Cassimatis et al (2 stars) analyzed behavioral interventions (excluding telemonitoring) and found some improvements in self-care, dietary and medication adherence, and physical activity, but the effect was inconsistent [22].

Multicomponent and Intensive Interventions

While it was clear that most telehealth interventions were complex multicomponent interventions, most of the included reviews either provided limited description of the interventions or did not specifically analyze the impact of individual components on the efficacy of the intervention as a whole. The only review to address explicitly the question of which components were associated with improved disease outcomes was Greenwood et al, weighted 2 stars [30]. They defined 7 separate components of telehealth (patient education, health care provider education, self-monitoring profile, blood glucose goals, use of blood glucose data to modify behavior, feedback to patients, and 2-way interaction). No interventions were found incorporating all 7 components; however, those including 5 or more were associated with significant improvement in HbA_{1c}. The authors concluded that a range of these components need to be incorporated into telehealth interventions for clinically significant improvements in diabetes self-management to be seen.

While the intensity of the interventions included in many of the reviews varied widely between the included RCTs (Multimedia Appendix 3), few reviews specifically analyzed the relationship between the intensity (in terms of either contact with health care professionals or the complexity or number of components in the intervention) and outcomes. Cassimatis et al highlighted in their analysis that more intensive lifestyle interventions appeared to have a greater impact on glycemic control [22]. Wu et al, who analyzed telephone interventions for type 2 diabetes, found no overall improvement in HbA_{1c} in a meta-analysis, but they did identify a significant improvement in a prespecified subgroup analysis of interventions providing intensive professional support [53]. By contrast, Inglis et al showed no significant impact on mortality of an exploratory subgroup analysis based on intensity of telemonitoring in heart failure [58].

Discussion

Statement of Main Findings

The individual long-term conditions considered in this metareview differed both in the quantity of evidence for telehealth interventions supporting self-management and in the findings and conclusions of the included systematic reviews.

XSL•FO
RenderX

Diabetes and heart failure constituted the greatest evidence base, with available data on cancer being very limited.

The impact of telehealth-supported self-management on disease control and health care utilization was inconsistent, with positive outcomes more frequently identified in type 2 diabetes and heart failure and often no effect demonstrated in other conditions. The highest-quality evidence for heart failure showed an overall improvement in mortality in meta-analyses of telemonitoring and telephone support [58]. In contrast, none of the reviews assessing mortality in COPD showed any significant improvement with telehealth [65,66,68,69]. None of the reviews, however, reported a negative impact of interventions employing telehealth for any condition. This should be treated with some caution, however, as few reviews specifically considered or assessed for publication bias and, of those that did, some found evidence to suggest bias [38,46]. Findings varied by disease group and by components of telehealth delivery. The highest-weighted evidence showed improvement in HbA_{1c} in type 2 diabetes with interventions remotely monitoring blood glucose and in some more intensive telephone interventions [31,38,53]. Physiological telemonitoring and telephone support for heart failure were associated with reduced mortality and heart failure-specific hospital admissions [58]. For both of these conditions, however, findings were inconsistent between reviews, and analyses of similar interventions reached different conclusions [25,55,57]. Interventions for type 1 diabetes did not improve glycemic control [49]. There was some evidence for reduced hospitalization with telehealth interventions in more severe asthma and COPD [63,68], but analyses of more specific, self-management-focused interventions showed insufficient or inconsistent evidence of benefit [27,62]. The interventions described incorporated a range of self-management components. No single component was consistently effective in any disease area, although none were associated with harm. Interventions with multiple components, or more intensive interventions, may be associated with greater benefits [30,53,63,68]. In most reviews, however, the description of self-management components and analysis of their relation to clinical outcomes was not sufficiently detailed to draw firm conclusions about which components or combinations were most beneficial, or in what conditions.

Strengths and Limitations

A strength of metareview methodology is that it allows a relatively rapid synthesis of a large body of primary literature and enables a broad overview of a subject area [15]. There are, however, limitations inherent in metareview methodology. Any systematic review is limited by the time delay from completion of a primary study to publication, and subsequently conducting the review itself. By reviewing systematic reviews, this lag time is further extended and, as such, the results risk being out of date, although we updated the review before completing the paper. The observation in 1 review that some more recent studies showed greater benefits in clinical outcomes [28] highlights the importance of up-to-date evidence in a fast-moving field such as telehealth. That we were unable to differentiate the RCT findings by year of publication is therefore a limitation of our methodology.

By relying on systematic review findings, the evidence is 1 step removed from the empirical evidence, and thus reliant on the interpretation of the review authors. This was mitigated to some extent by assessing the methodological quality of the reviews and using this to weight the evidence. This limitation is most evident, however, when addressing questions about strategies for developing and implementing telehealth interventions. For example, no reviews specifically addressed how potential participants were identified, recruited, and retained, a key issue if telehealth interventions are to be successfully implemented [73]. Few explicitly considered the importance of patient decision making in the interventions. Our consideration of the impact of the intensity of the interventions is also limited to a few systematic reviews, as the majority did not analyze the impact of this on disease control outcomes. By relying on the analysis of the review authors, we were also unable to adjust for important factors such as geographic location, age of participants, and socioeconomic variables such as educational status. A metareview such as this is thus well suited to forming an overview of the topic, but loses granularity and detail of the evidence, a limitation particularly evident when attempting to analyze components of self-management support.

There is also overlap in the included RCTs, which risks overrepresenting the results of a few RCTs included in several systematic reviews and giving a false impression of consensus. Overlap precludes meta-analysis, but the use of harvest plots provides a visual synthesis of the findings of each individual review. We designed the rules used to determine how each review was displayed in the harvest plots to ensure that the assessment of reviews was consistent and transparent. Specifying that more than 50% of RCTs in a narrative synthesis must show positive results for a review to be considered as having consistently positive findings avoids overly optimistic interpretation, but risks overlooking individual RCTs that may be particularly relevant to the review question.

There are also limitations specific to this metareview. This metareview considered only 6 long-term conditions and may have reached different conclusions had we selected a different set of conditions. However, we selected common conditions, 5 of which had a good evidence base for telehealth or self-management, or both, and included 1 area (cancer) in which these concepts are still at an early stage. The relative lack of data for cancer may reflect a lack of research in this field or simply that the available evidence has yet to be synthesized in a systematic review.

Not all the included systematic reviews explicitly focused on self-management. It could be argued that the inclusion of systematic reviews with an implied self-management focus was based on our subjective assessment. However, we developed clear rules for the inclusion criteria and undertook duplicate full-text screening, but this might have resulted in some inappropriate inclusions or exclusions due to limited description of interventions. In addition, we identified reviews that focused specifically on self-management and incorporated this into the weighting system, so as to minimize the impact of this limitation on the interpretation of the findings.

XSL•FO
RenderX

Comparison With Other Literature

Our finding of inconsistent evidence of benefit for disease control and health care utilization with telehealth-supported self-management is similar to the findings reported by other overviews of telehealth interventions that were not specifically focused on self-management. A recent metareview of heart failure telemonitoring interventions showed a reduction in all-cause mortality (relative risk 0.60-0.85) and heart failure hospitalizations (relative risk 0.64-0.86) in an analysis of 15 meta-analyses [74]. An overview of telehealth across a number of conditions noted modest improvements in HbA_{1c} in some reviews with others noting no overall effect [2]. A consistent conclusion, reflected in our findings, is that telehealth is not associated with worse outcomes. Taken together, it appears that, while not consistently superior to usual care, telehealth is a safe alternative mode of delivery for self-management support, particularly in conditions such as heart failure and type 2 diabetes, where the evidence base is more developed. The research agenda may therefore shift to demonstrating equivalence, or understanding the impact of offering choice, or evaluating other potential benefits of telehealth (such as improved access).

While components of the reported interventions mapped to Pearce et al's taxonomy of self-management support [14], few reviews analyzed how these related to the self-management process or clinical outcomes. In common with other overarching analyses of self-management support [12], we found no evidence to support 1 component as being consistently effective or essential for improved outcomes. The greatest volume of evidence was for educational interventions and monitoring of clinical data incorporating feedback, and several reviews, notably those with higher weighting, demonstrated improved outcomes in some (notably type 2 diabetes) but not all conditions (Figure 2). Clinical review, adherence support, and lifestyle interventions also showed some positive results, but with substantial inconsistency. There was some evidence for telehealth-delivered adherence support in asthma and diabetes, but this was variable and inconsistent. In heart failure, evidence linking knowledge, self-efficacy, or self-care behaviors to improvements in disease outcomes was either lacking or inconsistent [56]. Greenwood et al concluded that a range of components (including educational, self-monitoring, goal-driven, and interactive elements) were important in improving glycemic control [30] and emphasized the multifaceted nature of self-management support.

Analysis of the reasons for effectiveness (or lack thereof) of the interventions is limited in the included reviews. Such questions have been addressed using qualitative methodologies [75], which have highlighted that users valued convenience [76], a sense of being "watched over" [11,77], and finding the telemonitoring data reassuring [78]. Some felt empowered by these interventions to use the health care system more effectively [76,77]. Such findings correlate with the quantitative findings of this review showing that interventions with active monitoring of blood glucose data with feedback in type 2 diabetes were often associated with improved outcomes. However, qualitative explorations of some asthma and COPD interventions suggested that daily professional monitoring of telemonitored parameters (such as pulse oximetry or symptom scores) could engender reliance on professionals rather than supporting self-management [11,76,78,79]. In contrast, self-monitoring (including oximetry) without daily professional oversight gave patients greater confidence in following their self-management plans [80].

Conclusions

While telehealth interventions were not consistently found to be superior to usual care, none of the reviews reported any negative effects, suggesting that telehealth is a safe alternative mode of delivery for self-management support, particularly in conditions such as heart failure and type 2 diabetes, where the evidence base is more developed. Improvements may be more readily seen in those with more severe disease. The decision to adopt telehealth strategies will be determined not only by clinical or demographic circumstances but also by patient or clinician preference. Findings vary within and between different conditions, and further investigation is required to establish the role in conditions (such as cancer) where current evidence is limited.

We found little explicit evidence of a mediating role for self-management in telehealth interventions and the specific components that may encourage self-care. While some evidence suggests that, in the context of type 2 diabetes, more intensive interventions may be associated with greater improvements in glycemic control, such observations are limited to specific analyses from a small number of included reviews [22,30,53], and it is not clear whether this applies to other disease areas [58]. Larger-scale trials of self-management interventions delivered by telehealth, based on explicit self-management theory [81,82], linked with process evaluations that explore intermediary outcomes such as self-efficacy, and providing detailed description of the interventions, are needed before the extent to which telehealth technologies may be harnessed to support self-management at scale can be established.

Acknowledgments

Thanks to Marshall Dozier, University of Edinburgh, who advised on and reviewed the search strategy and choice of databases. This study received no specific funding.

Authors' Contributions

PH, HP, BM, CC, and DW devised the study and wrote the protocol. PH and HP piloted and refined the literature search and inclusion criteria. PH performed the literature search. LD independently checked a sample of the literature search. PH and LD

XSL•FO
RenderX

assessed all full-text articles and independently carried out R-AMSTAR quality assessments. PH extracted the data and LD performed a duplicate check of all extracted data. PH performed the analysis. PH wrote the first draft. All authors reviewed and commented on drafts of the manuscript. All authors approved the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms used in MEDLINE.

[PDF File (Adobe PDF File), 30KB - jmir_v19i5e172_app1.pdf]

Multimedia Appendix 2

List of randomized controlled trials in included systematic reviews.

[PDF File (Adobe PDF File), 226KB - jmir_v19i5e172_app2.pdf]

Multimedia Appendix 3

Study characteristics.

[PDF File (Adobe PDF File), 136KB - jmir_v19i5e172_app3.pdf]

Multimedia Appendix 4

A Measurement Tool to Assess Systematic Reviews (revised) quality assessment.

[PDF File (Adobe PDF File), 45KB - jmir_v19i5e172_app4.pdf]

Multimedia Appendix 5

Details of included studies.

[PDF File (Adobe PDF File), 75KB - jmir_v19i5e172_app5.pdf]

Multimedia Appendix 6

Results of included systematic reviews.

[PDF File (Adobe PDF File), 92KB - jmir_v19i5e172_app6.pdf]

References

1. 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. Lancet 2012 Jul 7;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
2. McLean S, Sheikh A, Cresswell K, Nurmatov U, Mukherjee M, Hemmi A, et al. The impact of telehealthcare on the quality and safety of care: a systematic overview. PLoS One 2013;8(8):e71238 [FREE Full text] [doi: 10.1371/journal.pone.0071238] [Medline: 23977001]
3. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
4. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
5. World Health Organization. Report on the Second Global Survey on eHealth. 2010. Telemedicine: opportunities and developments in member states URL: http://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2016-09-21] [WebCite Cache ID 6kgM3OU8G]
6. McGowan PT. Self-management education and support in chronic disease management. Prim Care 2012 Jun;39(2):307-325. [doi: 10.1016/j.pop.2012.03.005] [Medline: 22608868]
7. Department of Health. Our health, our care, our say: a new direction for community services. London, UK: The Stationary Office; 2006. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/272238/6737.pdf [accessed 2016-09-21] [WebCite Cache ID 6kgM6i3A3]

XSL•FO
RenderX

8. Department of Health. A new approach to chronic disease management for the 21st century. 2001. URL: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4018578.pdf [accessed 2016-09-21] [WebCite Cache ID 6kgMBHFXa]
9. World Health Organization. Innovative care for chronic conditions: building blocks for action. Geneva, Switzerland: WHO; 2012. URL: http://www.who.int/chp/knowledge/publications/icccglobalreport.pdf [accessed 2016-09-21] [WebCite Cache ID 6kgMFE2ZD]
10. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012;344:e3874 [FREE Full text] [Medline: 22723612]
11. Fairbrother P, Pinnock H, Hanley J, McCloughan L, Sheikh A, Pagliari C, et al. Exploring telemonitoring and self-management by patients with chronic obstructive pulmonary disease: a qualitative study embedded in a randomized controlled trial. Patient Educ Couns 2013 Dec;93(3):403-410. [doi: 10.1016/j.pec.2013.04.003] [Medline: 23647981]
12. Taylor S, Pinnock H, Epiphaniou E, Pearce G, Parke H, Schwappach A, et al. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS - Practical systematic Review of Self-Management Support for long-term conditions. Southampton (UK); NIHR Journals Library 2014:1-580. [Medline: 25642548]
13. Hoffmann TC, Erueti C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. BMJ 2013;347:f3755 [FREE Full text] [Medline: 24021722]
14. Pearce G, Parke HL, Pinnock H, Epiphaniou E, Bourne CLA, Sheikh A, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr;21(2):73-82. [doi: 10.1177/1355819615602725] [Medline: 26377727]
15. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]
16. 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]
17. Crowther M, Avenell A, MacLennan G, Mowatt G. A further use for the harvest plot: a novel method for the presentation of data synthesis. Res Synthesis Methods 2011;2(2):79-83. [doi: 10.1002/jrsm.37] [Medline: 26061676]
18. Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, et al. The harvest plot: a method for synthesising evidence about the differential effects of interventions. BMC Med Res Methodol 2008 Feb 25;8:8 [FREE Full text] [doi: 10.1186/1471-2288-8-8] [Medline: 18298827]
19. Baron J, McBain H, Newman S. The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. J Diabetes Sci Technol 2012 Sep;6(5):1185-1196 [FREE Full text] [Medline: 23063046]
20. 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]
21. Beatty L, Lambert S. A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. Clin Psychol Rev 2013 Jun;33(4):609-622. [doi: 10.1016/j.cpr.2013.03.004] [Medline: 23603521]
22. Cassimatis M, Kavanagh DJ. Effects of type 2 diabetes behavioural telehealth interventions on glycaemic control and adherence: a systematic review. J Telemed Telecare 2012 Dec;18(8):447-450. [doi: 10.1258/jtt.2012.GTH105] [Medline: 23209266]
23. Currell R, Urquhart C, Wainwright P, Lewis R. Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2000(2):CD002098. [doi: 10.1002/14651858.CD002098] [Medline: 10796678]
24. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012;12:CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
25. Farmer A, Gibson OJ, Tarassenko L, Neil A. A systematic review of telemedicine interventions to support blood glucose self-monitoring in diabetes. Diabet Med 2005 Oct;22(10):1372-1378. [doi: 10.1111/j.1464-5491.2005.01627.x] [Medline: 16176199]
26. Farmer AJ, McSharry J, Rowbotham S, McGowan L, Ricci-Cabello I, French DP. Effects of interventions promoting monitoring of medication use and brief messaging on medication adherence for people with type 2 diabetes: a systematic review of randomized trials. Diabet Med 2016 May;33(5):565-579. [doi: 10.1111/dme.12987] [Medline: 26470750]
27. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015;9:CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]

XSL•FO
RenderX

28. García-Lizana F, Sarría-Santamera A. New technologies for chronic disease management and control: a systematic review. J Telemed Telecare 2007;13(2):62-68. [doi: 10.1258/135763307780096140] [Medline: 17359568]
29. Graziano JA, Gross CR. The effects of isolated telephone interventions on glycemic control in type 2 diabetes: a literature review. ANS Adv Nurs Sci 2009;32(3):E28-E41. [doi: 10.1097/ANS.0b013e3181b0d6d6] [Medline: 19707085]
30. Greenwood DA, Young HM, Quinn CC. Telehealth remote monitoring systematic review: structured self-monitoring of blood glucose and impact on A1C. J Diabetes Sci Technol 2014 Feb 21;8(2):378-389. [doi: 10.1177/1932296813519311] [Medline: 24876591]
31. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
32. 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]
33. Huang Z, Tao H, Meng Q, Jing L. Management of endocrine disease. Effects of telecare intervention on glycemic control in type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. Eur J Endocrinol 2015 Mar;172(3):R93-101 [FREE Full text] [doi: 10.1530/EJE-14-0441] [Medline: 25227131]
34. Jaana M, Paré G. Home telemonitoring of patients with diabetes: a systematic assessment of observed effects. J Eval Clin Pract 2007 Apr;13(2):242-253. [doi: 10.1111/j.1365-2753.2006.00686.x] [Medline: 17378871]
35. Krishna S, Boren SA. Diabetes self-management care via cell phone: a systematic review. J Diabetes Sci Technol 2008 May;2(3):509-517 [FREE Full text] [Medline: 19885219]
36. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
37. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011 Nov;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]
38. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
39. Lieber BA, Taylor B, Appelboom G, Prasad K, Bruce S, Yang A, et al. Meta-analysis of telemonitoring to improve HbA1c levels: promise for stroke survivors. J Clin Neurosci 2015 May;22(5):807-811. [doi: 10.1016/j.jocn.2014.11.009] [Medline: 25791996]
40. Marcolino MS, Maia JX, Alkmim MBM, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]
41. Montori VM, Helgemoe PK, Guyatt GH, Dean DS, Leung TW, Smith SA, et al. Telecare for patients with type 1 diabetes and inadequate glycemic control: a randomized controlled trial and meta-analysis. Diabetes Care 2004 May;27(5):1088-1094. [Medline: 15111526]
42. Health Quality Ontario. Home telemonitoring for type 2 diabetes: an evidence-based analysis. Ont Health Technol Assess Ser 2009;9(24):1-38 [FREE Full text] [Medline: 23074529]
43. Mushcab H, Kernohan WG, Wallace J, Martin S. based remote monitoring systems for self-managing type 2 diabetes: a systematic review. Diabetes Technol Ther 2015 Jul;17(7):498-509. [doi: 10.1089/dia.2014.0296] [Medline: 25830528]
44. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]
45. Small N, Blickem C, Blakeman T, Panagioti M, Chew-Graham CA, Bower P. Telephone based self-management support by 'lay health workers' and 'peer support workers' to prevent and manage vascular diseases: a systematic review and meta-analysis. BMC Health Serv Res 2013;13:533 [FREE Full text] [doi: 10.1186/1472-6963-13-533] [Medline: 24370214]
46. Saffari M, Ghanizadeh G, Koenig HG. Health education via mobile text messaging for glycemic control in adults with type 2 diabetes: a systematic review and meta-analysis. Prim Care Diabetes 2014 Dec;8(4):275-285. [doi: 10.1016/j.pcd.2014.03.004] [Medline: 24793589]
47. Suksomboon N, Poolsup N, Nge YL. Impact of phone call intervention on glycemic control in diabetes patients: a systematic review and meta-analysis of randomized, controlled trials. PLoS One 2014;9(2):e89207 [FREE Full text] [doi: 10.1371/journal.pone.0089207] [Medline: 24586596]
48. Sutcliffe P, Martin S, Sturt J, Powell J, Griffiths F, Adams A, et al. Systematic review of communication technologies to promote access and engagement of young people with diabetes into healthcare. BMC Endocr Disord 2011;11:1 [FREE Full text] [doi: 10.1186/1472-6823-11-1] [Medline: 21210964]
49. Viana LV, Gomes MB, Zajdenverg L, Pavin EJ, Azevedo MJ, Brazilian Type 1 Diabetes Study Group. Interventions to improve patients' compliance with therapies aimed at lowering glycated hemoglobin (HbA1c) in type 1 diabetes: systematic review and meta-analyses of randomized controlled clinical trials of psychological, telecare, and educational interventions. Trials 2016;17:94 [FREE Full text] [doi: 10.1186/s13063-016-1207-6] [Medline: 26888087]

XSL•FO
RenderX

50. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May;4(3):666-684 [FREE Full text] [Medline: 20513335]
51. Verhoeven F, van Gemert-Pijnen GL, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]
52. Wens J, Vermeire E, Hearnshaw H, Lindenmeyer A, Biot Y, Van Royen P. Educational interventions aiming at improving adherence to treatment recommendations in type 2 diabetes: A sub-analysis of a systematic review of randomised controlled trials. Diabetes Res Clin Pract 2008 Mar;79(3):377-388. [doi: 10.1016/j.diabres.2007.06.006] [Medline: 17643546]
53. Wu L, Forbes A, Griffiths P, Milligan P, While A. Telephone follow-up to improve glycaemic control in patients with type 2 diabetes: systematic review and meta-analysis of controlled trials. Diabet Med 2010 Nov;27(11):1217-1225. [doi: 10.1111/j.1464-5491.2010.03113.x] [Medline: 20950378]
54. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]
55. Chaudhry SI, Phillips CO, Stewart SS, Riegel B, Mattera JA, Jerant AF, et al. Telemonitoring for patients with chronic heart failure: a systematic review. J Card Fail 2007 Feb;13(1):56-62 [FREE Full text] [doi: 10.1016/j.cardfail.2006.09.001] [Medline: 17339004]
56. Ciere Y, Cartwright M, Newman SP. A systematic review of the mediating role of knowledge, self-efficacy and self-care behaviour in telehealth patients with heart failure. J Telemed Telecare 2012 Oct;18(7):384-391. [doi: 10.1258/jtt.2012.111009] [Medline: 23019605]
57. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]
58. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015(10):CD007228. [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
59. Radhakrishnan K, Jacelon C. Impact of telehealth on patient self-management of heart failure: a review of literature. J Cardiovasc Nurs 2012;27(1):33-43. [doi: 10.1097/JCN.0b013e318216a6e9] [Medline: 21558862]
60. Schmidt S, Schuchert A, Krieg T, Oeff M. Home telemonitoring in patients with chronic heart failure: a chance to improve patient care? Dtsch Arztebl Int 2010 Feb;107(8):131-138 [FREE Full text] [doi: 10.3238/arztebl.2010.0131] [Medline: 20300221]
61. Jaana M, Paré G, Sicotte C. Home telemonitoring for respiratory conditions: a systematic review. Am J Manag Care 2009 May;15(5):313-320 [FREE Full text] [Medline: 19435399]
62. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013;11:CD010013. [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
63. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma. Cochrane Database Syst Rev 2010 Oct 06(10):CD007717. [doi: 10.1002/14651858.CD007717.pub2] [Medline: 20927763]
64. Bolton CE, Waters CS, Peirce S, Elwyn G, EPSRCMRC Grand Challenge Team. Insufficient evidence of benefit: a systematic review of home telemonitoring for COPD. J Eval Clin Pract 2011 Dec;17(6):1216-1222. [doi: 10.1111/j.1365-2753.2010.01536.x] [Medline: 20846317]
65. Cruz J, Brooks D, Marques A. Home telemonitoring effectiveness in COPD: a systematic review. Int J Clin Pract 2014 Mar;68(3):369-378. [doi: 10.1111/ijcp.12345] [Medline: 24472009]
66. Franek J. Home telehealth for patients with chronic obstructive pulmonary disease (COPD): an evidence-based analysis. Ont Health Technol Assess Ser 2012;12(11):1-58 [FREE Full text] [Medline: 23074421]
67. Lundell S, Holmner A, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015 Jan;109(1):11-26. [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]
68. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2011(7):CD007718. [doi: 10.1002/14651858.CD007718.pub2] [Medline: 21735417]
69. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-127. [doi: 10.1258/jtt.2009.090812] [Medline: 20197355]
70. McAlpine H, Joubert L, Martin-Sanchez F, Merolli M, Drummond KJ. A systematic review of types and efficacy of online interventions for cancer patients. Patient Educ Couns 2015 Mar;98(3):283-295. [doi: 10.1016/j.pec.2014.11.002] [Medline: 25535016]
71. Polisena J. Meta-analysis on COPD. J Telemed Telecare 2012 Jun;18(4):242.
72. McLean SC, McKinstry B. Meta-analysis on COPD: comment on 'Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. by J Polisena et al.'. J Telemed Telecare 2012 Jun;18(4):242; author reply 242. [doi: 10.1258/jtt.2012.110005] [Medline: 22645389]

XSL•FO
RenderX

73. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
74. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
75. Medical Research Council. Developing and evaluating complex interventions: new guidance. Swindon, UK: MRC; 2006. URL: http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/ [accessed 2016-09-21] [WebCite Cache ID 6kgLsaOnL]
76. Hanley J, Ure J, Pagliari C, Sheikh A, McKinstry B. Experiences of patients and professionals participating in the HITS home blood pressure telemonitoring trial: a qualitative study. BMJ Open 2013 May 28;3(5):e002671 [FREE Full text] [doi: 10.1136/bmjopen-2013-002671] [Medline: 23793649]
77. Ure J, Pinnock H, Hanley J, Kidd G, McCall SE, Tarling A, et al. Piloting tele-monitoring in COPD: a mixed methods exploration of issues in design and implementation. Prim Care Respir J 2012 Mar;21(1):57-64 [FREE Full text] [doi: 10.4104/pcrj.2011.00065] [Medline: 21785816]
78. Fairbrother P, Pinnock H, Hanley J, McCloughan L, Sheikh A, Pagliari C, et al. Continuity, but at what cost? The impact of telemonitoring COPD on continuities of care: a qualitative study. Prim Care Respir J 2012 Sep;21(3):322-328 [FREE Full text] [doi: 10.4104/pcrj.2012.00068] [Medline: 22875143]
79. Pinnock H, Slack R, Pagliari C, Price D, Sheikh A. Understanding the potential role of mobile phone-based monitoring on asthma self-management: qualitative study. Clin Exp Allergy 2007 May;37(5):794-802. [doi: 10.1111/j.1365-2222.2007.02708.x] [Medline: 17456228]
80. MacNab M, Lee SH, McCloughan L, Hanley J, McKinstry B, Pinnock H. Oximetry-supported self-management for chronic obstructive pulmonary disease: mixed method feasibility pilot project. BMC Health Serv Res 2015;15:485 [FREE Full text] [doi: 10.1186/s12913-015-1135-2] [Medline: 26503028]
81. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
82. McGowan PT. Self-management education and support in chronic disease management. Prim Care 2012 Jun;39(2):307-325. [doi: 10.1016/j.pop.2012.03.005] [Medline: 22608868]

Abbreviations

COPD: chronic obstructive pulmonary disease
HbA1c: hemoglobin A1c
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
PRISMS: Practical Reviews in Self-Management Support
R-AMSTAR: revised A Measurement Tool to Assess Systematic Reviews
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 26.09.16; peer-reviewed by N Kaufman, M Moy, S Mclean, B Holtz; comments to author 07.01.17; revised version received 22.02.17; accepted 09.03.17; published 17.05.17.

Please cite as:
Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H
Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer
J Med Internet Res 2017;19(5):e172
URL: http://www.jmir.org/2017/5/e172/
doi:10.2196/jmir.6688
PMID:28526671

©Peter Hanlon, Luke Daines, Christine Campbell, Brian McKinstry, David Weller, Hilary Pinnock. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 17.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health-Related Internet Use in People With Cancer: Results From a Cross-Sectional Study in Two Outpatient Clinics in Sweden

Susanne Mattsson[1], RN, MSc; Erik Martin Gustaf Olsson[2], PhD; Birgitta Johansson[3], RN, PhD; Maria Carlsson[1], RN, PhD

[1]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden

[2]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[3]Department of Immunology, Genetics and Pathology, Experimental and Clinical Oncology, Uppsala University, Uppsala, Sweden

Corresponding Author:
Susanne Mattsson, RN, MSc
Department of Public Health and Caring Sciences
Uppsala University
Box 564
Uppsala, 75122
Sweden
Phone: 46 18471 6622
Fax: 46 18471 6622
Email: susanne.mattsson@pubcare.uu.se

Abstract

Background: The access to various forms of support during the disease trajectory is crucial for people with cancer. The provision and use of Internet health services is increasing, and it is important to further investigate the preferences and demographical characteristics of its users. Investigating the uptake and perceived value of Internet health services is a prerequisite to be able to meet the needs in the targeted group.

Objective: The objective of this study was to investigate health-related Internet use among people with cancer.

Methods: The health online support questionnaire (HOSQ), examining the incentives for health-related Internet support use, was administered in two Swedish outpatient hospital clinics. Of the 350 copies of the questionnaire handed out, 285 (81.4%) were returned, answered by persons with cancer who had completed treatment or were under active surveillance or another medical treatment.

Results: A total of 215 (76.2%, 215/282) participants reported Internet use since being diagnosed with cancer. Internet-users were younger (P<.001), more likely to have a partner (P=.03), and had a higher level of education than nonusers (P<.001). The most common health-related activity on the Internet was searching for information (77.2%, 166/215), and users searched significantly more immediately after diagnosis compared with later on (P<.001). Use of My Healthcare Contacts was considered the most valuable Internet activity. Having a university degree (P .001) and being younger in age (P=.01) were associated with a significantly higher frequency of health- related Internet use.

Conclusions: People with cancer turn to the Internet for informational support that enables them to influence their care and to stay in touch with friends and relatives. Demographical differences regarding the uptake of Web-based support remains. This indicates a need for research on how to bridge this digital gap. By learning more about the use of health-related support on the Web among people with cancer, adequate support can be offered and potential strain reduced.

(J Med Internet Res 2017;19(5):e163) doi:10.2196/jmir.6830

KEYWORDS

Oncology; eHealth; support

Introduction

The treatment of cancer has gone through some important changes the past decades and one of them is an increase in outpatient care services coupled with a decrease in inpatient care [1,2]. This implies that patients spend less time in hospital, which may result in a decrease in various kinds of support delivered by health care staff and peer patients. Increased

XSL•FO
RenderX

outpatient care may be beneficial for patients who can spend less time at the hospital and even from an economic perspective, but presents challenges regarding the coordination of cancer care [3]. At the same time, the development toward a more empowered, self-determined, and partaking patient is continuing [4].

Internet-based technologies such as patient portals, websites, and apps managed by health care institutions, have been recognized as a significant lever to improve cancer care coordination [5]. Internet delivered support may also be a tool to increase patient empowerment [6] and has been found as cost effective as well as an important factor in reducing the need for support from the health care system [7].

The use of the Internet as a source of support is a trend that has increased rapidly among cancer patients during the past decade [8]. Motives for using the Internet as a source of support among patients with cancer are ease of communication and access to the most up-to-date information and peer support [9,10]. The Internet offers a wide range of websites delivering different kinds of support. In addition to searching for health information on the Web, people with cancer visit online peer support networks, blogs, and social networks [11,12]. This has been found to be a valuable source of social support [13]. Social support is associated with a better health-related quality of life, fewer stress symptoms, and better health [14-16], and might reduce anxiety and depression symptoms and increase quality of life in people with severe diseases such as cancer [17,18].

The access to the Internet in Europe has increased significantly during the past decades and is high in the Northern countries [19]. However, health-related Internet use is affected by sociodemographic characteristics such as age, gender, education, and civil and socioeconomic status, which should be further addressed in future research [20]. More knowledge is needed regarding cancer patients' current use and appreciation of the Internet as a tool for health-related support, since these patients are generally older which indicate a barrier for Internet use [3,20].

Due to potential long lasting effects of the disease and treatments, people with cancer are at a higher risk for comorbidities and psychosocial problems throughout their life [21]. An understanding of how this heterogeneous group uses the Internet for available health-related support may increase the possibilities to offer adequate interventions of Internet-based support to alleviate potential distress.

This study aims to investigate health-related Internet use and the perceived value of this among people with cancer. A further aim is to describe the incentives for Internet use and to explore associations between Internet use and medical and demographic variables.

Methods

Sample and Procedure

A convenience sample of persons with cancer was included from November 2014 to February 2015. Subjects were consecutively recruited at a scheduled visit to an oncology or urology outpatient clinic at a university hospital in Sweden. The inclusion criteria were an age of 18 years or older and completion of the initial treatment (surgery, chemotherapy, radiotherapy) or currently undergoing either hormone treatment, active surveillance or other medical treatment. This was in order for them to have gained some perspective on how they had used the Internet after being diagnosed. Exclusion criteria were inability to understand Swedish, cognitive impairment, or participation in an ongoing Internet-based intervention (U-CARE) at the hospital that could influence the reported use of support on the Internet [22]. Eligible patients were identified through the clinic visit list, approached, and handed the questionnaire in the waiting room. They were given oral and written information about the study and could choose to complete the questionnaire at the clinic or at home and return it by mail in a prepaid envelope. Consent to participate was implied by completion and return of the questionnaire.

Ethics approval was granted by the regional ethical review board in Uppsala (2013-11-20; diary number 2013/436).

Data Collection

The first question in the questionnaire asked about whether the patients had used the Internet. The group of Internet users was defined as the group who reported that they had used the Internet on a computer, mobile phone or a tablet, at least once or twice since being diagnosed with cancer. Internet users were asked if they had visited information sites; social media sites such as Twitter, Facebook, and Instagram; discussion forums and blogs regarding health, diseases, treatments, lifestyle or similar; and whether they had created their own blogs and/or discussion threads and/or commented on others. The questions were rated from 0=never, to 1=once or twice ever, 2=at least once a year, 3=at least once a month, 4=at least once a week, and 5=daily or almost daily. Subjects were further asked a question about the use of an eHealth service (My Healthcare Contacts) that allows patients to request, cancel, or reschedule appointments; read their medical record; and renew their prescriptions. This question was rated from 0=never, to 1=sometimes, and 2=several times. The questions regarding the frequency of using the apps were asked in relation to both the time immediately after diagnosis, and later on. All patients gave their own definition of how long the time immediately after diagnosis was. How valuable they considered the use of different apps to handle their health situation was, was rated on a scale ranging from 0=not valuable at all, to 10=very valuable.

The incentives for Internet use were investigated using the health online support questionnaire, (HOSQ) [23]. The HOSQ was developed and primarily tested in two Swedish samples (one nonclinical sample and one clinical cancer sample) [23]. The HOSQ is validated in Swedish and consists of 18 questions regarding the incentives for health-related Internet use. They are scored on a 6-point Likert scale describing the frequency of use ranging from 0=not relevant or never, to 5=on a daily basis, and the highest possible total score is 90 (Cronbach alpha=.92). The HOSQ can be divided into two subscales: the reading scale and the interacting scale. The subscales contain 9 questions each and the total possible score is 45 for each subscale (Cronbach alpha: reading=.88; interacting=.77). The initial

XSL•FO
RenderX

validation study of the HOSQ depicted a response pattern revealing expected differences both between the interaction and reading scales and according to age, gender, education, and health problems, hence showed a good face- and construct validity [23].

Demographic data on age, sex, civil status (having a partner or being single), birth country, educational level, diagnosis, time since diagnosis, and cancer treatment (completed, hormone, active surveillance, other) were collected using project-specific questions answered by both Internet users and nonusers.

Statistical Analysis

The statistical analyses were performed using IBM SPSS Statistics (version 20.0).

The level of significance in this study was $P \leq .05$.

Internet use, the perceived value of this, and the incentives for Internet use were analyzed descriptively. Statistical comparisons between groups were conducted with chi-square tests (sex, education, civil status, treatment, and birth country) and *t*-tests (age). Comparisons between the use of apps (information sites, discussion forums, blogs, and so on) immediately after diagnosis and later on were conducted with Wilcoxon signed rank test due to positively skewed data. Correlations between age and the HOSQ total score and the subscales were conducted with Spearman rho. The associations between Internet use and demographic and medical data were analyzed with multiple logistic regressions including age, sex, education, civil status, and completion of treatment as independent variables. This analysis was done on the HOSQ total scale and separately for each subscale. Due to positively skewed data, all HOSQ scales were dichotomized based on the median of the respective scale.

Results

Internet Use and Demographic Characteristics

In total, 350 questionnaires were handed out and 285 (81.4%, 285/350) were answered and returned. Three were excluded due to missing data. Two hundred and fifteen (76.2%, 215/282) of the participants reported Internet use after being diagnosed with cancer. Internet users were younger, more often had a partner, and had higher education compared with nonusers. The most common type of diagnosis was prostate or breast cancer and the median time since diagnosis was 3 years (Table 1).

The Health Online Support Questionnaire (HOSQ) Scores Among the Internet Users

The median score of the HOSQ reading subscale was higher compared with the HOSQ interacting subscale median score (Table 2).

Associations Between Web-Based Support and Demographic and Medical Variables

The multiple regression analysis showed that lower age was significantly associated with higher score on HOSQ total (*P*=.01) and the reading subscale (*P* .001). Having a university degree was significantly associated with a higher score on all three scales (reading: *P*=.01; interacting, *P*=.05; Total: *P* .001) compared with elementary school. Completion of treatment was significantly associated with a lower score on the HOSQ reading subscale (*P*=.05; Table 3).

Incentives for Internet Use

The primary incentives for using the Internet for health-related support was to search for information that could improve the overall health, enable more informed decisions, and to get the best possible health care. Instrumental support such as searching for scheduled appointments, addresses, or phone numbers to health care was also reported by many Internet users. The most reported incentives from the interactive scale was staying in touch with friends and colleagues and reading about other peoples' experiences of a similar situation (see Table 4).

XSL•FO
RenderX

Table 1. Demographic characteristics of Internet users (n=215) and nonusers (n=67).

Demographic and medical characteristics	Internet users, n=215	Nonusers, n=67	*P* value
Age (years): mean (SD), range	60.5 (13.8) 20-84	70 (8.9) 39-90	<.001
Sex[a], n (%)			.44
Male	119 (55.3)	35 (52.2)	
Female	94 (43.7)	32 (47.7)	
Education[a], n (%)			<.001
Elementary school[b]	34 (15.8)	31 (46.3)	
Secondary school	72 (33.5)	21 (31.3)	
University	105 (48.8)	13 (19.4)	
Civil status[a], n (%)			.03
Having a partner	181 (84.2)	48 (71.6)	
Single	31 (14.4)	17 (25.4)	
Birth country: Sweden	189 (87.9)	57 (85.0)	.94
Diagnosis[a], n (%)			
Prostate cancer	69 (32.0)	16 (23.8)	
Breast cancer	34 (15.8)	12 (17.9)	
Gastro-intestinal cancer	15 (6.9)	7 (10.4)	
Malignant melanoma	11 (5.1)	4 (5.9)	
Lung cancer	2 (0.9)	7 (10.4)	
Lymphoma	30 (13.9)	6 (8.9)	
CNS[c]tumor	12 (5.6)	1 (1.5)	
Gynecological cancer	8 (3.7)	1 (1.5)	
Head and neck cancer	2 (0.9)	2 (2.9)	
Sarkoma	9 (4.2)	1 (1.5)	
Hematological cancer	4 (1.9)	0 (0.0)	
Other cancer	6 (2.8)	2 (2.9)	
Time since diagnosis, median, (range), years	3 (1-50)	3 (1-29)	
Treatment condition[a], n (%)			.17
Completed treatment	127 (59.1)	35 (52.2)	
Hormone treatment	37 (17.2)	16 (23.9)	
Active surveillance	18 (8.4)	2 (2.9)	
Other treatment	26 (12.1)	12 (17.9)	

[a]Due to occasional missing data in the questionnaires, the sum of the subgroups may be lower than the corresponding total numbers of individuals.

[b]1-9th grade.

[c]CNS: central nervous system.

XSL•FO
RenderX

Table 2. The median and interquartile range of health online support questionnaire (HOSQ) scores among the Internet users (n=215).

Demographic and medical characteristics	Total (Max: 90)		HOSQ[a] reading (Max: 45)		HOSQ interacting (Max: 45)	
	Median	Interquartile range	Median	Interquartile range	Median	Interquartile range
All	12	19	8	15	2	6.75
Sex						
Male	9	16.25	7	13	1	4
Female	14.5	20	8	14.5	4	10
Education						
Elementary school	7	15	4	8.75	1	4
Secondary school	7	16	6	10	1.50	5
University	16	20	11	14	3	10
Civil status						
Having a partner	12	28.75	8	13.75	2	6
Single	12	24.5	8	16.5	1	8
Completed treatment						
Yes	12	22	7	16	2	7
No	12	17.5	9	14	2.5	7

[a]HOSQ: health online support questionnaire.

Table 3. Multiple regression analyses of the health online support questionnaire (HOSQ) scores and demographic and medical variables in the group of Internet users (n=215).

Demographic and medical variables	Total		HOSQ[a] reading		HOSQ interacting	
	OR[b]	CI	OR	CI	OR	CI
Age, years	0.96	0.94-0.99	.95	0.93-0.98	0.99	0.96-1.01
Sex						
Female (ref)						
Male	0.59	0.41-1.67	1.14	0.57-2.30	0.69	0.35-1.36
Education						
Elementary school (ref)						
Secondary school (1)	1.58	0.55-4.54	1.71	0.59-4.91	1.58	0.61-4.11
University (2)	5.37	1.98-14.55	4.74	1.74-12.9	2.75	1.12-6.77
Civil status						
Having a partner (ref)						
Single	0.99	0.40-2.43	1.07	0.43-2.66	0.59	0.25-1.38
Completed treatment						
Yes (ref)						
No	1.74	0.85-3.54	2.03	1.01-4.11	1.11	0.57-2.15

[a]HOSQ: health online support questionnaire.

[b]OR: odds ratio.

XSL•FO
RenderX

Table 4. The frequency of participants reporting never, once or twice ever, or more than sometimes on the items respectively of the health online support questionnaire.

Item	Never n (%)	Once or twice ever n (%)	More than sometimes n (%)
"Since I was diagnosed with cancer I have used the Internet..."			
To search for information that can improve my overall health (R[a])	60 (27.9)	63 (29.3)	81 (37.7)
To search for scheduled appointments, addresses or phone numbers to health care providers (R)	76 (35.3	59 (27.4)	65 (30.2)
To be able to make more informed decisions regarding my illness or health condition (R)	78 (35.3)	59 (27.4)	64 (29.8)
To search for the very latest research regarding my health situation (R)	84 (39.1)	51 (23.7)	65 (30.2)
To search for information so I can better understand physicians and other health care personnel (R)	92 (42.8)	57 (26.5)	52 (24.2)
To search for information from various sources so I can get the best possible health care (R)	98 (45.6)	44 (20.5)	66 (30.7)
To seek further information when I feel worried (R)	98 (45.6)	41 (19.1)	61 (28.4)
To read about other people's experience of a particular illness or health condition or treatment (R)	106 (49.3)	44 (20.5)	50 (23.3)
To find out whether symptoms I have discovered are dangerous or not (R)	107 (49.8)	46 (21.4)	47 (21.9)
To keep friends and relatives informed about how I'm feeling (I[b])	111 (51.6)	36 (16.7)	53 (24.7)
To stay in touch with friends and colleagues when I'm sick or not feeling well (I)	114 (53.1)	25 (11.6)	59 (27.4)
To get feedback from friends and relatives on how I'm handling my illness or health situation (I)	143 (66.5)	28 (13.0)	27 (12.6)
To get feedback from people who have or have had the same health problem as I have (I)	150 (69.8)	28 (13.0)	20 (9.3)
To talk about a treatment for an illness or health condition that I've been through (I)	152 (70.7)	20 (9.3)	27 (26.5)
To share practical advice and suggestions about illness or health (I)	154 (71.6)	23 (10.7)	22 (10.2)
To express my opinion regarding health or illness or care (I)	165 (76.7)	17 (9.9)	16 (7.4)
To look for compassion when I'm not feeling well (I)	170 (79.1)	13 (6.1)	16 (7.4)

[a]R: reading scale.

[b]I: interacting scale.

The Use of Apps on the Internet

Among the Internet users, 166 (77.2%, 166/215) reported that they had used the Internet to search for health information. Forty five (20.9%, 45/215) had visited health-related discussion forums, 39 (18.1%, 39/215) had visited blogs, only one (0.5%, 1/215) had taken part in psychological treatment on the Internet, 73 (33.9%, 73/215) reported use of social media. My Healthcare Contacts had been used by 82 (38.1%, 82/215) and 38 (17.7%, 38/215) of them had taken part of information and test results in the medical e-record services, 8 (3.7%, 8/215) had scheduled appointments, 12 (5.6%, 12/215) had renewed prescriptions, and 5 (2.3%, 5/215) had chosen a general practitioner (GP). My Healthcare Contacts was considered the most valuable Internet health resource (mean=6.7; standard deviation [SD]=3) compared with information sites (mean=6.2; SD=2.6), forums (mean=6.1; SD=2.9), blogs (mean=5.8; SD=3.2), and social media (mean=4.1; SD=3.3).

The daily use of information apps immediately after diagnosis was higher (21.9%, 47/215) than later on (5.1%, 11/215), $P<.001$. One hundred and twelve (52.1%, 112/215) considered the length of the time immediately after diagnosis to be somewhere between 1 day and 3 months. The total range of reports was 1 day to 5 years and the median time was 3 months. No differences were found regarding use of other apps than information between the time immediately after the diagnosis and thereafter. The vast majority of users visiting information sites (86%) and discussion forums (81%) did so regarding their own health. Use of the other apps was less related to their own health (blogs 58%; social media 28%).

Discussion

Principal Findings

This study found that the persons with cancer who use the Internet for health-related support mainly search for information that enables them to improve their overall health and get the

XSL•FO
RenderX

best possible health care. The health care delivered tool My Healthcare Contacts was considered very valuable among the ones who used it, hence, seems to be a satisfactory app on the Web. Younger and higher educated used the Internet significantly more than the older and less educated.

Information sites were the most frequented sites compared with social media, discussion forums, and blogs. This finding converges with other studies on health-related Internet use [9,20,24]. Reported reasons in previous studies are perceived lack of information from health care, the efficiency of the Internet, and as also found in this study, a desire to stay updated on the most recent information about disease-related matters [5,8,24]. It is well known that the need for information is associated with health-related variables [11]. This study found that the search for information was significantly higher during the time immediately after diagnosis than later. Hence, it is of major importance for the health care system to provide patients with adequate informational support during this phase. As this study shows, information found on the Internet is considered valuable. It has been reported that well-informed patients report greater engagement in care decisions and an increased confidence in their interactions with health care providers [24].

Peer support is often referred to as an important factor for people with cancer, even though it is not considered as one of the most important health activities on the Internet [3]. The use of blogs and discussion forums was not that high in this study even though the development of these apps as well as their use has increased over the past decade [25]. The relatively low frequency, in this study, of visiting these sites could be explained by the relatively high age in this group since older patients are less likely to use Web-based tools [3]. The score on the HOSQ interacting scale was lower than the score on the reading scale, which corroborates other studies reporting that taking part of information rather than also sharing information is more common [26,27]. It could also be that this group prefers face-to-face support since only a third of the Internet users reported use of social media. By addressing cancer patients' preferences and providing customized support in navigating the Internet, which has been found necessary, the Internet might become a significant source of peer support for this group.

The demographic factors found to predict Internet use in this study were young in age and having a university education. The variables that frequently appear to influence health-related Internet use are age, gender, educational level, perceived health, and socioeconomic and civil status [3,20,28]. According to the unified theory of acceptance and use of technology (UTAUT), there are factors such as gender and age that mediate actual usage of technology [29]. In this study, still being under treatment or active surveillance predicted a higher frequency of reading on the Internet. Other research has shown that there is also a need for support after a patient have finished treatment and are in clinical remission, at which point the contact with the health care decreases [30]

One third of the Internet users reported use of My Healthcare Contacts, where taking part of information and finding out test results in the medical e-record services, was the most frequent activity. This was the app that was reported as the most valuable. There are findings suggesting that the majority of cancer patients consider improved access to their medical records as something that should be prioritized [3] Scheduling appointments, renewing prescriptions, and choosing one's GP was also reported, which can be highly efficient for patients as well as the health care system in reducing the workload.

Despite just over 20 years of Internet, the number of people using it is still expanding in Sweden. The access as well as Internet use is still increasing even though the vast majority of people in Sweden already use the Internet. In particular, the group of younger pensioners using the Internet is growing. Currently, almost 100% of people between the ages of 12-55 years, and approximately 90% of people aged 60-65 years use the Internet in Sweden [31]. Future patients with cancer will be more active Internet users compared with today's patients. This puts demands on the health care system regarding the development of relevant Web apps to meet their needs.

Strengths and Limitations

The response rate was high (81%) and 75% of the participants were Internet users. To avoid selection bias the questionnaire was handed out at the clinic instead of being administered in a Web-based context to Internet users only. This way we were able to compare demographic and disease-related variables between the Internet users and the nonusers. The heterogeneity regarding diagnosis, age, educational level, and sex was satisfactory. Thus, the results may be fairly representative for people with cancer but limited to patients that have completed their initial treatment. However, there is a need for diagnose specific studies as well.

Many of the patients who chose not to answer the questionnaire said that they did not use the Internet, hence, did not think it was relevant for them to respond. Therefore, the percentage of Internet users in this study may be higher than what is representative for the population examined.

The patients' definition of the time immediately after the diagnosis varied greatly. Therefore, results showing that the need for information was significantly higher immediately after diagnosis compared with later should be interpreted with caution, since the time "immediately after" the diagnosis might overlap the time "later on."

A recall bias among the Internet users should also be taken into consideration since the median time since diagnosis is 3 years, meaning that the frequencies of Internet use may be under- or overestimated, which decreases the reliability of the findings.

The questionnaire used in this study, the HOSQ, has not been used previously, so the validity and reliability are uncertain. However, it has been psychometrically tested and validated regarding face and content validity in two samples where the response pattern revealed expected differences both between the interaction and reading scales and according to age, gender, education, and health problems [23].

According to UTAUT [29], there are factors mediating actual usage of technology that has not been collected in this study. It should therefore be taken into consideration that there may be

XSL•FO
RenderX

other variables than the one collected that may have an impact on the results in this study.

Conclusions

This study has found that patients turn to the Internet primarily for informational support that enable them to improve their overall health, make more informed decisions, and to get the best possible health care. Also to stay in touch with friends and colleagues and take part of other peoples' experiences of a similar situation. Instrumental support such as searching for scheduled appointments, phone numbers, and addresses to health care was also something that they used the Internet for. The perceived value of the apps examined in this study was generally high. The use of instrumental support such as My Healthcare Contacts was considered the most valuable activity on the Web. Internet use was associated with having a university degree and being younger. This may indicate that the threshold for health-related activities on the Web is higher for older and less educated individuals, which needs to be addressed in future research.

Practice Implications

A better understanding of health-related Internet use in different groups is a prerequisite to the provision of adequate Internet delivered support. By learning more about the incentives for health-related Internet use in various contexts of people with cancer, we may be able to develop tailored support that may alleviate potential distress, save costs, and reduce health care workload.

Acknowledgments

We are grateful to all the patients who took part in the survey. Also to the staff at the urology- and oncology department who made it possible for us to administer the questionnaire. This research was funded by a grant to the U-CARE program, hosted by the group Clinical psychology in health care, PI Louise von Essen. MC participated in the work within her employment at Uppsala University. BJ was partly funded by the U-CARE and partly within her employment at Uppsala University.

Authors' Contributions

SM contributed to the study design and drafted the manuscript. EO, MC, and BJ contributed to the study design and helped to draft the manuscript.

Conflicts of Interest

None declared.

References

1. American Society of Clinical Oncology. The state of cancer care in America, 2014: a report by the American Society of Clinical Oncology. J Oncol Pract 2014 Mar;10(2):119-142. [doi: 10.1200/JOP.2014.001386] [Medline: 24618075]
2. Jönsson B, Hofmarcher T, Lindgren P, Wilking N. The cost and burden of cancer in the European Union 1995-2014. Eur J Cancer 2016 Oct;66:162-170. [doi: 10.1016/j.ejca.2016.06.022] [Medline: 27589247]
3. Girault A, Ferrua M, Lalloué B, Sicotte C, Fourcade A, Yatim F, et al. Internet-based technologies to improve cancer care coordination: current use and attitudes among cancer patients. Eur J Cancer 2015 Mar;51(4):551-557. [doi: 10.1016/j.ejca.2014.12.001] [Medline: 25661828]
4. Notisum. 2014. Svensk författningssamling: Patientlagen (2014:821) URL: http://www.notisum.se/rnp/sls/sfs/20140821.pdf [accessed 2017-05-08] [WebCite Cache ID 6qIvY5W3j]
5. Clauser SB, Wagner EH, Aiello BE, 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]
6. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten Wim H. Empowerment of cancer survivors through information technology: an integrative review. J Med Internet Res 2015;17(11):e270 [FREE Full text] [doi: 10.2196/jmir.4818] [Medline: 26614438]
7. Elbert NJ, van OH, van RW, Ekeland AG, Hakkaart-van RL, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
8. Carlsson ME. Cancer patients seeking information from sources outside the health care system: change over a decade. Eur J Oncol Nurs 2009 Sep;13(4):304-305. [doi: 10.1016/j.ejon.2009.03.005] [Medline: 19362053]
9. Yli-Uotila T, Rantanen A, Suominen T. Motives of cancer patients for using the internet to seek social support. Eur J Cancer Care (Engl) 2013 Mar;22(2):261-271. [doi: 10.1111/ecc.12025] [Medline: 23320398]
10. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]

XSL•FO
RenderX

11. Beckjord EB, Arora NK, McLaughlin W, Oakley-Girvan I, Hamilton AS, Hesse BW. Health-related information needs in a large and diverse sample of adult cancer survivors: implications for cancer care. J Cancer Surviv 2008 Sep;2(3):179-189. [doi: 10.1007/s11764-008-0055-0] [Medline: 18792791]
12. Fox S, Purcell K. Pew Research Center. 2010. Chronic disease and the internet URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2017-05-08] [WebCite Cache ID 6qIvktBd9]
13. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel patient community on social media. BMC Res Notes 2012;5:699 [FREE Full text] [doi: 10.1186/1756-0500-5-699] [Medline: 23270426]
14. Giordano GN, Björk J, Lindström M. Social capital and self-rated health--a study of temporal (causal) relationships. Soc Sci Med 2012 Jul;75(2):340-348. [doi: 10.1016/j.socscimed.2012.03.011] [Medline: 22537483]
15. Fiorillo D, Sabatini F. Quality and quantity: the role of social interactions in self-reported individual health. Soc Sci Med 2011 Dec;73(11):1644-1652. [doi: 10.1016/j.socscimed.2011.09.007] [Medline: 22001229]
16. Gerich J. Effects of social networks on health from a stress theoretical perspective. Soc Indic Res 2013;118(1):349-364.
17. Klemm P. Effects of online support group format (moderated vs peer-led) on depressive symptoms and extent of participation in women with breast cancer. Comput Inform Nurs 2012 Jan;30(1):9-18. [doi: 10.1097/NCN.0b013e3182343efa] [Medline: 22240564]
18. Pinar G, Okdem S, Buyukgonenc L, Ayhan A. The relationship between social support and the level of anxiety, depression, and quality of life of Turkish women with gynecologic cancer. Cancer Nurs 2012;35(3):229-235. [doi: 10.1097/NCC.0b013e31822c47bd] [Medline: 21946902]
19. Europa. Digital economy and society URL: http://ec.europa.eu/eurostat/web/digital-economy-and-society [accessed 2017-04-24] [WebCite Cache ID 6pxiFROiT]
20. 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]
21. Elliott J, Fallows A, Staetsky L, Smith PW, Foster CL, Maher EJ, et al. The health and well-being of cancer survivors in the UK: findings from a population-based survey. Br J Cancer 2011 Nov 8;105 Suppl 1:S11-S20 [FREE Full text] [doi: 10.1038/bjc.2011.418] [Medline: 22048028]
22. Mattsson S, Alfonsson S, Carlsson M, Nygren P, Olsson E, Johansson B. U-CARE: Internet-based stepped care with interactive support and cognitive behavioral therapy for reduction of anxiety and depressive symptoms in cancer--a clinical trial protocol. BMC Cancer 2013;13:414 [FREE Full text] [doi: 10.1186/1471-2407-13-414] [Medline: 24024826]
23. Mattsson S, Olsson EM, Alfonsson S, Johansson B, Carlsson M. Measuring use of health-related support on the internet: development of the health online support questionnaire (HOSQ). J Med Internet Res 2015 Nov 20;17(11):e266 [FREE Full text] [doi: 10.2196/jmir.4425] [Medline: 26589638]
24. Dolce MC. The Internet as a source of health information: experiences of cancer survivors and caregivers with healthcare providers. Oncol Nurs Forum 2011 May;38(3):353-359. [doi: 10.1188/11.ONF.353-359] [Medline: 21531685]
25. Coiera E. Social networks, social media, and social diseases. BMJ 2013;346:f3007. [Medline: 23697672]
26. Magnezi R, Bergman YS, Grosberg D. Online activity and participation in treatment affects the perceived efficacy of social health networks among patients with chronic illness. J Med Internet Res 2014;16(1):e12 [FREE Full text] [doi: 10.2196/jmir.2630] [Medline: 24413148]
27. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
28. 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]
29. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478.
30. Garrett K, Okuyama S, Jones W, Barnes D, Tran Z, Spencer L, et al. Bridging the transition from cancer patient to survivor: pilot study results of the Cancer Survivor Telephone Education and Personal Support (C-STEPS) program. Patient Educ Couns 2013 Aug;92(2):266-272 [FREE Full text] [doi: 10.1016/j.pec.2013.04.002] [Medline: 23647980]
31. Findahl O. Svenskarna och internet (The Swedes and the Internet 2014). Stockholm: Stiftelsen för internet infrastruktur (The Institute of Internet Infrastructure); 2014.

Abbreviations

HOSQ: health online support questionnaire

XSL•FO
RenderX

Edited by G Eysenbach; submitted 16.10.16; peer-reviewed by JL Lopez-Guerra, J Apolinário-Hagen, M Cox, E Børøsund; comments to author 29.11.16; revised version received 18.02.17; accepted 24.03.17; published 15.05.17.

Please cite as:
Mattsson S, Olsson EMG, Johansson B, Carlsson M
Health-Related Internet Use in People With Cancer: Results From a Cross-Sectional Study in Two Outpatient Clinics in Sweden
J Med Internet Res 2017;19(5):e163
URL: http://www.jmir.org/2017/5/e163/
doi:10.2196/jmir.6830
PMID:28506959

©Susanne Mattsson, Erik Martin Gustaf Olsson, Birgitta Johansson, Maria Carlsson. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Diversity in Older Adults' Use of the Internet: Identifying Subgroups Through Latent Class Analysis

Leonieke C van Boekel[1], PhD; Sebastiaan TM Peek[1], MSc; Katrien G Luijkx[1], PhD

Tilburg School of Social and Behavioral Sciences, Department Tranzo, Tilburg University, Tilburg, Netherlands

Abstract

Background: As for all individuals, the Internet is important in the everyday life of older adults. Research on older adults' use of the Internet has merely focused on users versus nonusers and consequences of Internet use and nonuse. Older adults are a heterogeneous group, which may implicate that their use of the Internet is diverse as well. Older adults can use the Internet for different activities, and this usage can be of influence on benefits the Internet can have for them.

Objective: The aim of this paper was to describe the diversity or heterogeneity in the activities for which older adults use the Internet and determine whether diversity is related to social or health-related variables.

Methods: We used data of a national representative Internet panel in the Netherlands. Panel members aged 65 years and older and who have access to and use the Internet were selected (N=1418). We conducted a latent class analysis based on the Internet activities that panel members reported to spend time on. Second, we described the identified clusters with descriptive statistics and compared the clusters using analysis of variance (ANOVA) and chi-square tests.

Results: Four clusters were distinguished. Cluster 1 was labeled as the "practical users" (36.88%, n=523). These respondents mainly used the Internet for practical and financial purposes such as searching for information, comparing products, and banking. Respondents in Cluster 2, the "minimizers" (32.23%, n=457), reported lowest frequency on most Internet activities, are older (mean age 73 years), and spent the smallest time on the Internet. Cluster 3 was labeled as the "maximizers" (17.77%, n=252); these respondents used the Internet for various activities, spent most time on the Internet, and were relatively younger (mean age below 70 years). Respondents in Cluster 4, the "social users," mainly used the Internet for social and leisure-related activities such as gaming and social network sites. The identified clusters significantly differed in age (P<.001, ω^2=0.07), time spent on the Internet (P<.001, ω^2=0.12), and frequency of downloading apps (P<.001, ω^2=0.14), with medium to large effect sizes. Social and health-related variables were significantly different between the clusters, except social and emotional loneliness. However, effect sizes were small. The minimizers scored significantly lower on psychological well-being, instrumental activities of daily living (iADL), and experienced health compared with the practical users and maximizers.

Conclusions: Older adults are a diverse group in terms of their activities on the Internet. This underlines the importance to look beyond use versus nonuse when studying older adults' Internet use. The clusters we have identified in this study can help tailor the development and deployment of eHealth intervention to specific segments of the older population.

(J Med Internet Res 2017;19(5):e180) doi:10.2196/jmir.6853

KEYWORDS

Internet; aged; cluster analysis; health

Introduction

In Western societies, Internet use is widespread and is increasingly important in diverse aspects of everyday life. For instance, Internet is indispensable in communication; access to news and information; and administrative applications such as applying for allowance, tax declaration, or Internet banking. Age is known to be strongly related to the likelihood that individuals use the Internet. In the Netherlands, among adults aged 65 years and older, 77.8% have access to the Internet, compared with 94% of Internet users among the whole population [1]. In the United States, Internet use among older adults is lower compared with the Dutch older adults, namely, 64% in 2016 [2]. Compared with other European countries, Internet access and using the Internet for Internet banking and social media is higher among older adults in the Netherlands [3]. Internet use among older adults in the Netherlands has increased in the last decade, although 50% of adults aged 75 years or older have never used the Internet [1]. Knowledge and insight in the Internet use of older adults is of importance since the Internet may be related to self-management, such as instrumental activities of daily living (iADL) [4]. The Internet has the potential to enhance social capital among older adults, for instance expanding or maintaining social contacts, decreasing loneliness [5], and enlarging access to information [6]. In addition, the Internet may be used for communication with

XSL•FO
RenderX

health professionals or informal care givers, which also enhances self-care.

The digital divide framework [7,8] suggests a societal gap between individuals who use and who do not use the Internet. Groups that are traditionally more disadvantaged in socioeconomic sense also appear to be at a greater likelihood to not having access to the Internet (eg, [9,10]). However, this framework focuses on a comparison of users and nonusers of the Internet, and it does not account for the diversity in use of the Internet. Conversely, the usage gap suggests that the potential benefit of the Internet may be related to the activities for which the Internet is being used. For instance, Internet activities related to information, career, or education offer Internet users more chances and resources than Internet activities which are limited to entertainment purposes (eg, [7,11,12]). Higher educated people more actively use the Internet, and they mainly use the Internet for searching information whereas lower educated people use the Internet more frequently for entrainment purposes [13]. In addition, age is also related to usage of the Internet; older adults more often use the Internet for a shorter period of time compared with younger persons and are less likely to use the Internet for activities such as email and Web-based shopping [9]. Compared with younger persons, older adults gain fewer benefits from using the Internet [14]. These findings underline the to investigate the purposes and activities older adults use the Internet for. Until now, research has merely focused on comparing older adults' use versus nonuse of the Internet (eg, [9,15,16]), overlooking the activities and purposes for which the Internet is being used. Older adults are a heterogeneous group, and the Internet use may be very diverse among this group. Understanding the diversity of the use of Internet use among older adults provides insight in the potential benefits of using the Internet for different subgroups of older adults.

The objective of this study was to identify and describe the diversity in older adults' activities on the Internet and whether this diversity is related to social and health-related variables. The following research questions were formulated: (1) Which subgroups or clusters can be identified among older adults based on their Internet activities? (2) What are the features of these subgroups and how do they differ in their Internet activities, time spent on the Internet, and demographic variables? (3) Is there a difference between the subgroups concerning social and emotional loneliness, psychological health, and activities of daily living (ADL) of older adults?

Methods

Recruitment

We used data collected by an existing Internet panel that is representative of the Dutch population, namely, longitudinal Internet studies for social sciences (LISS) panel. This panel is administered by CentERdata, a Dutch research institute specialized in data collection. Panel members receive questionnaires every month and completed questionnaires are rewarded. The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Households are invited to participate in the panel, and people without an appropriate computer or Internet connection are provided equipment, insuring a representative sample. The LISS panel consists of 4500 households with approximately 7000 individuals.

For this study, data from 2 different questionnaires that are annually completed by LISS panel members were combined, namely, the LISS core studies "social integration and leisure" (data collection in October and November 2015) and "health" (data collection in July and August 2015). Demographic information such as age, gender, and marital status were measured in November 2015. Panel members aged 65 years and older were selected if they completed the "social integration and leisure" questionnaire that included questions regarding Internet use (N=1608). In addition, respondents were included in the analyses when they reported to have access and use the Internet, which was the majority 88.18% (1418/1608).

Measures

All measures were taken from annual core studies among the LISS panel members and therefore, questions were developed and tested by CentERdata. The core study, "social integration and leisure," provided information regarding Internet use of the respondents. Web-based activities were assessed by 17 dichotomous items (never or ever spend time on this particular functionality or Web-based activity). Web-based activities included financially related activities (eg, "comparing products and searching product information," "Internet banking"), functional and more traditional activities ("emailing," "searching for information"), and social and leisure-related Internet activities (eg, "reading and viewing social media," "playing Internet or Web-based games"). The amount of time spent on the Internet was asked by the following items: "Can you indicate how many hours you use the Internet on a computer or laptop/tablet/smartphone per week, on average (including emailing), besides when completing questionnaires of this panel? These items were added up to an overall amount of hours using the Internet. Answers ranged between 0 and 175 h per week. Since 175 h per week is an obvious outlier, we categorized the answers into 7 categories ranging from 1: ≤5 h, 2: 5-10 h, 3: 10-15 h...7: ≥30 h per week. Frequency of downloading apps was assessed by the items: "How often do you download apps on your tablet / smartphone" (1=never to 7=almost every day). We calculated mean scores based on both items with higher scores representing more frequent downloading apps.

Social and emotional loneliness was measured with the 6-item version of the loneliness scale of de Jong Gierveld [17]. Two examples of items are: "I experience a general sense of emptiness" and "There are many people I can trust completely" (answers: yes, no, more, or less). Overall scores for social as well as emotional loneliness were calculated ranging from 0-3, and high scores indicated more loneliness. The Mental Health Inventory 5 [18] was used to measure psychological well-being. This scale asks respondents how they felt in the last 4 weeks, for example, did you feel "calm and peaceful" or "depressed and gloomy" (6-point scale never—continuously). Sum scores were calculated in which higher scores indicated better psychological well-being [19]. ADL and iADL were measured with 2 standardized and frequently used questionnaires (eg,

XSL•FO
RenderX

[20,21]) that assess limitations in general daily activities due to health problems. ADL was measured with 7 items on a 5-point scale (without any trouble—not at all): "Can you indicate, for each activity, whether you can perform it, dressing and undressing including shoes and socks / walking across the room / bathing or showering / eating, such as cutting you food into small bits / getting in and out of bed / using the toilet, including sitting down and standing up, reading a map to find your way in an unfamiliar area." iADL was measured with 6 items on the same question and scale: preparing a hot meal / shopping telephoning / taking medicines / performing housekeeping work or maintaining the garden / taking care of financial affairs such as paving bills and keeping track of expenditure. Higher scores represent having more problems with performing iADL and ADL. Finally, demographic information that was collected included: age, gender, marital status, education level, urbanization of place of residence, and ethnicity.

Statistical Analysis: Latent Class Analysis

The first step in the analyses was performing a latent class analysis (LCA) to identify underlying structure of the categorical data about Internet use among older people. LCA is a statistical and probabilistic method that can be used to classify individuals from a heterogeneous group into smaller more homogenous unobserved subgroups [22]. LCA was performed using the program Latent GOLD Choice 5.0 [23]. Variables included in the LCA were Web-based activities mentioned by at least 15% of the study population in order to avoid inclusion of activities that were rarely done. The Web-based activities (dichotomous yes/no) included in the LCA were (1) email; (2) searching for information; (3) searching for and comparing products or product information; (4) purchasing items; (5) Internet banking; (6) reading Web-based news and magazines; (7) reading and viewing social media (eg, Facebook, Instagram, Twitter, YouTube, LinkedIn, Google+, Pinterest, Flickr, or similar services); (8) chatting or video calling or sending messages via social media such as Instagram or Skype; (9) playing Internet games or Web-based gaming); (10) watching Web-based films or TV programs; (11) newsgroups; and (12) posting messages, photos, and short films on social media yourself (eg, Facebook, Instagram, Twitter, YouTube, LinkedIn, Google+, Pinterest, Flickr, or similar services). Model fit indices were used to select the latent model and number of clusters that is not too complex, yet also had a good fit with the data. Bayesian information criterion (BIC) and the Akaike's information criterion 3 (AIC3) were used, that are both relative indicators of model fit. Lower values indicate better fit of the model to the data. Classification error represents the chance that a respondent is assigned to the wrong cluster and should be ideally around 10%. In case bivariate residuals were high, the assumption of local independence may be violated. Therefore, direct effects between variables were added post-hoc to the model one by one in order to reach a solution in which bivariate residuals are ≤10.

Statistical Analysis: Describing and Comparing Identified Clusters

The second step in the analyses was describing and comparing the identified clusters. SPSS version 22 was used to conduct these analyses. A probability level of $P \leq .01$ was used. Descriptive analyses were carried out to describe the different clusters on Internet use variables and demographic variables. Chi-square test (categorical variables) and analysis of variances (ANOVAs) (continuous variables) were conducted to compare the identified clusters on Internet use variables, demographic variables, and social and emotional loneliness, psychological health, ADL, and iADL. For the ANOVA, the robust Welch *F* test was used in case the assumption of homogeneity of variances was violated. To compare the clusters pairwise, the following post-hoc tests were used: Bonferroni and Games-Howell post-hoc tests, in case assumption of homogeneity of variance was violated. Omega squared (ω^2) was calculated as effect size estimate, as ω^2 is robust when one of the assumptions is being violated [24]. Interpretation of ω^2 was as follows: $\omega^2 \leq 0.06$ small effect, $\omega^2 > 0.06 < 0.14$ medium effect, and $\omega^2 \geq 0.14$ large effect.

Results

Description of Study Sample

In total, N=1418 respondents were included for the analyses who were individuals aged 65 years and older and using the Internet. Of these respondents, 52.82% were men (749/1418). Mean age of the respondents was 71.8 (standard deviation, SD 5.7). Of the selected sample, 8.04% (114/1418) was provided with equipment from LISS panel to be able to fill in the questionnaires monthly. Table 1 shows detailed background information of the study sample.

XSL•FO
RenderX

Table 1. Background data of the study sample (N=1418).

Variable	n (%)
Age (mean 71.79, SD 5.68, range 65-93)	1418 (100.00)
Gender	1418 (100.00)
Men	749 (52.82)
Women	669 (47.18)
Marital status	1418 (100.00)
Married	931 (65.66)
Separated	4 (0.28)
Divorced	173 (12.20)
Widow or widower	235 (16.57)
Never been married	75 (5.29)
Level of education[a]	1416 (99.86)
Low education	618 (43.64)
Middle education	346 (24.44)
High education	452 (31.92)
Ethnicity	1409 (99.37)
Dutch background	1240 (88.01)
First generation foreign, Western background	61 (4.33)
First generation foreign, non-Western background	20 (1.42)
Second generation foreign, Western background	82 (5.82)
Second generation foreign, non-Western background	6 (0.43)
Urbanization of place of residence	1413 (99.65)
Extremely urban	160 (11.32)
Very urban	378 (26.75)
Moderately urban	310 (21.94)
Slightly urban	357 (25.27)
Not urban	208 (14.72)

[a]Low education refers to primary education or prevocational secondary education. Middle education refers to preuniversity education or secondary vocational education. High education refers to higher professional education or university education.

Results Latent Class Analysis

We compared the model fit indices, number of parameters, and classification error for models ranging from 1-8 clusters (see Table 2). The 4-cluster model was chosen as the most appropriate model since BIC (16602.92) was lowest and also AIC3 (16385.81) was low. The classification error was appropriate for the 4-cluster model (0.15). We included 4 direct effects in the model in order to decrease bivariate residuals: (1) newsgroups—reading Web-based news and magazines, (2) searching for information—email, (3) product information—searching for information, (4) reading Web-based news and magazines—watching Web-based films or TV programs. In this model with direct effects, bivariate residuals were all below 7. The entrophy R^2 of the 4-cluster model with direct effects was 0.71.

XSL•FO
RenderX

Table 2. Results of the latent class analysis (N=1418).

Model	LL[a]	BIC[b] (LL)	AIC3[c] (LL)	# parameters	Classification error
1-cluster	−9040.46	18168.00	18116.91	12	0
2-cluster	−8380.13	16941.69	16835.27	25	0.08
3-cluster	−8182.31	16640.38	16478.61	38	0.13
4-cluster	−8116.41	16602.92	16385.81	51	0.15
5-cluster	−8076.15	16616.74	16344.29	64	0.16
6-cluster	−8040.57	16639.93	16312.14	77	0.19
7-cluster	−8014.25	16681.63	16298.50	90	0.20
4-cluster with direct effects[d]	−8039.69	16478.51	16244.37	55	0.15

[a]LL: Log likelihood.

[b]BIC: Bayesian information criterion.

[c]AIC3: Akaike's information criterion 3.

[d]Four direct effects were included in the model based on bivariate residuals, namely (1) newsgroups—reading Web-based news and magazines, (2) searching for information—email, (3) product information—searching for information, and (4) reading Web-based news and magazines—watching Web-based films or TV programs.

Results Describing and Comparing Identified Clusters

Cluster 1 included 36.88% (523/1418) of the respondents, and these respondents can be described as the "practical users." The majority of respondents in this cluster used the Internet for functional and financially related activities such as "comparing products and searching product information," "purchasing items," and "Internet banking." In addition, "email" and "searching for information" was a frequently mentioned activity for which these practical users used the Internet. Among the practical users, the amount of men was high compared with the other clusters (65.0%, 340/523, P<.001) although effect size was small (Cramer V=.21). Cluster 2 comprised 32.23% (457/1418) of the respondents and can be labeled as the "minimizers." Respondents in this group spent the lowest amount of hours per week on the Internet compared with the other clusters (see Tables 4 and 5) and reported the lowest frequencies of Internet activities compared with the other clusters (see Table 3). The minimizers mainly used the Internet for traditional purposes such as "email" and "searching for information" although the frequency of these activities were lowest compared with the other clusters. The mean age of the minimizers (mean 73.8, SD 6.3) was significantly higher compared with the other clusters with a medium effect size (Welch F=36.7, P<.001, ω^2=0.07). Respondents in cluster 3 were labeled as the "maximizers" and included 17.77% (252/1418) of the respondents. The diversity of reported Internet activities in this cluster was high; and for many Internet activities, the maximizers reported the highest frequency of an Internet activity of all clusters, for instance, "watching Web-based films or TV programs," "downloading software," "reading Web-based news or magazines," and "chatting, video calling, sending messages." The mean age of the maximizers was lower compared with other clusters (mean 69.6, SD 4.4) and they spent the highest amount of hours per week on the Internet compared with the other clusters with a medium effect size (mean 3.4, SD 2.1, Welch F=63.3, P<.001, ω^2=0.12). The maximizers downloaded apps on their devices significantly more often compared with the other clusters with a large effect size (Welch F=81.2, P<.001, ω^2=0.14). Among the maximizers, the amount of men and women as well as different education levels is quite equally distributed. In sum, the maximizers more frequently used the Internet and used the Internet for a great diversity of Internet activities. Finally, cluster 4 comprised 13.12% (186/1418) of the respondents and can be described as the "social users." The social users mainly used the Internet for social and leisure-related Internet activities. For instance, "reading and viewing social media," "playing Internet or Web-based games," and "posting messages or photos or short films on social media" was frequently mentioned as Internet activity by social users. The amount of women is high among the social users (63.98%, 119/186). The social users are most comparable with the practical users (cluster 1) in terms of age, time spent on the Internet, and frequency of downloading apps (see Tables 4 and 5). The majority of the social users, as was the case among the minimizers, were respondents with lower education (respectively 54.84%, 102/186 and 56.24% 257/457).

ANOVA tests were carried out to compare the 4 clusters on social and health-related variables, namely, social and emotional loneliness, psychological well-being, ADL, and iADL. Overall, no big differences were found in social and health-related variables between the identified clusters since effects sizes were all rather small (see Tables 4 and 5). Nevertheless, the practical users reported significant higher psychological well-being compared with the minimizers (practical users: mean 79.9, SD 13.6; minimizers: mean 76.7, SD 15.4). Additionally, in iADL, significant differences were found between the minimizers on the one hand (mean 9.2, SD 3.4) and the maximizers and practical users on the other hand (practical users: mean 8.3, SD 2.3; maximizers: mean 8.2, SD 2.1). This indicated that the minimizers had more problems with iADL compared with the practical users and the maximizers. Finally, the minimizers scored significantly lower in their experienced health compared with the maximizers (minimizers: mean 2.8, SD 0.7; maximizers: mean 3.0, SD 0.7).

XSL•FO
RenderX

Table 3. Frequency (%) of respondents ever spending time on an Internet activity per cluster.

Internet activity	Practical users n=523	Minimizers n=457	Maximizers n=252	Social users n=186	Chi-square *P* value	Cramer *V*
Email	99.6	83.2	100	98.9	<.001	.33
Searching for information	98.5	79.0	98.0	91.4	<.001	.31
Comparing products or product information	94.8	33.9	100	53.2	<.001	.64
Purchasing items	81.8	8.5	100	14.0	<.001	.79
Watching Web-based films or TV programs	15.1	4.6	38.5	17.2	<.001	.31
Downloading software or music or films[a]	15.1	2.8	28.2	10.2	<.001	.26
Internet banking	98.1	46.4	95.6	68.8	<.001	.55
Playing Internet or Web-based games	20.8	19.3	40.5	52.7	<.001	.27
Reading Web-based news or magazines	55.3	19.7	73.0	48.9	<.001	.39
Newsgroups	18.4	10.1	29.8	24.7	<.001	.18
Reading and viewing social media	23.5	8.5	99.6	93.5	<.001	.77
Reading or writing blogs[a]	7.3	1.8	21.8	14.0	<.001	.25
Posting messages or photos or short films on social media	1.0	2.4	59.9	57.5	<.001	.67
Chatting or video calling or sending messages	33.3	5.9	80.6	52.7	<.001	.55
Dating websites[a]	1.5	0.9	2.8	3.8	.05	.07
Visiting forums and communities[a]	3.3	0.9	11.9	3.8	<.001	.19
Other activities	15.1	5.9	32.1	14.0	<.001	.25

[a]Not included in the latent class analysis because frequency of activity mentioned by <15% of the respondents.

Table 4. Comparison (chi-square tests) of the identified clusters on demographic variables.

Demographic variables	Practical users	Minimizers	Maximizers	Social users	X^2	*P* value	Cramer *V*
Gender, n (%)					63.4	<.001	.21
Men	340 (65.0)	206 (45.1)	136 (54.0)	67 (36.0)			
Women	183 (35.0)	251 (54.9)	116 (46.0)	119 (64.0)			
Marital status, n (%)					27.4	.007	.08
Married	348 (66.5)	301 (65.9)	170 (67.5)	112 (60.2)			
Separated	1 (0.2)	-	2 (0.8)	1 (0.5)			
Divorced	64 (12.2)	37 (8.1)	39 (15.5)	33 (17.7)			
Widow or widower	81 (15.5)	93 (20.4)	29 (11.5)	32 (17.2)			
Never married	29 (5.6)	26 (5.7)	12 (4.8)	8 (4.3)			
Level of education, n (%)					90.0	<.001	.18
Low education	187 (35.8)	257 (56.2)	72 (28.6)	102 (54.8)			
Middle education	121 (23.1)	96 (21.0)	83 (32.9)	46 (24.7)			
High education	215 (41.1)	104 (22.8)	96 (38.1)	37 (19.9)			

XSL•FO
RenderX

Table 5. Comparison (analysis of variances) of the identified clusters on age, Internet variables, and social and health-related variables.

Variables, mean (SD[a])	Practical users[1f]	Minimizers[2f]	Maximizers[3f]	Social users[4f]	Welch F or F ratio (df[b])	P value	ω^2
Age	71.3 (5.3)[2,3]	73.8 (6.3)[1,3,4]	69.6 (4.4)[1,2,4]	71.1 (4.9)[2,3]	36.7[e] (3610)	<.001	0.07
Amount of hours spend on Internet per week	2.4 (1.6)[2,3]	1.6 (1.3)[1,3,4]	3.4 (2.1)[1,2,4]	2.5 (1.7)[2,3]	63.3[e] (3550)	<.001	0.12
Frequency downloading apps	1.7 (1.8)[2,3]	0.7 (1.4)[1,3,4]	2.6 (1.7)[1,2,4]	1.5 (1.8)[2,3]	87.2[e] (3567)	<.001	0.14
Psychological well-being	79.9 (13.6)[2]	76.7 (15.4)[1]	78.7 (14.6)	76.3 (15.5)	5.0[e] (3558)	.002	0.01
Emotional loneliness	0.5 (0.9)	0.5 (0.9)	0.5 (0.9)	0.6 (1.0)	1.7[e] (3576)	.17	0.00
Social loneliness	1.1 (1.2)	1.0 (1.2)	1.0 (1.12)	1.0 (1.1)	0.1 (31,411)	.96	0.00
ADL[c]	6.8 (1.9)	7.1 (2.2)	6.7 (1.8)	6.8 (1. 6)	3.9[e] (3591)	.009	0.01
iADL[d]	8.3 (2.3)[2]	9.2 (3.4)[1,3]	8.2 (2.1)[2]	8.8 (2.4)	9.5[e] (3588)	<.001	0.02
Experienced health	2.9 (0.7)	2.8 (0.7)[3]	3.0 (0.7)[2]	2.9 (0.7)	4.0 (31,383)	.007	0.01

[a]SD: standard deviation.

[b]df: degrees of freedom.

[c]ADL: activities of daily living.

[d]iADL: instrumental activities of daily living.

[e]Welch F test and Games-Howell post hoc test were used since for these variables assumption of homogeneity of variances were violated.

[f]The superscript numbers 1-4 indicate significant differences (<.01) between the clusters on Bonferroni and Games-Howell posthoc test.

Discussion

Principal Findings

The results of this study show that older adults are a diverse group concerning their activities on the Internet. We identified 4 clusters of older adults based on the activities for which they use the Internet. First, the minimizers are the oldest respondents (mean age 74 years old) and spend the least time on the Internet. The minimizers report the lowest frequency on most of the Internet activities and mainly use the Internet for traditional purposes such as email. Second, on the other end of the spectrum are the maximizers, who are relatively young (mean age below 70 years old), spend the most time on the Internet, and are reported to spend time on almost all of the Internet activities. Among the maximizers, the amount of men and women, as well as the different education levels, are equally distributed. The third and fourth clusters are in between the minimizers and maximizers: the practical users and social users. These clusters score in between the maximizers and minimizers regarding both time spent on the Internet and their age (mean age 71 years old). The practical users, in contrast with the social users, use the Internet mainly for financial and practical matters such as searching for information, comparing products, and Internet banking. As opposed to practical users, the social users mainly use the Internet for social and leisure related activities (social media, games, etc). The amount of men is higher among the practical users and the amount of women higher among the social users.

The clusters did not differ to a large extent in social and health-related variables. However, the minimizers reported lower psychological well-being compared with the practical users, more problems with iADL in comparison with the maximizers and practical users, and lower experienced health compared with the maximizers. In sum, it appeared that the minimizers show a somewhat lower health, but this cluster also comprised the oldest respondents (mean age 74 years old). As causality between the variables is unclear, it is unknown whether age causes older adults to be only minimally active on the Internet or that a lower health status causes lower Internet activity.

It has been established that physical and mental limitations may form a barrier for older adults to use computers and the Internet [25]. In our study sample, the majority of older adults belonged to the practical users and the minimizers. It would be worthwhile to replicate this study in the upcoming years to study whether the distribution over the clusters will change. Possibly the amount of maximizers will increase considering the fact that a new generation will be more active in using the Internet. Longitudinal research may provide insight in whether older adults shift from one cluster to the other when they become older. In addition, it is of interest to further investigate the association between health (decline), age, and older adults' Internet activities. In particular, previous research has suggested that aging processes may lead older adults to consciously or unconsciously limit the number of Web-based activities they engage in [26]. This is in line with Baltes and Baltes' [27] concept of "selection" that describes how seniors, who are confronted with more options than their internal and external resources can handle, are forced to concentrate their energy on a subset of those options. Further research is needed to comprehend why older adults would limit or expand their number of Web-based activities.

XSL•FO
RenderX

Comparison With Prior Work

One study [28] also investigated differences for which adults aged 65 years and older use the Internet, although only four Internet activities were included. In line with our findings it was found that the oldest respondents, aged 80 years and older, mainly use the Internet for practical purposes categorized as email or texting [28]. As mentioned before, most research has been focused on comparing older adults who use and who do not use the Internet. However, patterns of Internet use among the general population have been investigated more extensively. As mentioned earlier, previous findings have indicated that lower educated people use the Internet more often for entertainment-related purposes [13]. This was in line with our findings—among the minimizers and the social users education level was somewhat lower. One study found indications that digital literacy, that is, using the Internet and email, was related to lower iADL impairment [29]. The results of this study found more iADL impairments among the minimizers compared with the maximizers and practical users. Furthermore, we found that in general the oldest respondents, above 73 years old, spend a smaller amount of time on the Internet compared with younger respondents (between 65 and 70 years old) which was in line with prior work [9]. In the same study it was found that being older was related to being less active on the Internet [9]. One of the reasons for this is possibly the fact that older adults are from a different technological generation and therefore find it more difficult to use nowadays technological devices [30]. In this study we also found a significant age differences between the different clusters based on older adult's activities on the Internet. Nevertheless, age differences were rather small and the study sample consisted only of adults aged 65 years and older by which a comparison with younger respondents was not part of the study. Therefore, longitudinal research and replication of this study is recommended to study whether results are equal in the upcoming years when the younger respondents (below 70 years old) become the older ones.

Limitations

To our knowledge this is the first study that identifies clusters of older adults based on the activities for which older adults use the Internet. The large sample size strengthens the findings of this study. We strongly recommend other studies to consider using LISS panel data since the quality of the data is excellent. The Dutch population is considered to be comparable to other Western populations in terms of Internet use; therefore, we expect that the findings of this study apply to a large extent to other Western populations. Nevertheless, attention should be paid to the following limitations. Data of two surveys were combined and data collection took place on two different moments within a time span of 3 months. We are of the opinion that the variables included in this study are quite stable and are not expected to fluctuate to a large extent in a period of 3 months. The information on which the LCA was based was limited to dichotomous information whether respondents ever spend time on a particular Internet activity or not. We did not have information about the time spent on each of the Internet activities, nor did we have information about attitudes of the respondents with regard to Internet use. In addition, no information was available about support that older adults receive in using technologies which is known to be related to older adults' use of technologies [31,32]. Finally, due to the cross-sectional design it was impossible to investigate causality between the included variables.

Practical Implications

In the Netherlands, considerable emphasis is placed on increasing the use of eHealth, in particular among older adults and patient with chronic illnesses [33]. However, research shows that awareness of eHealth services among Dutch users of primary care is rather low and could be improved [33,34]. The identified clusters in this study can be of use in increasing the awareness and use of eHealth interventions among older adults in two ways. First, our results provide information on which channels can be employed to raise awareness of eHealth applications. Choosing the right channels to reach older adults is of importance since our results show that older adults are only on the Web a couple of hours per week. Raising awareness through social media campaigns is likely to be effective for reaching older adults belonging to the cluster social users and maximizers. In contrast, practical users may be reached by advertising campaigns on Web-based shops or product comparison websites. Minimizers may be reached by advertising in search engines, but offline methods also need to be considers for this particular cluster. Second, the identified clusters are an indication of Internet experience among older adults. This information is useful in choosing and designing effective components of eHealth interventions. For instance, practical users and minimizers have less experience in engaging in Web-based discussions than social users and maximizers. In case an eHealth intervention contains a social discussion component, it should be taken into account that particular older adults are familiar with social media interfaces whereas other are not familiar with this. Another important difference between the clusters is the fact that practical users and maximizers are more experiences in purchasing on the Web compared with the other two clusters. This can have implications for determining the range and type of payment options of eHealth services for older adults. Finally, knowledge about the amount of older adults that engage in different Internet activities may be useful for health care organizations in their marketing strategies.

Conclusions

The findings of this study establish that older adults are a diverse group in terms of their activities on the Internet. This underlines the importance to look beyond use versus nonuse when investigating older adults' Internet use. The heterogeneity in activities for which older adults use the Internet is widespread and is vital to consider when attempting to stimulate or facilitate Internet use among older adults. The clusters we have identified in this study can be useful in creating awareness of eHealth interventions among specific segments of the older population.

XSL•FO
RenderX

Acknowledgments

In this paper we used data from the longitudinal Internet studies for the social sciences (LISS) panel administered by CentERdata (Tilburg University, The Netherlands). The authors would like to thank CentERdata for providing the data.

Conflicts of Interest

None declared.

References

1. Centraal Bureau voor de Statistiek. 2016. ICT-gebruik van huishoudens en personen (ICT) (ICT use among households and individuals) URL: https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/ict-gebruik-van-huishoudens-en-personen--ict-- [accessed 2017-05-10] [WebCite Cache ID 6qM09xAoU]
2. Pew Research Center. 2017 Feb 24. Internet/Broadband fact sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2017-05-10] [WebCite Cache ID 6qM0DZU9Q]
3. Europa. Luxembourg: Eurostat; 2014. Proportion of the population using the internet, by age, 2014 URL: http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Proportion_of_the_population_using_the_internet,_by_age,_2014_(%25)_PF15.png [accessed 2017-05-10] [WebCite Cache ID 6qM0HEExj]
4. Melrose RJ, Brommelhoff JA, Narvaez T, Natta L, Riskin-Jones HH, Sakhai S, et al. The use of Information and Communication technology when completing instrumental activities of daily living. Comput Human Behav 2016 Oct;63:471-479. [doi: 10.1016/j.chb.2016.05.045]
5. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
6. Blit-Cohen E, Litwin H. Elder participation in cyberspace: a qualitative analysis of Israeli retirees. J Aging Stud 2004 Nov;18(4):385-398. [doi: 10.1016/j.jaging.2004.06.007]
7. van Dijk JA. The deepening divide. In: Inequality in the Information Society. London: Sage Publications, Inc; 2005.
8. Kuttan A, Peters L. From Digital Divide to Digital Opportunity. Lanham, MD: Scarecrow Press; 2003.
9. van Deursen AJ, Helsper EJ. A nuanced understanding of Internet use and non-use among the elderly. Eur J Commun 2015 Apr;30(2):171-187. [doi: 10.1177/0267323115578059]
10. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
11. Hargittai E, Hinnant A. Digital inequality: differences in young adults' use of the internet. Comm Res 2008 Aug 04;35(5):602-621. [doi: 10.1177/0093650208321782]
12. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. New Media & Soc 2014 May;16(3):507-526. [doi: 10.1177/1461444813487959]
13. Bonfadelli H. The internet and knowledge gaps: a theoretical and empirical investigation. Eur J Commun 2002 Mar;17(1):65-84. [doi: 10.1177/0267323102017001607]
14. van Deursen AJ, Helsper EJ. The third-level digital divide: who benefits most from being online? In: Robinson L, Cotten SR, Schulz J, Hale TM, Williams A, editors. Communication and Information Technologies Annual (Studies in Media and Communications, Volume 10). Bingley: Emerald Group Publishing; 2015:29-52.
15. Wang J, Bennett K, Probst J. Subdividing the digital divide: differences in internet access and use among rural residents with medical limitations. J Med Internet Res 2011 Mar 03;13(1):e25 [FREE Full text] [doi: 10.2196/jmir.1534] [Medline: 21371989]
16. Gracia E, Herrero J. Internet use and self-rated health among older people: a national survey. J Med Internet Res 2009 Dec 02;11(4):e49 [FREE Full text] [doi: 10.2196/jmir.1311] [Medline: 19955041]
17. de Jong Gierveld J, van Tilburg T. A 6-item scale for overall, emotional, and social loneliness: confirmatory tests on survey data. Res Aging 2006 Sep 01;28(5):582-598. [doi: 10.1177/0164027506289723]
18. Berwick DM, Murphy JM, Goldman PA, Ware JE, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. Med Care 1991 Feb;29(2):169-176. [Medline: 1994148]
19. van der Zee K, Sanderman R, Heyink J. De psychometrische kwaliteiten van de mos 36-item Short Form Health Survey (SF-36) in een Nederlandse populatie. Tijdschrift voor Sociale Gezondheidszorg 1993;71:183-191.
20. Hardy SE, Dubin JA, Holford TR, Gill TM. Transitions between states of disability and independence among older persons. Am J Epidemiol 2005 Mar 15;161(6):575-584. [doi: 10.1093/aje/kwi083] [Medline: 15746474]
21. Netuveli G, Wiggins RD, Hildon Z, Montgomery SM, Blane D. Quality of life at older ages: evidence from the English longitudinal study of aging (wave 1). J Epidemiol Community Health 2006 Apr;60(4):357-363 [FREE Full text] [doi: 10.1136/jech.2005.040071] [Medline: 16537355]

XSL•FO
RenderX

22. Vermunt JK, Magidson J. Latent class models for classification. Comput Stat Data Anal 2003 Jan;41(3-4):531-537. [doi: 10.1016/S0167-9473(02)00179-2]
23. Vermunt J, Magidson J. Latent GOLD 4.0 choice user's guide. Belmont, MA: Statistical Innovations Inc; 2005.
24. Volker MA. Reporting effect size estimates in school psychology research. Psychol Schools 2006 Jul;43(6):653-672.
25. Gatto SL, Tak SH. Computer, internet, and e-mail use among older adults: benefits and barriers. Educ Gerontol 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
26. Peek ST, Luijkx KG, Aarts S, Nieboer MW, van der Voort CS, van Hoof J, et al. Use one, use all? Selective use of apps by seniors. In: Gerontechnology. 2016 Presented at: ISG 2016; September 2016; Nice. [doi: 10.4017/gt.2016.15.s.662.00]
27. Baltes PB, Baltes MM. Psychological perspectives on successful aging: the model of selective optimization with compensation. In: Baltes PB, Baltes MM, editors. Successful aging: Perspectives from the behavioral sciences. New York: Cambridge University Press; 1990:1-34.
28. Choi NG, Dinitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. J Med Internet Res 2013 May 16;15(5):e97 [FREE Full text] [doi: 10.2196/jmir.2333] [Medline: 23681083]
29. d'Orsi E, Xavier AJ, Steptoe A, de Oliveira C, Ramos LR, Orrell M, et al. Socioeconomic and lifestyle factors related to instrumental activity of daily living dynamics: results from the English Longitudinal Study of Ageing. J Am Geriatr Soc 2014 Sep;62(9):1630-1639. [doi: 10.1111/jgs.12990] [Medline: 25243677]
30. Lim CSC. Designing inclusive ICT products for older users: taking into account the technology generation effect. J Eng Design 2010 Apr;21(2):189-206. [doi: 10.1080/09544820903317001]
31. Luijkx K, Peek S, Wouters E. "Grandma, You Should Do It--It's Cool" older adults and the role of family members in their acceptance of technology. Int J Environ Res Public Health 2015 Dec 05;12(12):15470-15485 [FREE Full text] [doi: 10.3390/ijerph121214999] [Medline: 26690188]
32. Lee C, Coughlin JF. PERSPECTIVE: Older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
33. Krijgsman J, Swinkels I, van Lettow B, de Jong J, Out K, Friele R, et al. Nictiz. 2016. E-health monitor URL: https://www.nictiz.nl/ehealth/ehealth-monitor [accessed 2017-05-15] [WebCite Cache ID 6qTaKNN47]
34. Peeters JM, Krijgsman JW, Brabers AE, Jong J, Friele RD. Use and uptake of ehealth in general practice: a cross-sectional survey and focus group study among health care users and general practitioners. JMIR Med Inform 2016 Apr 06;4(2):e11 [FREE Full text] [doi: 10.2196/medinform.4515] [Medline: 27052805]

Abbreviations

AIC3: Akaike's Information Criterion 3
ADL: activities of daily living
ANOVA: analysis of variance
BIC: Bayesian Information Criterion
iADL: instrumental activities of daily living
ICT: information and communication technology
LCA: latent class analysis
LISS: longitudinal Internet studies for social sciences
LL: log likelihood

Edited by G Eysenbach; submitted 20.10.16; peer-reviewed by M Jordan-Marsh, J Apolinário-Hagen; comments to author 21.01.17; revised version received 07.03.17; accepted 23.03.17; published 24.05.17.

Please cite as:
van Boekel LC, Peek STM, Luijkx KG
Diversity in Older Adults' Use of the Internet: Identifying Subgroups Through Latent Class Analysis
J Med Internet Res 2017;19(5):e180
URL: http://www.jmir.org/2017/5/e180/
doi:10.2196/jmir.6853
PMID:28539302

©Leonieke C van Boekel, Sebastiaan TM Peek, Katrien G Luijkx. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete

XSL•FO
RenderX

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

XSL•FO
RenderX

Original Paper

Patient-Centered Decision Support: Formative Usability Evaluation of Integrated Clinical Decision Support With a Patient Decision Aid for Minor Head Injury in the Emergency Department

Edward R Melnick[1], MHS, MD; Erik P Hess[2], MSc, MD; George Guo[3], BA; Maggie Breslin[4], MDes; Kevin Lopez[1], MS; Anthony J Pavlo[5], PhD; Fuad Abujarad[1], PhD; Seth M Powsner[1,5], MD; Lori A Post[6], PhD

[1]Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, United States

[2]Department of Emergency Medicine, Mayo Clinic, Rochester, MN, United States

[3]Yale School of Medicine, New Haven, CT, United States

[4]School of Visual Arts, New York, NY, United States

[5]Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States

[6]Department of Emergency Medicine, Feinberg School of Medicine, Chicago, IL, United States

Corresponding Author:
Edward R Melnick, MHS, MD
Department of Emergency Medicine
Yale School of Medicine
464 Congress Ave, Suite 260
New Haven, CT,
United States
Phone: 1 203 737 6454
Fax: 1 203 785 4580
Email: edward.melnick@yale.edu

Abstract

Background: The Canadian Computed Tomography (CT) Head Rule, a clinical decision rule designed to safely reduce imaging in minor head injury, has been rigorously validated and implemented, and yet expected decreases in CT were unsuccessful. Recent work has identified empathic care as a key component in decreasing CT overuse. Health information technology can hinder the clinician-patient relationship. Patient-centered decision tools to support the clinician-patient relationship are needed to promote evidence-based decisions.

Objective: Our objective is to formatively evaluate an electronic tool that not only helps clinicians at the bedside to determine the need for CT use based on the Canadian CT Head Rule but also promotes evidence-based conversations between patients and clinicians regarding patient-specific risk and patients' specific concerns.

Methods: User-centered design with practice-based and participatory decision aid development was used to design, develop, and evaluate patient-centered decision support regarding CT use in minor head injury in the emergency department. User experience and user interface (UX/UI) development involved successive iterations with incremental refinement in 4 phases: (1) initial prototype development, (2) usability assessment, (3) field testing, and (4) beta testing. This qualitative approach involved input from patients, emergency care clinicians, health services researchers, designers, and clinical informaticists at every stage.

Results: The *Concussion or Brain Bleed* app is the product of 16 successive iterative revisions in accordance with UX/UI industry design standards. This useful and usable final product integrates clinical decision support with a patient decision aid. It promotes shared use by emergency clinicians and patients at the point of care within the emergency department context. This tablet computer app facilitates evidence-based conversations regarding CT in minor head injury. It is adaptable to individual clinician practice styles. The resultant tool includes a patient injury evaluator based on the Canadian CT Head Rule and provides patient specific risks using pictographs with natural frequencies and cues for discussion about patient concerns.

Conclusions: This tool was designed to align evidence-based practices about CT in minor head injury patients. It establishes trust, empowers active participation, and addresses patient concerns and uncertainty about their condition. We hypothesize that, when implemented, the *Concussion or Brain Bleed* app will support—not hinder—the clinician-patient relationship, safely reduce CT use, and improve the patient experience of care.

XSL•FO
RenderX

(J Med Internet Res 2017;19(5):e174) doi:10.2196/jmir.7846

KEYWORDS

clinical decision support; decision aids; head injury, minor; medical informatics; spiral computed tomography; health services overuse; patient-centered outcomes research

Introduction

After a patient sustains a minor head injury, computed tomography (CT) imaging can diagnose structural brain injuries like hemorrhages but cannot detect the presence or severity of concussion [1]. The Canadian CT Head Rule, a clinical decision rule designed to safely reduce imaging in minor head injury, was developed and validated but has not decreased CT use [2-4]. This rule is 100% sensitive for predicting the need for neurosurgical intervention and more specific than other guidelines [3,5-7]. It should decrease CT use by one-third, making care more affordable, efficient, and safer [8-12]. The American Board of Internal Medicine and the American College of Emergency Physicians Choosing Wisely Initiative recommends avoiding unnecessary head CTs in emergency department (ED) patients with minor head injuries as the top national priority for addressing overuse in emergency care [13]. Conversely, CT imaging rates increased, and clinical decision support (CDS) implementation efforts have only had a modest effect (5%-8%) on decreasing CT use [4,14,15]. Research on nonclinical factors that influence overuse of CT revealed clinicians and patients identified establishing trust, patient engagement, and reassurance as essential to decreasing overuse of imaging [16,17].

Empathic care requires tools that facilitate conversation between patient and clinician [17-19]. Unfortunately, contemporary electronic health records (EHRs) tend to impede conversation [20-24]. The EHR interface physically separates the clinician from the patient, compromising communication. It distracts and decreases eye contact, touch, and decreases patient time with clinicians [20-23] and focuses almost entirely on physician behavior even if it is patient-specific (and evidence-based). Informing patients directly has rarely been part of the effort [25-27]. CDS is most effective when it is part of the clinician workflow at the time and location of decision making [27,28]. Patient decision aids, on the other hand, focus on patients, trying to help them decide among options by clarifying patient values, preferences, and goals and providing the best scientific evidence available to increase understanding of possible risks, benefits, alternatives, and their associated outcomes [29]. A successful decision aid facilitates conversation between the patient and clinician and improves patient engagement [18].

Current EHRs prohibit empathic care. Technology must support—not hinder—the clinician-patient relationship. Although paper charts were intuitive and simple, they were criticized for being disorganized and illegible, leading to medical errors. EHRs promised to improve patient safety and outcomes by reducing errors. In the rush to adopt EHRs to qualify for federal incentive payments, clinicians and hospitals adopted products with poor usability and poor integration that impede clinical workflow [30]. The EHR's potential for improving care has not yet been realized [27,28]. A large-scale study of EHR implementation found no negative association with mortality or adverse events across 17 hospitals [31]. EHR implementation has done harm in other ways [21,24,30]. Ratanawongsa et al [21] found high computer use by clinicians to be associated with lower patient satisfaction and communication. Sinsky et al [22] also found that physicians only spend 27% of their time face to face with patients, with 49% of their time spent on the EHR and desk work. In addition, EHR documentation requires an additional 1 to 2 hours daily of after-hour charting. A productivity analysis in the emergency care setting found that data entry accounted for 43% of physician time, requiring 4000 mouse clicks per shift [23]. Furthermore, EHRs in their current form physically obstruct and separate the clinician and patient, denying patients time with their clinician as well as compromising communication and human connection by distracting and decreasing eye contact and touch [20-23]. We propose that the patient-centered decision support presented here is the first step toward a more empathic medical interface that can support the clinician-patient relationship.

We developed a computerized, user-centered decision support tool called *Concussion or Brain Bleed* [32] for use on tablet computers (with 1536 × 2048 resolution) that integrates a patient decision aid and CDS at the bedside for decisions about CT use in ED patients with minor head injury. Herein is the design, development, and user experience and user interface (UX/UI) evaluation of *Concussion or Brain Bleed*. *Concussion or Brain Bleed* aims to engage patients in their care by giving them an understanding of their condition and helping them trust their clinician to safely reduce CT use in minor head injury.

Methods

Design

A user-centered design approach based on UX/UI industry standards was followed to develop a decision tool to promote shared decision making [33-35]. User-centered design is an iterative, multistage design and evaluation approach that is driven and refined by user input and customizes the interface based upon an explicit understanding of users, tasks, and environments [36]. UX design refers to user experience, while UI design stands for user interface. Both elements are crucial to app development. UX/UI refers to different aspects of the design. UX design is more comprehensive than UI—encompassing user needs, values, abilities, and limitations as they relate to the user's interaction with and perception of the design product. UI design focuses on ensuring that the graphical interface has elements that can be used, accessed, and understood based on user needs. UX/UI development was adopted with a goal of creating a tool that deviates from traditional CDS (eg, alerts and reminders).

UX/UI design elicits feedback and input from a multidisciplinary team, here including patients, emergency care clinicians

XSL•FO
RenderX

(attending physicians, residents, physician assistants, and nurse practitioners), health services researchers, interaction design experts, and health systems information specialists (including a software system engineer and a computer programmer) to make incremental refinements to the prototypes. The development process involved successive iterations of the prototype within 4 UX/UI phases: (1) initial prototype development, (2) usability testing, (3) field testing, and (4) beta testing (Figure 1). Each phase continued until thematic saturation [37]. The initial prototype, including review and synthesis of the evidence and analysis of usual practice, has been previously described [16,19]. In the second phase, we performed formative usability evaluation in a simulated clinical environment using clinicians with standardized patients to maximize ease of use and clinical integration. Next, the prototype was field-tested by the research team with ED patients and, finally, the tool was beta-tested during clinical care by physician users.

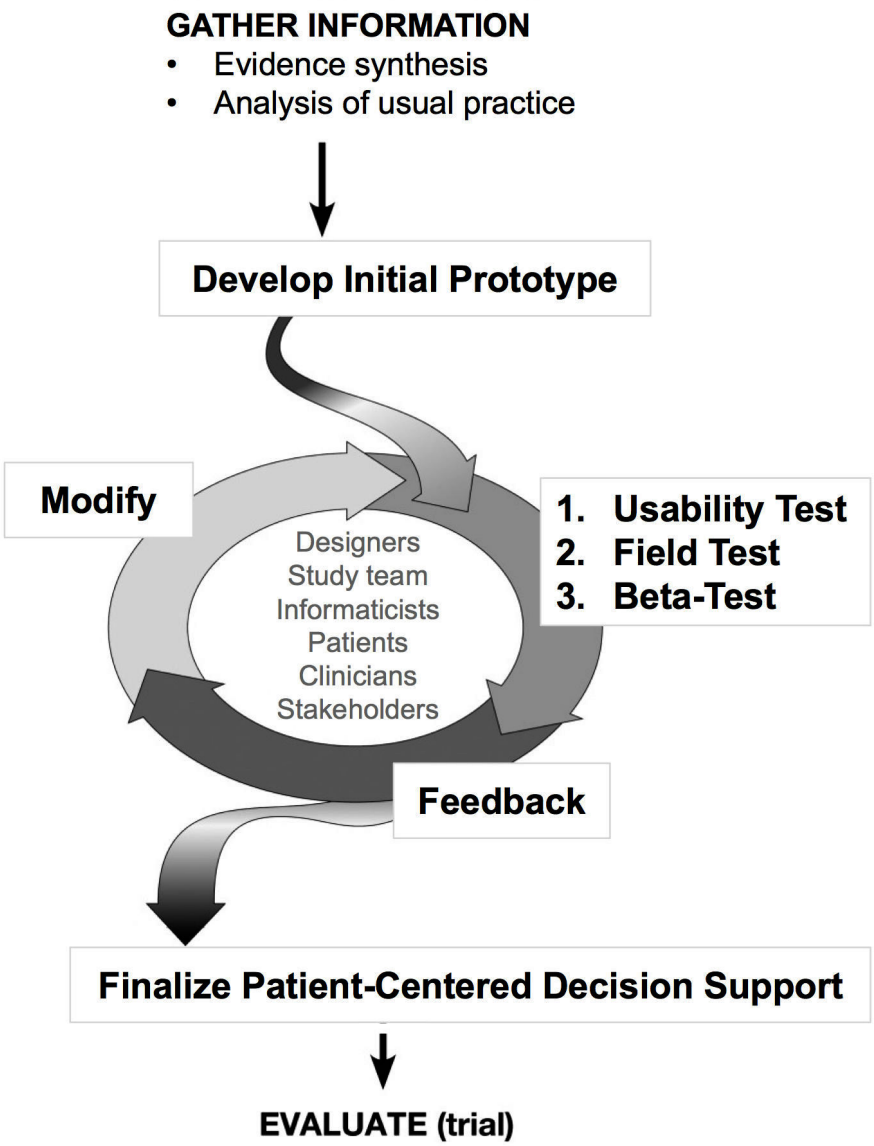
Study Setting and Population

Participants were patients and clinicians recruited from an urban, academic Level I trauma center ED with 103,000 patient visits per year and a satellite ED with 24,000 patient visits per year. Clinicians were recruited from the 48 attending physician faculty, 58 resident physicians, and 47 midlevel providers.

Protocol

The study protocol was approved by the hospital institutional review board (IRB). All participants provided their verbal consent as specified by the IRB. Some portions of the evaluation were performed at an outside institution. The protocol was also approved by that institution's IRB. Usability evaluation subjects were compensated for their time and travel with $100 gift cards. In beta testing, physicians were compensated for their time with $50 gift cards for each patient enrolled. Patients were not compensated in beta testing during their ED visit.

Figure 1. Patient-centered decision support development process.



XSL•FO
RenderX

User-Centered Design

Development of Initial Prototype

The full details of the initial development process including review and synthesis of the evidence and analysis of usual practice are reported elsewhere [16,19]. To identify nonclinical human factors that promote or inhibit appropriate use of CT in patients presenting to the ED with minor head injury, we performed qualitative studies in 3 phases: (1) patient focus groups, (2) clinician focus groups, and, (3) cognitive task analysis with direct ED observation and individual semistructured interviews using the critical decision method [16]. Next, a multidisciplinary team applied the findings from the qualitative study as user requirements for the initial prototype [19]. Primary goals were to promote smooth navigation through screens while completing tasks of patient education, risk communication, and shared decision making in the ED.

Usability Evaluation

Formative usability evaluations were conducted in a simulated environment to observe, record, and analyze a standardized clinician-patient encounter with the prototype. Using a "think aloud" protocol, scripted simulations of patient encounters with clinicians and standardized patients were observed and analyzed [38]. Attending emergency physicians were given a case study (Multimedia Appendix 1) to use the prototype while commenting on what they saw, thought, did, and felt. Inferences were made about the reasoning process behind task completion. Afterwards, a usability feedback questionnaire and semistructured interview (Multimedia Appendix 2) were conducted to determine the tool's ease-of-use, usefulness, and how the decision-making process was affected by the tool.

Field Testing

To optimize naturalistic decision making under the constraints of the complex, high-pressure ED, field testing was conducted by the research team. ED patients available and amenable to participation were identified by the treating clinicians on duty. The prototype was implemented and reviewed by patients during their clinical encounter when they were not actively under evaluation. Patterns of conversation were analyzed while issues and challenges with the tool's use were noted; all notes and experiences were shared and used to track the performance of successive iterations of the prototype based on content and quality of the conversation between the study clinician and the patient. Patients completed a semistructured interview (Multimedia Appendix 3) regarding the tool's content and format within the ED context. The tool was iteratively refined according to ecological interface design to optimize communication of patient-specific risk [39,40]. After thematic saturation, the wireframe prototype was programmed for use as a Web app on an iPad (Apple Inc). Technical specifications and system requirements were similar to the initial prototype [19].

Beta Testing

Beta testing was conducted by emergency physicians using the interactive prototype during clinical care of ED patients with minor head injury. Physicians described their experience to improve workflow. Structured email interviews were conducted after physicians had seen multiple patients. Survey responses informed the final prototype.

Results

Concussion or Brain Bleed underwent 16 successive revisions with content, process, and format adjustment based on usability, field, and beta testing.

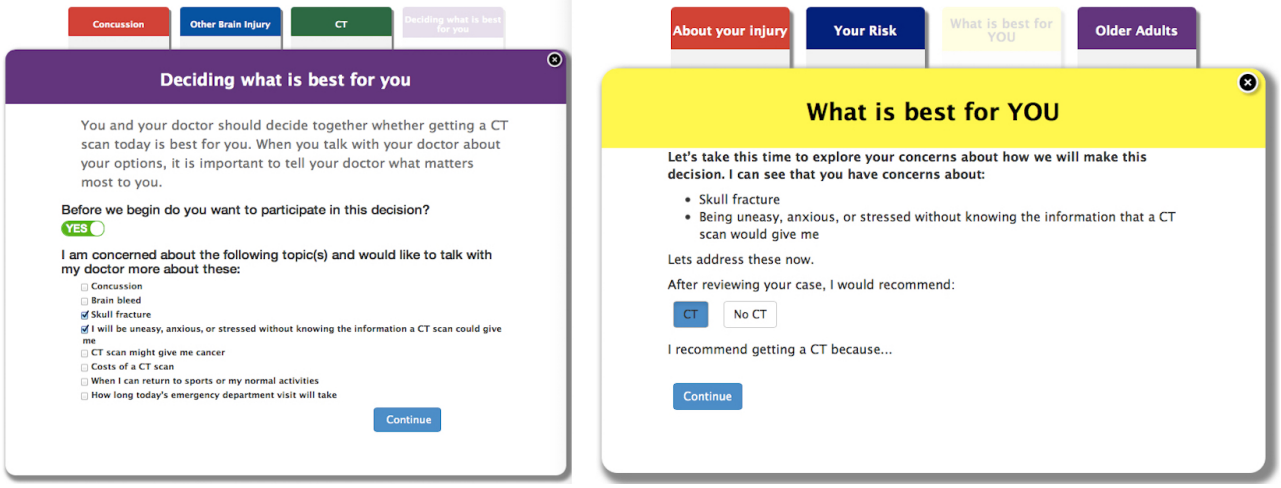
Development of Initial Prototype

The initial results of the prototype were previously reported [16,19]. Cognitive task analysis (critical decision method interviews and 150 hours of direct observation in the ED of peer-nominated senior emergency physicians recognized for their skill in safely minimizing testing while maintaining patient safety and engagement) revealed 5 core domains: trust, anxiety, constraints, influence of others, and patient expectations [16].

The initial prototype followed a visual metaphor of design reminiscent of decision aids on paper cards [19]. After the patient filled out eligibility and questionnaire forms to autopopulate subjective components of the clinical decision rule, 3 sections followed. The first section centered around patient education (information about concussions, CT scans) to be used by the patient alone prior to the clinician's evaluation and gave the patient the opportunity to flag concerns on a digital checklist. These concerns would later show up in the second section to be used by the clinician with the patient (screen capture of this section displayed in Figure 2). After completing a CDS checklist, the tool generated patient-specific risk estimates for pertinent outcomes and risk of cancer from a head CT. The final section involved a process of shared decision making in which patients and clinicians decided together whether to obtain a CT scan, to continue to be observed in the ED, or to go home.

XSL•FO
RenderX

Figure 2. Initial prototype screen capture of patient concern screens.

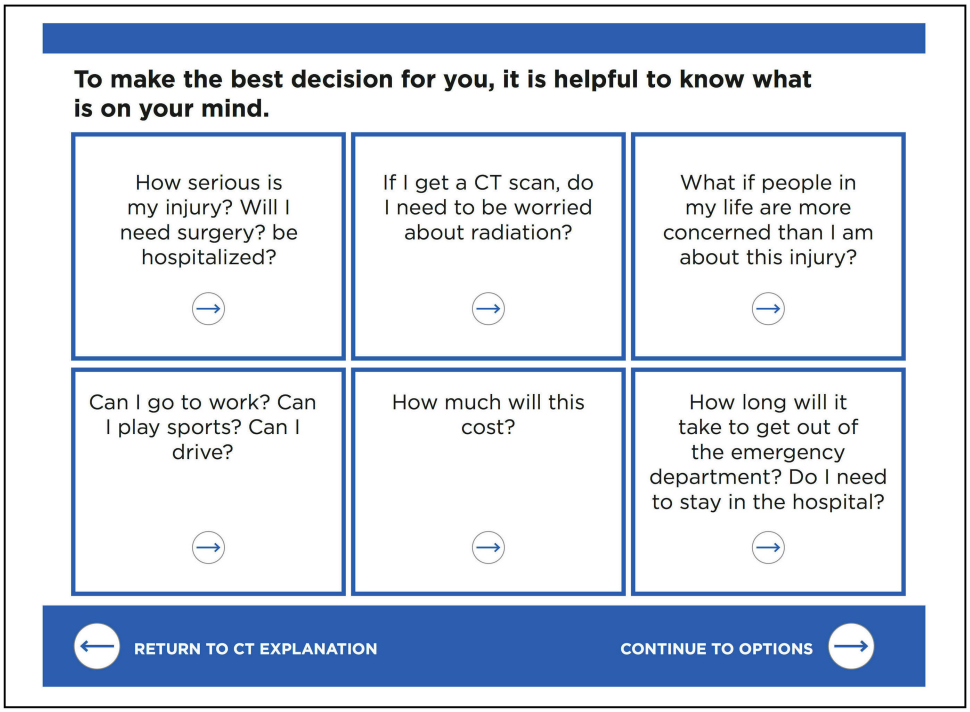


Usability Evaluation

Usability evaluation was conducted 3 times with 9 users. Observation revealed the tool required modification to facilitate conversation between the patient and clinician to be incorporated seamlessly into the clinical workflow [18,19]. Therefore, the initial user-centered design was augmented by interaction design using patient-centered and participatory decision aid development [18,19,41-44]. An interaction designer (MB) joined the research team [18,45,46]. Subsequent rounds involved rapid prototyping and low-fidelity wireframing.

This enhanced approach focused on tool usefulness (and lack of use by test subjects). Interview responses revealed users were not using the tool because the tool was overly prescriptive with too much text on the screen that interrupted or distracted from conversation with patients. Earlier prototypes were overdesigned, which forced clinicians to give more attention to the tool than the patient or to abandon the tool. Eliminations included the patient section with educational materials for patient review prior to the clinician's evaluation (based on previous qualitative findings that patients come to the ED for a clinician's expert evaluation) and a patient demographic survey and questionnaire about the injury. Revisions dramatically reduced the number of screen taps, checkboxes, and data entry. Furthermore, the Concerns section expanded to 6 boxes a patient could select to discuss (Figure 3). This minimalist version allowed clinicians to adapt the tool to their practice style and patient-specific education. It reassured patients by providing structure to the clinical conversation with cues (eg, How soon can I get back to work?). The tool was less prescriptive and increased the likelihood of implementation.

Figure 3. Revised Concerns section after initial round of usability evaluation.



XSL•FO
RenderX

Field Testing

Field testing was conducted with 10 patients. Additional incremental revisions were made to the prototype. Observation and analysis of use in the ED context and application of ecological interface design principles distilled the workflow for the final *Concussion or Brain Bleed* app (Figure 4). This further elucidated important patient issues. The final app now supports the clinician's decision and patient engagement and education around patient-specific risk about head injuries, CT imaging, counseling, and patient concerns.

Data entry was streamlined, and explicit user input was nearly eliminated. Grouping risk categories provides the clinician with the patient's individualized risk assessment by a single tap of the screen (Figure 5). This efficient Canadian CT Head Rule display gives the clinician more time for risk communication with the patient.

The risk visualization format and content underwent revisions from the initial prototype through usability and field testing (Figure 6 a-d). The initial prototype used text-based risks (eg, clinically important brain injury) [2,19,47]. Later versions used pictographs, plain language, absolute risks with a constant denominator, and a color scheme to differentiate the 4 categories of patient-centered outcomes [42,44,48,49].

A key finding was how important it is to teach and emphasize that a concussion is not visible on CT. The tool evolved into helping patients understand specific recommendations and their implications. The Risk Discussion section offers plain language on the utility (or lack thereof in low-risk patients) of CT as well as cues to discuss concussion and the individual patient's concerns (Figure 7).

Figure 4. Conceptualization of the workflow and potential pathways for the Concussion or Brain Bleed application.

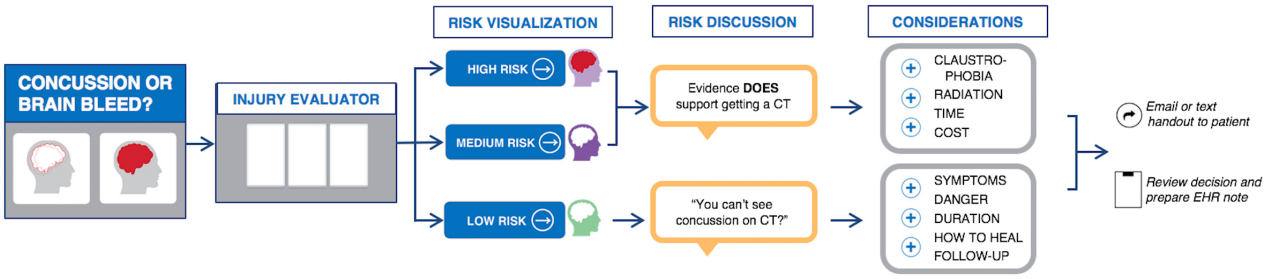


Figure 5. Clinical Decision Support portion of app after field testing.



XSL•FO
RenderX

Figure 6. Risk visualization for low-risk patients from the initial prototype (top left) through usability testing (early, top right; late, bottom left) and field testing (bottom right).

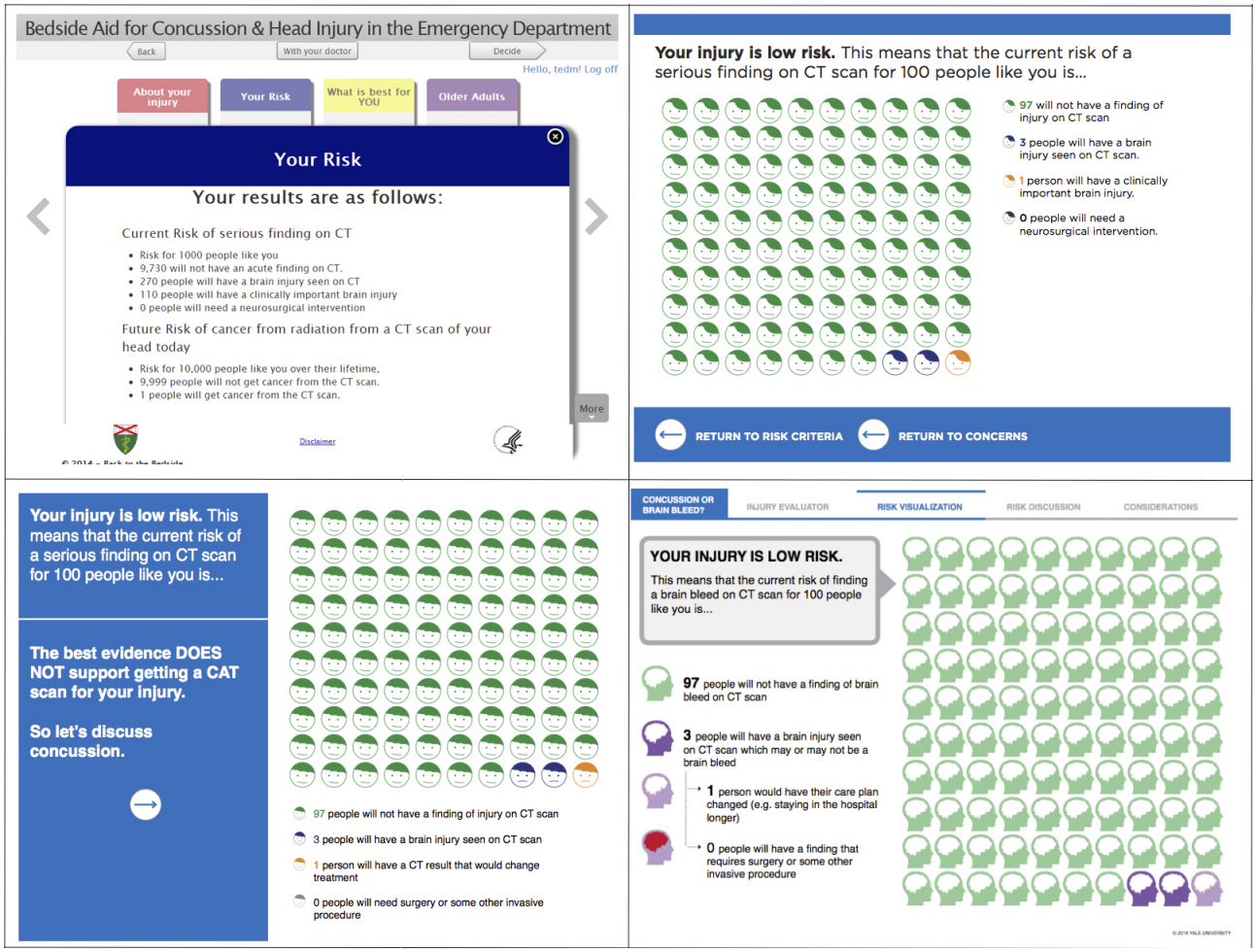
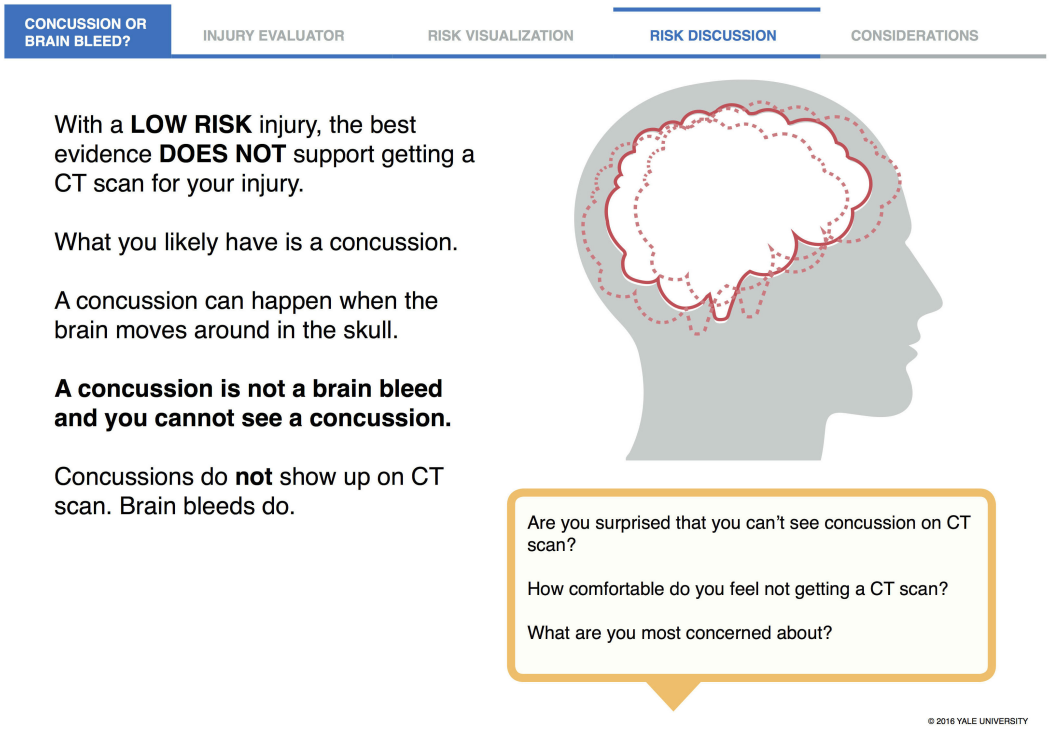


Figure 7. Risk discussion screen for low-risk patients after field testing.



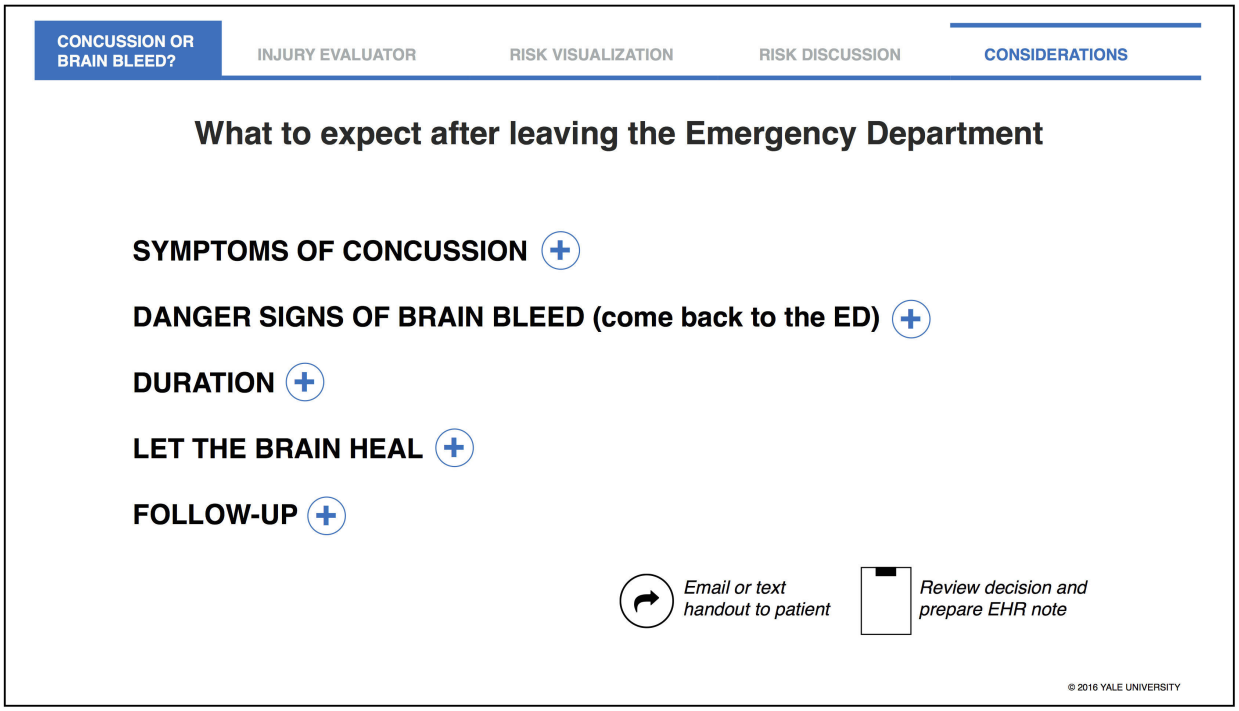
XSL•FO
RenderX

Beta Testing

Beta testing was conducted over 6 weeks with 4 attending emergency physicians in the care of 7 low-risk, minor head injury ED patients. The final Considerations section for low-risk patients was revised based on user feedback that it was too busy. Prior to beta testing, this section had a wall of text including a large inventory of sections that could be discussed at the clinician's discretion. Beta testing revealed just a checklist with the option to expand sufficed. The section's content remained relatively unchanged with the format converted to a checklist with single-tap dropdown options that provided more information (via hypertext) when specifically selected (Figure 8). Readability increased with limited distractions while remaining flexible to differing clinician practice styles and individual patient needs.

We developed a work-around for integration with EHR workflow using Epic (Epic Systems Corp) SmartPhrases (Multimedia Appendix 4). This charting tool allows clinicians to autopopulate text using shorthand. SmartPhrases allow rapid documentation of use of the *Concussion or Brain Bleed* app in the EHR.

Figure 8. Considerations screen for low-risk patients after beta testing.



Discussion

Principal Findings

A total of 16 successive iterations have resulted in a tool that integrates Canadian CT Head Rule CDS at point of care with a patient decision aid to promote conversation around individualized risk and patients' specific concerns. Design, development, and formative evaluation were informed by the philosophy that technology can accelerate the provision of evidence-based care that is efficient and empathic, effectively reducing unnecessary care [17,19,24]. The user can traverse the app in its entirety in 3 to 5 screen taps. *Concussion or Brain Bleed* addresses the human factors that research demonstrates are critical for optimizing CT use in minor head injury by creating the time and space for conversation between patients and their clinician [16]. The app equips clinicians to foster trust and manage patient expectations in a fast-paced ED environment characterized by uncertainty and high emotions.

We elected a formative evaluation inclusive of usability evaluation—namely, the well-established practice-based, patient-centered, and participatory decision aid development process adapted for our computerized tool [18,19,41-44]. It allowed for problematic elements of the prototype to be rapidly identified and addressed and usability of unchanged elements to be tested by multiple users through existing and subsequent iterations. The final product is refined based on user input and represents the culmination of rigorous testing in simulations and real-world clinical encounters. Feedback from a multidisciplinary team has been incorporated with the express goal of practicality and usability. Our tool addresses multiple items across the dimensions of the International Patient Decision Aids Standards including the use of a systematic development process, presenting information on probabilities of outcomes and using the scientific literature on which content is based, conveyed with plain language [43]. Involving end-users and a variety of clinicians in both simulated and real-world clinical environments in an iterative process ensures that the format and information content of our tool is responsive to user preferences and the complexities of decision context.

Comparison With Prior Work

CDS is most effective when it is part of the clinician's workflow at the time and location of decision making [27,28]. Decision support strategies to date have focused on physician behavior [25-27]. By bringing CDS to the point of care and integrating

XSL•FO
RenderX

it with a decision aid on a tablet computer shared by the patient and clinician, *Concussion or Brain Bleed* could improve both the quantity and quality of time at the beside. Patient decision aids have been established as an effective way to translate evidence-based care into practice [29,50]. Visualizing benefits and harms can lead to increased patient knowledge and involvement in decision making, greater satisfaction with the decision-making process, and optimal health outcomes consistent with patient values and preferences [29]. Patient involvement begins with the development of a partnership and includes participation in information exchange, deliberation, and decision making [51]. Patients—even older patients with more experience in a historically paternalistic mode—report great interest in getting involved in similar decision making again [52,53]. While paper-based patient decision aids are beneficial, they do not provide CDS for the clinician and are static in nature with regard to patient-specific estimates and concerns [18,29,44,50]. Tablet computers retain the portability and usability of a paper-based decision aid while also providing the customization and flexibility of computerized CDS with regard to patient-specific risk visualization. In developing a digital tool, we also have the additional benefit of having a database to collect, edit, store, and retrieve data generated by the tool and further reduce workflow burden through direct integration with EHR systems.

Conclusions

The fight to stem medical overuse will require the use of disruptive technologies—often innovative but simple, high-value solutions that can be widely adopted and easily used. In creating this patient-centered clinical decision support tool, we aim to decrease CT use for minor head injury. This tool combines evidence-based practices with patient engagement that establishes trust, empowers active participation, and addresses patient concerns and uncertainty about their condition at the point of care. It helps clinicians to determine who needs a CT and then helps patients to understand why. We hypothesize that, when implemented, the *Concussion or Brain Bleed* app will support—not hinder—the clinician-patient relationship, safely reduce CT use, and improve the patient experience of care.

Acknowledgments

This project was supported by grant number K08HS021271 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. The authors would like to express their gratitude to Sundas Khan, Lauren McCullagh, and Dr Thomas McGinn for facilitating usability evaluation at the Northwell Health Center for Learning and Innovation; Dr Richard Shiffman for methodological mentorship; and Caitlin Johnson for facilitating the IRB process.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Usability evaluation standardized cases.

[PDF File (Adobe PDF File), 44KB - jmir_v19i5e174_app1.pdf]

Multimedia Appendix 2

Usability evaluation postuse semistructured interview guide.

[PDF File (Adobe PDF File), 42KB - jmir_v19i5e174_app2.pdf]

Multimedia Appendix 3

Head computed tomography decision tool for the emergency department prototype research questions.

[PDF File (Adobe PDF File), 44KB - jmir_v19i5e174_app3.pdf]

Multimedia Appendix 4

Beta testing instructions for using the app and Epic SmartPhrases to document use of the tool in the electronic health record.

[PDF File (Adobe PDF File), 172KB - jmir_v19i5e174_app4.pdf]

References

1. Dematteo CA, Hanna SE, Mahoney WJ, Hollenberg RD, Scott LA, Law MC, et al. “My child doesn't have a brain injury, he only has a concussion”. Pediatrics 2010 Feb;125(2):327-334. [doi: 10.1542/peds.2008-2720] [Medline: 20083526]
2. Stiell IG, Wells GA, Vandemheen K, Clement C, Lesiuk H, Laupacis A, et al. The Canadian CT Head Rule for patients with minor head injury. Lancet 2001 May 05;357(9266):1391-1396. [Medline: 11356436]

XSL•FO
RenderX

3. Stiell IG, Clement CM, Rowe BH, Schull MJ, Brison R, Cass D, et al. Comparison of the Canadian CT Head Rule and the New Orleans Criteria in patients with minor head injury. JAMA 2005 Sep 28;294(12):1511-1518. [doi: 10.1001/jama.294.12.1511] [Medline: 16189364]
4. Stiell IG, Clement CM, Grimshaw JM, Brison RJ, Rowe BH, Lee JS, et al. A prospective cluster-randomized trial to implement the Canadian CT Head Rule in emergency departments. CMAJ 2010 Oct 05;182(14):1527-1532 [FREE Full text] [doi: 10.1503/cmaj.091974] [Medline: 20732978]
5. Smits M, Dippel DW, de Haan GG, Dekker HM, Vos PE, Kool DR, et al. External validation of the Canadian CT Head Rule and the New Orleans Criteria for CT scanning in patients with minor head injury. JAMA 2005 Sep 28;294(12):1519-1525. [doi: 10.1001/jama.294.12.1519] [Medline: 16189365]
6. Jagoda AS, Bazarian JJ, Bruns JJ, Cantrill SV, Gean AD, Howard PK, American College of Emergency Physicians, Centers for Disease ControlPrevention. Clinical policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting. Ann Emerg Med 2008 Dec;52(6):714-748. [doi: 10.1016/j.annemergmed.2008.08.021] [Medline: 19027497]
7. Papa L, Stiell IG, Clement CM, Pawlowicz A, Wolfram A, Braga C, et al. Performance of the Canadian CT Head Rule and the New Orleans Criteria for predicting any traumatic intracranial injury on computed tomography in a United States Level I trauma center. Acad Emerg Med 2012 Jan;19(1):2-10 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01247.x] [Medline: 22251188]
8. Melnick ER, Szlezak CM, Bentley SK, Dziura JD, Kotlyar S, Post LA. CT overuse for mild traumatic brain injury. Jt Comm J Qual Patient Saf 2012 Nov;38(11):483-489. [Medline: 23173394]
9. Parma C, Carney D, Grim R, Bell T, Shoff K, Ahuja V. Unnecessary head computed tomography scans: a level 1 trauma teaching experience. Am Surg 2014 Jul;80(7):664-668. [Medline: 24987897]
10. Korley FK, Morton MJ, Hill PM, Mundangepfupfu T, Zhou T, Mohareb AM, et al. Agreement between routine emergency department care and clinical decision support recommended care in patients evaluated for mild traumatic brain injury. Acad Emerg Med 2013 May;20(5):463-469 [FREE Full text] [doi: 10.1111/acem.12136] [Medline: 23672360]
11. Sharp AL, Nagaraj G, Rippberger EJ, Shen E, Swap CJ, Silver MA, et al. Computed tomography use for adults with head injury: describing likely avoidable emergency department imaging based on the Canadian CT Head Rule. Acad Emerg Med 2017 Dec;24(1):22-30. [doi: 10.1111/acem.13061] [Medline: 27473552]
12. Melnick ER. Big versus small data and the generalizability of the rate of computed tomography overuse in minor head injury. Acad Emerg Med 2017 Mar;24(3):391-392. [doi: 10.1111/acem.13084] [Medline: 27617902]
13. American College of Emergency Physicians. Ten Things Physicians and Patients Should Question. 2014 Nov 30. URL: http://www.choosingwisely.org/societies/american-college-of-emergency-physicians/ [accessed 2017-05-13] [WebCite Cache ID 6qRLP54PP]
14. Ip IK, Raja AS, Gupta A, Andruchow J, Sodickson A, Khorasani R. Impact of clinical decision support on head computed tomography use in patients with mild traumatic brain injury in the ED. Am J Emerg Med 2015 Mar;33(3):320-325. [doi: 10.1016/j.ajem.2014.11.005] [Medline: 25572644]
15. Sharp A, Huang B, Tang T, Shen E, Melnick E, Venkatesh A. Implementation of the Canadian CT Head Rule and its effect on utilization of computed tomography among patients with head injury. Ann Emerg Med (in press) 2017:2.
16. Melnick ER, Shafer K, Rodulfo N, Shi J, Hess EP, Wears RL, et al. Understanding overuse of computed tomography for minor head injury in the emergency department: a triangulated qualitative study. Acad Emerg Med 2015 Dec;22(12):1474-1483 [FREE Full text] [doi: 10.1111/acem.12824] [Medline: 26568523]
17. Melnick ER. How to make less more: empathy can fill the gap left by reducing unnecessary care. BMJ 2015 Nov 04;351:h5831. [Medline: 26537887]
18. Montori VM, Breslin M, Maleska M, Weymiller AJ. Creating a conversation: insights from the development of a decision aid. PLoS Med 2007 Aug;4(8):e233 [FREE Full text] [doi: 10.1371/journal.pmed.0040233] [Medline: 17683195]
19. Melnick ER, Lopez K, Hess EP, Abujarad F, Brandt CA, Shiffman RN, et al. Back to the bedside: developing a bedside aid for concussion and brain injury decisions in the emergency department. EGEMS (Wash DC) 2015;3(2):1136. [doi: 10.13063/2327-9214.1136] [Medline: 26290885]
20. Toll E. A piece of my mind. The cost of technology. JAMA 2012 Jun 20;307(23):2497-2498. [doi: 10.1001/jama.2012.4946] [Medline: 22797449]
21. Ratanawongsa N, Barton JL, Lyles CR, Wu M, Yelin EH, Martinez D, et al. Association between clinician computer use and communication with patients in safety-net clinics. JAMA Intern Med 2016 Jan;176(1):125-128 [FREE Full text] [doi: 10.1001/jamainternmed.2015.6186] [Medline: 26619393]
22. Sinsky C, Tutty M, Colligan L. Allocation of Physician Time in Ambulatory Practice. Ann Intern Med 2017 May 02;166(9):683-684. [doi: 10.7326/L17-0073] [Medline: 28460382]
23. Hill RG, Sears LM, Melanson SW. 4000 clicks: a productivity analysis of electronic medical records in a community hospital ED. Am J Emerg Med 2013 Nov;31(11):1591-1594. [doi: 10.1016/j.ajem.2013.06.028] [Medline: 24060331]
24. Gellert G, Webster S, Gillean J, Melnick E, Kanzaria H. Should US doctors embrace electronic health records? BMJ 2017 Dec 24;356:j242. [Medline: 28119282]
25. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998 Oct 21;280(15):1339-1346. [Medline: 9794315]

XSL•FO
RenderX

26. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
27. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 2;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
28. Karsh B. Clinical practice improvement and redesign: how change in workflow can be supported by clinical decision support. Rockville: Agency for Healthcare Research and Quality; 2009. URL: https://healthit.ahrq.gov/sites/default/files/docs/biblio/09-0054-EF-Updated_0.pdf [accessed 2017-05-13] [WebCite Cache ID 6qRLwQsDv]
29. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2014;1:CD001431. [doi: 10.1002/14651858.CD001431.pub4] [Medline: 24470076]
30. Rosenbaum L. Transitional chaos or enduring harm? The EHR and the disruption of medicine. N Engl J Med 2015 Oct 22;373(17):1585-1588. [doi: 10.1056/NEJMp1509961] [Medline: 26488690]
31. Barnett ML, Mehrotra A, Jena AB. Adverse inpatient outcomes during the transition to a new electronic health record system: observational study. BMJ 2016 Jul 28;354:i3835 [FREE Full text] [Medline: 27471242]
32. Concussion or Brain Bleed?.: Yale University URL: https://b2b.med.yale.edu/ [accessed 2017-05-12] [WebCite Cache ID 6qPTI68mY]
33. Dong T, Churchill E, Nichols J. Understanding the challenges of designing and developing multi-device experiences. 2901851: ACM; 2016 Presented at: Proceedings of the 2016 ACM Conference on Designing Interactive Systems; 2016; Brisbane p. 62-72.
34. Panhale M. Introduction to Mobile Application Development Ecosystems: Beginning Hybrid Mobile Application Development. Berkeley: Apress; 2016:1-13.
35. Chung K, Kim J, Park RC. Knowledge-based health service considering user convenience using hybrid Wi-Fi P2P. Inf Technol Manag 2015 Jul 24;17(1):67-80. [doi: 10.1007/s10799-015-0241-5]
36. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]
37. Curry LA, Nembhard IM, Bradley EH. Qualitative and mixed methods provide unique contributions to outcomes research. Circulation 2009 Mar 17;119(10):1442-1452 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.742775] [Medline: 19289649]
38. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 1993 Nov;3(4):430-441. [doi: 10.1177/104973239300300403]
39. Floch M, Walker W, Ringel Y. An ecological approach to interface design. 1998 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 1998; Albuquerque.
40. Vicente K. Ecological interface design: progress and challenges. Hum Factors 2002;44(1):62-78.
41. Hess EP, Wyatt KD, Kharbanda AB, Louie JP, Dayan PS, Tzimenatos L, et al. Effectiveness of the head CT choice decision aid in parents of children with minor head trauma: study protocol for a multicenter randomized trial. Trials 2014 Jun 25;15:253 [FREE Full text] [doi: 10.1186/1745-6215-15-253] [Medline: 24965659]
42. Melnick ER, Probst MA, Schoenfeld E, Collins SP, Breslin M, Walsh C, et al. Development and testing of shared decision making interventions for use in emergency care: a research agenda. Acad Emerg Med 2016 Dec;23(12):1346-1353. [doi: 10.1111/acem.13045] [Medline: 27457137]
43. Elwyn G, O'Connor AM, Bennett C, Newcombe RG, Politi M, Durand M, et al. Assessing the quality of decision support technologies using the International Patient Decision Aid Standards instrument (IPDASi). PLoS One 2009;4(3):e4705 [FREE Full text] [doi: 10.1371/journal.pone.0004705] [Medline: 19259269]
44. Coylewright M, Shepel K, Leblanc A, Pencille L, Hess E, Shah N, et al. Shared decision making in patients with stable coronary artery disease: PCI choice. PLoS One 2012;7(11):e49827 [FREE Full text] [doi: 10.1371/journal.pone.0049827] [Medline: 23226223]
45. Breslin M, Mullan RJ, Montori VM. The design of a decision aid about diabetes medications for use during the consultation with patients with type 2 diabetes. Patient Educ Couns 2008 Dec;73(3):465-472. [doi: 10.1016/j.pec.2008.07.024] [Medline: 18771876]
46. Hess EP, Knoedler MA, Shah ND, Kline JA, Breslin M, Branda ME, et al. The chest pain choice decision aid: a randomized trial. Circ Cardiovasc Qual Outcomes 2012 May;5(3):251-259 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.111.964791] [Medline: 22496116]
47. Melnick E, Szlezak C, Dziura J, Stiell I. Risk estimates for the Canadian CT Head Rule in patients with minor head injury. 2013 Presented at: SAEM Annual Meeting; 2013; Atlanta.
48. Kurz-Milcke E, Gigerenzer G, Martignon L. Transparency in risk communication: graphical and analog tools. Ann NY Acad Sci 2008 Apr;1128:18-28. [doi: 10.1196/annals.1399.004] [Medline: 18469211]

XSL•FO
RenderX

49. Kent DM, Shah ND. Personalizing evidence-based primary prevention with aspirin: individualized risks and patient preference. Circ Cardiovasc Qual Outcomes 2011 May;4(3):260-262 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.111.961136] [Medline: 21586722]
50. Hess EP, Hollander JE, Schaffer JT, Kline JA, Torres CA, Diercks DB, et al. Shared decision making in patients with low risk chest pain: prospective randomized pragmatic trial. BMJ 2016 Dec 05;355:i6165 [FREE Full text] [Medline: 27919865]
51. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [Medline: 9032835]
52. Mullan RJ, Montori VM, Shah ND, Christianson TJ, Bryant SC, Guyatt GH, et al. The diabetes mellitus medication choice decision aid: a randomized trial. Arch Intern Med 2009 Sep 28;169(17):1560-1568. [doi: 10.1001/archinternmed.2009.293] [Medline: 19786674]
53. Weymiller AJ, Montori VM, Jones LA, Gafni A, Guyatt GH, Bryant SC, et al. Helping patients with type 2 diabetes mellitus make treatment decisions: statin choice randomized trial. Arch Intern Med 2007 May 28;167(10):1076-1082. [doi: 10.1001/archinte.167.10.1076] [Medline: 17533211]

Abbreviations

CDS: clinical decision support
CT: computed tomography
ED: emergency department
EHR: electronic health record
IRB: institutional review board
UI: user interface
UX: user experience

Edited by G Eysenbach; submitted 11.04.17; peer-reviewed by I Hargraves, P Boisrond; comments to author 03.05.17; revised version received 03.05.17; accepted 05.05.17; published 19.05.17.

Please cite as:
Melnick ER, Hess EP, Guo G, Breslin M, Lopez K, Pavlo AJ, Abujarad F, Powsner SM, Post LA
Patient-Centered Decision Support: Formative Usability Evaluation of Integrated Clinical Decision Support With a Patient Decision Aid for Minor Head Injury in the Emergency Department
J Med Internet Res 2017;19(5):e174
URL: http://www.jmir.org/2017/5/e174/
doi:10.2196/jmir.7846
PMID:28526667

©Edward R Melnick, Erik P Hess, George Guo, Maggie Breslin, Kevin Lopez, Anthony J Pavlo, Fuad Abujarad, Seth M Powsner, Lori A Post. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 19.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Virtual Care Climate Questionnaire: Development and Validation of a Questionnaire Measuring Perceived Support for Autonomy in a Virtual Care Setting

Eline Suzanne Smit[1], PhD; Alexandra Lelia Dima[1,2], PhD; Stephanie Annette Maria Immerzeel[1], BSc; Bas van den Putte[1,3], PhD; Geoffrey Colin Williams[4,5], MD, PhD

[1]Amsterdam School of Communication Research/ASCoR, Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands

[2]Health Services and Performance Research (HESPER EA 7425), University Claude Bernard Lyon 1, Lyon, France

[3]Trimbos Institute, Netherlands Institute for Mental Health and Addiction, Utrecht, Netherlands

[4]Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY, United States

[5]Healthy Living Center, Center for Community Health, Department of Medicine, University of Rochester, Rochester, NY, United States

Corresponding Author:
Eline Suzanne Smit, PhD
Amsterdam School of Communication Research/ASCoR
Department of Communication Science
University of Amsterdam
PO Box 15791
Amsterdam, 1001 NG
Netherlands
Phone: 31 (0)6 25258558
Email: E.S.Smit@uva.nl

Abstract

Background: Web-based health behavior change interventions may be more effective if they offer autonomy-supportive communication facilitating the internalization of motivation for health behavior change. Yet, at this moment no validated tools exist to assess user-perceived autonomy-support of such interventions.

Objective: The aim of this study was to develop and validate the virtual climate care questionnaire (VCCQ), a measure of perceived autonomy-support in a virtual care setting.

Methods: Items were developed based on existing questionnaires and expert consultation and were pretested among experts and target populations. The virtual climate care questionnaire was administered in relation to Web-based interventions aimed at reducing consumption of alcohol (Study 1; N=230) or cannabis (Study 2; N=228). Item properties, structural validity, and reliability were examined with item-response and classical test theory methods, and convergent and divergent validity via correlations with relevant concepts.

Results: In Study 1, 20 of 23 items formed a one-dimensional scale (alpha=.97; omega=.97; *H*=.66; mean 4.9 [SD 1.0]; range 1-7) that met the assumptions of monotonicity and invariant item ordering. In Study 2, 16 items fitted these criteria (alpha=.92; *H*=.45; omega=.93; mean 4.2 [SD 1.1]; range 1-7). Only 15 items remained in the questionnaire in both studies, thus we proceeded to the analyses of the questionnaire's reliability and construct validity with a 15-item version of the virtual climate care questionnaire. Convergent validity of the resulting 15-item virtual climate care questionnaire was confirmed by positive associations with autonomous motivation (Study 1: r=.66, P<.001; Study 2: r=.37, P<.001) and perceived competence for reducing alcohol intake (Study 1: r=.52, P<.001). Divergent validity could only be confirmed by the nonsignificant association with perceived competence for learning (Study 2: r=.05, P=.48).

Conclusions: The virtual climate care questionnaire accurately assessed participants' perceived autonomy-support offered by two Web-based health behavior change interventions. Overall, the scale showed the expected properties and relationships with relevant concepts, and the studies presented suggest this first version of the virtual climate care questionnaire to be reasonably valid and reliable. As a result, the current version may cautiously be used in future research and practice to measure perceived support for autonomy within a virtual care climate. Future research efforts are required that focus on further investigating the virtual climate care questionnaire's divergent validity, on determining the virtual climate care questionnaire's validity and reliability

XSL•FO
RenderX

when used in the context of Web-based interventions aimed at improving nonaddictive or other health behaviors, and on developing and validating a short form virtual climate care questionnaire.

(J Med Internet Res 2017;19(5):e155) doi:10.2196/jmir.6714

KEYWORDS

questionnaire design; validation studies; psychometrics; personal autonomy; Internet; health behavior; health promotion; self-determination theory

Introduction

Unhealthy lifestyle behaviors—for example, smoking tobacco or cannabis, consuming too much alcohol, overeating calories or not eating a sufficient amount of fruit and vegetables, or being insufficiently physically active—are a major cause of chronic illnesses like cancer, diabetes, and cardiovascular diseases [1]. They have a detrimental effect on quality of life, decrease work productivity, and put an enormous, preventable strain on health care [2]. Developing effective interventions to promote healthy lifestyle behaviors that prevent or delay these diseases, as well as making these interventions available on a large scale and to a great variety of people, is thus important.

To improve one's lifestyle, self-determination theory (SDT) [3] proposes that an autonomous form of motivation and perceiving competence for changing are imperative. When it concerns the improvement of health-related behaviors like those mentioned before, people are assumed to perceive themselves to be autonomous in their motivation to change when the behavior is accompanied by an experience of psychological freedom of choice. Supportive of this theoretical assumption, earlier research has found autonomous motivation to be an important predictor of health behavior change, its maintenance, and subsequent positive health outcomes [4]. According to SDT, someone's autonomous motivation can be increased by providing **s**upport for autonomy. The concept of autonomy-support was first of all applied to a face-to-face context where one person, such as a health professional, interacts with another person, such as a patient, and tools have been developed to measure this concept in face-to-face settings. When provided by a health care professional, such as a general practitioner or lifestyle counselor, support for autonomy involves strategies like eliciting and acknowledging a person's perspective, providing a clear rationale for change, offering choice, and using noncontrolling language [5,6]. Perceived support for autonomy from a health care professional is measured through instruments like the health care climate questionnaire (HCCQ) [7]. Similarly, perceived autonomy-support from a physical education teacher can be assessed by the perceived autonomy support scale for exercise settings (PASSES) [8].

Nowadays, interventions aimed at a healthy lifestyle are increasingly delivered via the Internet [9,10], creating a virtual instead of face-to-face care climate. Web-based health behavior change interventions can successfully promote a healthy lifestyle [11] and have several advantages over and above face-to-face interventions; they are highly accessible, participants can use them at any convenient time, and many people can be reached at minimal cost [9]. It can be assumed to be equally important to provide support for autonomy within a virtual care climate in order to increase people's autonomous motivation for initiating and maintaining health behavior change. However, instead of human interaction, Web-based interventions deliver this support in different formats and rather make use of computer-human interaction. For interventions that make use of virtual health care providers, like a virtual clinician [12] or computerized personal trainer [13], the interaction between the virtual health care provider and receiver of the intervention might resemble the interaction between two humans. Consequently, only slight adjustments to questionnaires like the HCCQ and PASSES might be sufficient. There is, however, an abundance of Web-based interventions that do not involve virtual care providers, but in which autonomy-support is integrated in the structure of the Web-based tool (for examples of such interventions, see [14-16]). In this context, the operationalization of the concept of perceived autonomy-support would need to be reconsidered to a larger extent and available measurement instruments would need to be adapted to this new context. To illustrate, one of the items from the HCCQ reads, "My physician listens to how I would like to do things;" whereas a physician or other person is indeed able to listen, a Web-based intervention where no virtual health care provider is involved does not possess this ability. As a consequence, the item would not be applicable.

Although substantial evidence is available for the positive effects of perceived autonomy-support within the face-to-face setting [4,6], as well as some evidence for these effects when it concerns virtual health care providers [12,13], no such evidence exists for the role of autonomy-support in virtual care settings without virtual health care providers involved. To develop an evidence base for this growing field of Web-based interventions, an instrument is needed with adequate measurement properties specifically for assessing perceived support for autonomy in such settings. Therefore, the objective of this study was to develop and validate the virtual care climate questionnaire (VCCQ): the first measurement instrument of perceived support for autonomy in virtual care settings. By developing the VCCQ, this study aimed to fulfill the need for such tools and thus enable further research into how Web-based health behavior change interventions can successfully support autonomy, increase autonomous motivation, and ultimately promote a healthy lifestyle.

Methods

The VCCQ was constructed in three steps: item development, pretesting, and psychometric validation.

XSL•FO
RenderX

Item Development

The items of the VCCQ were developed primarily based on the HCCQ [7]. The 15 items of the HCCQ were adapted to measure perceived autonomy-support in a virtual instead of face-to-face care setting. The adaptation of the items consisted of rewording items from a setting in which there is direct contact with a physician (eg, "My physician answers my questions fully and carefully") to a setting in which the respondent interacts with a Web-based intervention (eg, "< *name intervention*> answers my questions fully and carefully"). Additionally, the target behavior was added when applicable to increase the specificity of the items (eg, "I feel that <name intervention> provides me with effective possibilities to <target behavior>"). The name of the new questionnaire, that is, virtual care climate questionnaire, aims to recognize the basis of most of its items, that is, the health care climate questionnaire. Yet, to ensure that a wide variety of autonomy-supportive strategies were represented in the VCCQ, four additional items were included based on the PASSES [8] (items 16 and 17) and a discussion with experts on motivation and Web-based health behavior change interventions (items 18 and 19).

Similar as in the HCCQ, a 7-point response scale was used with *totally disagree* (1) and *totally agree* (7) as endpoints. All items and possible responses were translated from English to Dutch.

Pretesting

To investigate the face validity of the VCCQ, a pretest was conducted among experts. Moreover, a pretest was conducted among the target population in order to identify improving alterations to the questionnaire.

Sample and Procedure

Pretests were conducted in the context of a Web-based, computer-tailored intervention aimed at reducing alcohol intake (*Drinktest* [15]) and took place among five experts as well as five Dutch adults who (occasionally) drink alcohol, with varying age and socioeconomic status.

The experts were identified through the first author's professional network and were considered experts when they had a track record in the field of motivation, health behavior, or Web-based health behavior change interventions. Experts were asked whether they thought the 19-item VCCQ measured the same concept as the HCCQ, whether items were properly reworded to fit a virtual care setting, and whether they thought the VCCQ was a comprehensive instrument to measure perceived autonomy-support in a virtual care setting. Furthermore, experts were asked to indicate any perceived ambiguities in wording and whether they felt any aspects of autonomy-support were not covered. Experts also gave feedback on the translation of the items from English to Dutch.

The Dutch adults were invited to complete the VCCQ after they had visited and consulted the Web-based intervention Drinktest, and reflected on whether the questions and instructions in the VCCQ were clearly formulated and whether they thought the questionnaire had an acceptable length . During their participation, they were asked to take notes, which were discussed with them afterwards.

Pretest Results

Most experts indicated that, generally due to the rewording of the HCCQ-items to a virtual care setting, the VCCQ items did not always measure the underlying concept of perceived support for autonomy. To ensure the VCCQ truly measured perceived autonomy-support in a virtual care setting, experts suggested to primarily use the content of the HCCQ-items as a basis and reformulate items when necessary, instead of a literal translation and rewording to the virtual context. Furthermore, experts indicated the VCCQ to be reasonably comprehensive. Yet, to ensure the VCCQ included a wide range of autonomy-supportive strategies, four additional items were included based on their feedback (items 20-23). Moreover, based on both expert input and feedback from the participating Dutch adults, the wording of 14 VCCQ-items was simplified to guarantee comprehensibility by the target group. This resulted in the final 23-item questionnaire presented in Table 1.

XSL•FO
RenderX

Table 1. The 23 items of the initial VCCQ (virtual care climate questionnaire) with their description and source.

No.	Item	Item description	Source
1	VCCQ_choice[a]	I feel that <name intervention> has provided me with choices and options	HCCQ[b]
2	VCCQ_understood	I felt understood by <name intervention>	HCCQ
3	VCCQ_honest	I am able to be honest and open when completing questions about <target behavior> on <name intervention>	HCCQ
4	VCCQ_confidence[a]	<name intervention> conveys confidence in my ability to <target behavior>	HCCQ
5	VCCQ_judgment[a]	I feel that <name intervention> does not judge me	HCCQ
6	VCCQ_knowledge	Because of <name intervention> I really understand what I need to do to <target behavior>	HCCQ
7	VCCQ_answers[a]	<name intervention> encourages me to search for answers to the questions that I have	HCCQ
8	VCCQ_trust[a]	I feel a lot of trust in <name intervention>	HCCQ
9	VCCQ_questions[a]	<name intervention> answers my questions fully and carefully	HCCQ
10	VCCQ_input[a]	<name intervention> allows me to provide input on how I would like to do things	HCCQ
11	VCCQ_emotions[a]	<name intervention> takes into account my emotions in the advice given	HCCQ
12	VCCQ_care[a]	I feel that <name intervention> cares about me as a person	HCCQ
13	VCCQ_communication	I don't feel very good about the way <name intervention> communicates with me	HCCQ
14	VCCQ_see[a]	<name intervention> tries to incorporate how I see things in the advice given	HCCQ
15	VCCQ_feelings	I feel that <name intervention> asks enough questions about my feelings	HCCQ
16	VCCQ_stimulant[a]	<name intervention> encourages me to <target behavior>	PASSES[c]
17	VCCQ_feedback	<name intervention> provides me with positive feedback when I do something to <target behavior>	PASSES
18	VCCQ_steering	I feel the advice of <name intervention> is directive	Experts
19	VCCQ_effective[a]	I feel that <name intervention> provides me with effective possibilities to <target behavior>	Experts
20	VCCQ_way[a]	<name intervention> gives me the feeling that I can choose a way to <target behavior> myself	Experts
21	VCCQ_idea[a]	<name intervention> explains to me why it would be a good idea to <target behavior>	Experts
22	VCCQ_use[a]	<name intervention> asks me the right questions about <target behavior>	Experts
23	VCCQ_must	I feel that <name intervention> is telling me what to do about <target behavior> without my having a say	Experts

[a]Items were included in the final VCCQ (virtual care climate questionnaire) based on the psychometric validation; the Dutch version of the final VCCQ is available upon request.

[b]HCCQ: health care climate questionnaire.

[c]PASSES: perceived autonomy support scale for exercise settings.

Psychometric Validation

To validate the VCCQ psychometrically, two studies were conducted. Study 1 was conducted in the context of a Web-based, computer-tailored intervention to reduce alcohol intake (Drinktest [15,17]) among Dutch adults who drink alcohol (occasionally). Study 2 was conducted in the context of a Web-based, computer-tailored intervention to reduce cannabis intake (*Weed-check* [18]) among a sample of Dutch students.

Sample and Procedure

Study 1

The sample consisted of 230 Dutch adults who drink alcohol (occasionally). Participants were recruited via the ISO-certified (International Organization for Standardization, ISO) research panel *PanelClix* [19] and, after providing informed consent, completed the study through the Web. Before filling out the VCCQ, participants visited and consulted Drinktest [17], a Dutch, Web-based, computer-tailored intervention to reduce alcohol intake [15]. When consulting this intervention, participants answered a variety of questions concerning their current alcohol consumption as well as their cognitions related to this behavior, based on which they received feedback that is tailored to their personal situation and beliefs. This feedback was presented to participants on their computer screen, and participants were also to send this feedback to an email address of their choice as well as to print it. To ensure that participants made proper use of the Web-based intervention, they were

XSL•FO
RenderX

instructed to imagine they wanted to reduce their alcohol intake and that they used the Web-based intervention to this end. For participants to be able to imagine wanting to reduce their alcohol intake, an inclusion criterion was set based on participants' alcohol intake. Participants who indicated never to have drunk alcoholic beverages before or those who had not drunk any alcoholic beverages in the past 12 months were excluded from participation—as were respondents who did not complete the survey entirely, had missing data on key variables, did not engage with the intervention, or participated for too long in the survey (ie, had z-scores >3 for participation time). After visiting and consulting the Web-based intervention, they were informed that they would be asked for their opinions about—and personal experiences with—this intervention in a subsequent questionnaire (ie, the VCCQ) and to answer several related questions regarding their motivation and perceived competence for changing. Participants were rewarded for their approximately 15-minute participation by means of 100 *Clix*, the usual incentive used by PanelClix, worth about €1.25.

Study 2

The sample consisted of 228 Dutch students using cannabis. Participants were recruited via the student participation website of the faculty of Social and Behavioral Sciences of the University of Amsterdam and, after providing informed consent, completed the study through the Web. As part of the survey, participants visited and consulted Weed-check [18], a Dutch, Web-based, computer-tailored intervention to reduce cannabis consumption. Similar as in the intervention used for Study 1, as part of their consultation of this intervention participants answered questions concerning their cannabis consumption and their cognitions associated with cannabis consumption, based on which they received feedback that was tailored to their personal situation and beliefs. Again, feedback was visible on the participants' computer screen, and options were provided to send the feedback to an email address and to print it. To ensure that participants made proper use of the Web-based intervention, they were instructed to imagine they wanted to have an overview of their cannabis usage and that they used the Web-based intervention to this end. In order for participants to be able to imagine wanting an overview of their cannabis consumption, an inclusion criterion was set based on participants' cannabis usage. Participants who indicated never to have used cannabis before or those who had used cannabis but only for medical reasons were excluded—as were respondents who did not complete the survey entirely, participated twice, indicated to have no understanding of the intervention, participated for too long in the survey, or had a too high number of neutral answers (ie, had z-scores >3 for participation time and the number of neutral answers, respectively). After visiting and consulting the Web-based intervention, participants were informed that they would be asked for their opinions about—and personal experiences with—this intervention and to answer several related questions regarding their motivation and perceived competence for changing. As a reward for their approximately 20-minute participation, students could choose their own incentive (ie, 0.5 research credits or € 2.50).

Measures

Both studies assessed several common and study-specific background variables, autonomous motivation to change, and perceived competence concepts, as detailed below.

Background Variables

Participants' age and gender were assessed via single items. Moreover, in Study 1 several characteristics of alcohol consumption were measured: the total number of weekdays consuming alcoholic beverages, average number of alcoholic beverages consumed during a typical weekday, total number of weekend days consuming alcoholic beverages, average number of alcoholic beverages consumed during a typical weekend day, and the total number of times consuming more than four alcoholic beverages in a day within the last 6 months. In Study 2, four characteristics of cannabis consumption were measured: number of times consuming cannabis in the last 12 months, cannabis use in the last 30 days, number of times consuming cannabis in the last 30 days, total number of joints on a typical day.

Autonomous Motivation

The degree to which one's motivation to change is autonomous (vs controlled) was measured by the treatment self-regulation questionnaire (TSRQ) [20], consisting of 16 items relating to autonomous or controlled motivation, or amotivation. The autonomous motivation subscale (6 items) showed good internal consistency (Study 1: Cronbach alpha=.95 and omega=.93; Study 2: Cronbach alpha=.91 and omega=.91), as did the controlled motivation subscale (7 items; Study 1: Cronbach alpha=.93 and omega=.93; Study 2: Cronbach alpha =.92 and omega=.92).

Perceived Competence for Changing

Participants' feelings of competence to change were measured by the perceived competence scale (PCS) [21] for reducing alcohol intake (Study 1) and cannabis intake (Study 2), respectively. The PCS asked participants the extent to which they agreed with the four statements, such as "I feel confident in my ability to reduce my alcohol intake" (Study 1), or "I am capable to reduce my cannabis consumption" (Study 2). All items were measured on a scale ranging from *strongly disagree* (1) to *strongly agree* (7), and the scale proved to be reliable (Study 1: Cronbach alpha=.95 and omega=.95; Study 2: Cronbach alpha=.93 and omega=.93). For divergent validity analysis purposes, in Study 2 perceived competence for learning was also assessed using the PCS [22] (Cronbach alpha=.94 and omega=.94).

Analyses

All statistical analyses were conducted using R statistical package [23]. Item selection followed several steps following the criteria described below.

Item Descriptives and Inter-Item Correlations

The first step was to investigate the distribution of responses to VCCQ items via descriptive statistics and examine the strength of association between items based on Spearman correlations. Items were flagged if they showed inadequate distributions (eg, <10 responses per response option [24]) or low associations

XSL•FO
RenderX

with other items (eg, nonsignificant correlations with other items), but were kept in further analyses for detailed diagnosis.

Nonparametric Item Response Theory Analyses

To examine the VCCQ's structural validity taking into account variations in item difficulty [25], nonparametric item response theory (NIRT) analyses (Mokken scale analysis, MSA) were conducted using the *mokken* package in R [26]. MSA examines whether an item set orders respondents accurately on a continuum representing a single latent trait (ie, in this study perceived autonomy-support). According to MSA, items can be considered a scale (fit the monotone homogeneity model) if three conditions are met: unidimensionality, monotonicity, and local independence (LI). If the items also meet a fourth criterion, invariant item ordering (IIO), the scale can be used for group comparison [25,27].

Unidimensionality entails that items can be located on a single latent continuum in terms of probabilities of obtaining high scores [25]. It was examined by assessing coefficients of homogeneity (*H*; ranging from 0 to 1, from no association to perfect association considering item distributions) and via an automated item selection procedure (*aisp*) algorithm [26]. Loevinger's scalability coefficients estimated item and scale homogeneity of the items (H_i coefficients and *H* coefficients) [28]. The *aisp* analysis performs an exploration of the scale unidimensionality at increasing levels of homogeneity [28]. Monotonicity implies that the probability of obtaining high scores on an item does not decrease as latent trait scores increase; *crit* values >40 are considered as violating monotonicity requirements [29] and then the recommended strategy is to remove the item with the most serious (highest) violation and rerun the analysis. LI means that associations between items are explained only by their relationship with the latent trait (perceived autonomy-support) [26]. At the moment, LI is still under research as a new MSA test [30]; until very recently, no tests were even available to test this assumption. Therefore, in this study no exclusion decisions were made based on LI results. If the items also show IIO, the scale represents a "person-free" item hierarchy in terms of their difficulty, that is, the order of items remains the same at different levels of the latent dimension [31]. LI and IIO are tested via dedicated functions that flag items that violate these criteria to be considered for exclusion. For monotonicity and IIO, minimum group size was set at 50; LI was examined with default parameter values.

Confirmatory Factor Analysis and Reliability

After selecting an item set fitting the MSA criterion, a confirmatory factor analysis (CFA) was conducted using the function *cfa* in the *lavaan* R package [32] to examine whether the unidimensional structure is supported by this alternative analysis, and thus allows comparisons with prior literature on autonomy-support scales; default parameter values were used (eg, maximum likelihood estimator, covariance matrix). Model diagnostics, parameter estimates, and goodness-of-fit indices were examined. Model fit was judged against the following criteria: Tucker-Lewis index (TLI) and Comparative Fit Index (CFI) >0.95; root mean square error of approximation (RMSEA) <0.06; and chi-square *P* value >.05 [33,34]. This analysis was followed by assessing the internal consistency of the final scale using Cronbach alpha and omega [35].

Construct Validity

The VCCQ's construct validity was determined by examining its convergent and divergent validity, based on expectations based on theory [3] and evidence from earlier research [8]. In terms of convergent validity, the VCCQ was expected to positively correlate with the autonomous motivation subscale of the TSRQ as well as with perceived competence for changing (ie, for reducing one's alcohol intake or cannabis consumption). In terms of divergent validity, the VCCQ was expected to show nonsignificant or negative correlations with the controlled motivation subscale of the TSRQ and with the nonhealth-related concept of perceived competence for learning (Study 2 only). Both studies had 90% power to detect bivariate correlations *r*=.21 and 80% power for *r*=.18, at alpha=.05.

Results

Samples

Sample characteristics of respondents participating in Study 1 are presented in Table 2. For Study 2, sample characteristics are presented in Table 3. For both studies, the flow of respondents is presented Multimedia Appendix 1.

XSL•FO
RenderX

Table 2. Demographic characteristics of respondents participating in Study 1 (N=230).

Variable	Response categories	n (%)
Gender	Female	110 (47.8)
	Male	120 (52.2)
Age, mean (SD)		46.09 (15.29)
Number of weekdays alcoholic beverages are used	4	49 (21.3)
	3	16 (7.0)
	2	40 (17.4)
	1	46 (20)
	Less than 1	42 (18.3)
	I never drink alcoholic beverages during weekdays	37 (16.1)
How many glasses do you usually drink during a weekday?	11 or more	3 (1.3)
	7-10	4 (1.7)
	6	9 (3.9)
	5	7 (3.0)
	4	11 (4.8)
	3	21 (9.1)
	2	79 (34.3)
	1	59 (25.7)
Number of weekend days alcoholic beverages are used	3	78 (33.9)
	2	68 (29.6)
	1	55 (23.9)
	Less than 1	29 (12.6)
How many glasses do you usually drink during a weekend day?	11 or more	8 (3.5)
	7-10	19 (8.3)
	6	10 (4.3)
	5	13 (5.7)
	4	31 (13.5)
	3	45 (19.6)
	2	69 (30.0)
	1	35 (15.2)
How often have you drunk four or more glasses in a day within the last 6 months?	Every day	17 (7.4)
	5-6 times per week	3 (1.3)
	3-4 times per week	18 (7.8)
	1-2 times per week	31 (13.5)
	1-3 times per month	34 (14.8)
	3-5 times per 6 months	25 (10.9)
	1-2 times per 6 months	43 (18.7)
	Never	59 (25.7)

XSL•FO
RenderX

Table 3. Demographics characteristics of respondents participating in Study 2 (N=228).

Variable	Response categories	n (%) or mean (SD)
Gender	Female	157 (68.9)
	Male	71 (31.1)
Age in years, mean (SD)		21.44 (2.30)
How often have you used cannabis in the last 12 months? n (%)	1 time	25 (11.0)
	2 times	45 (19.7)
	3 times	36 (15.8)
	4 times	24 (10.5)
	5 times or more	98 (43.0)
Used cannabis in the last 30 days, n (%)	Yes	118 (51.8)
	No	110 (48.2)
How often have you used cannabis in the last 30 days? n (%)	Daily	14 (6.1)
	More times per week	10 (4.4)
	At least 1 time per week	21 (9.2)
	Less than 1 time per week	73 (32.0)
Number of joints used on a typical day, mean (SD)		1.2 (0.88)

Item Descriptives and Inter-Item Correlations

In Study 1, participants' average scores on the 23 VCCQ-items ranged between 3.68 (SD 1.66) and 5.59 (SD 1.29) (Table 4). Scores covered the entire range on the scale from 1 to 7. All items were skewed toward agreement that the virtual care climate is autonomy-supportive, though varied in the extent to which agreement was expressed (Multimedia Appendix 2). Only two items had ≥10 answers in each of the two response categories at the lower end of the scale. Inter-item correlations ranged from −.11 to .79 (P<.001) (Multimedia Appendix 3) . Two negatively worded items (13 and 23) showed nonsignificant associations with other items.

In Study 2, participants' average scores on the 23 VCCQ-items ranged between 3.70 (SD 1.61) and 6.01 (SD 1.12) (Table 4). As in Study 1, scores covered the entire range of the scale. When compared with Study 1, items were less skewed toward agreement that the virtual care climate is autonomy-supportive; three items had <10 answers for the response category at the lower extreme and 14 items had <10 answers for the highest response category (Multimedia Appendix 2). Inter-item correlations (Spearman) ranged from −.22 (P<.001) to .67 (P<.001) (Multimedia Appendix 3) . The same two negatively worded items (13 and 23) as in Study 1 as well as item 18 and item 3 showed weak or no associations with other items.

Nonetheless, all items in both studies were included in the next steps for a more detailed diagnosis.

XSL•FO
RenderX

Table 4. Descriptive statistics for the 23 VCCQ (virtual care climate questionnaire) items in Study 1 (N=230) and Study 2 (N=228).

No.	VCCQ (virtual care climate questionnaire) items	Study 1		Study 2	
		Mean (SD)	Skew	Mean (SD)	Skew
1	VCCQ_choice	5.06 (1.32)	−0.34	4.46 (1.78)	−0.49
2	VCCQ_understood	4.97 (1.29)	−0.24	3.79 (1.69)	−0.13
3	VCCQ_honest	5.59 (1.29)	−0.75	6.01 (1.12)	−1.49
4	VCCQ_confidence	4.70 (1.32)	−0.19	4.01 (1.49)	−0.35
5	VCCQ_judgment	5.19 (1.37)	−0.54	4.58 (1.84)	−0.43
6	VCCQ_knowledge	4.78 (1.33)	−0.17	3.73 (1.51)	−0.10
7	VCCQ_answers	4.73 (1.34)	−0.13	3.70 (1.61)	−0.04
8	VCCQ_trust	4.96 (1.27)	−0.21	4.09 (1.58)	−0.30
9	VCCQ_questions	4.91 (1.24)	−0.11	4.21 (1.51)	−0.38
10	VCCQ_input	4.72 (1.21)	0.07	3.75 (1.50)	−0.11
11	VCCQ_emotions	4.96 (1.32)	0.02	4.03 (1.56)	−0.29
12	VCCQ_care	4.88 (1.32)	−0.21	3.86 (1.68)	−0.29
13	VCCQ_communication	3.68 (1.66)	0.05	3.85 (1.66)	0.09
14	VCCQ_see	4.70 (1.14)	0.24	4.10 (1.51)	−0.43
15	VCCQ_feelings	4.53 (1.33)	0.04	4.14 (1.47)	−0.24
16	VCCQ_stimulant	4.77 (1.33)	−0.13	4.19 (1.78)	−0.25
17	VCCQ_feedback	4.84 (1.22)	−0.01	4.28 (1.33)	−0.57
18	VCCQ_steering	4.59 (1.24)	−0.03	4.92 (1.42)	−0.68
19	VCCQ_effective	4.79 (1.18)	−0.14	4.11 (1.50)	−0.39
20	VCCQ_way	4.90 (1.21)	−0.20	4.28 (1.42)	−0.51
21	VCCQ_idea	5.08 (1.24)	−0.17	5.12 (1.45)	−0.94
22	VCCQ_use	5.10 (1.21)	−0.22	4.51 (1.59)	−0.62
23	VCCQ_must	4.19 (1.46)	−0.20	4.25 (1.65)	−0.05

Nonparametric Item Response Theory Analyses

In Study 1, items 13 and 23 violated the assumption of unidimensionality (ie, item *H*<.30; Table 5); therefore, these items were excluded from further analyses. Table 5 therefore shows results for the resulting 21-item unidimensionality analysis. All items and the total scale had scores >.30 (*H*=.660, standard error, SE=0.029), hence the 21-item VCCQ scale could be considered unidimensional. In other words, its items measured a single underlying concept, as intended. No significant violations of monotonicity were found for the 21-item VCCQ (ie, all *crit* values <.40). Violations of IIO were only found for item 6; therefore, this item was excluded. All 20 remaining items fitted the IIO criterion.

In Study 2, items 13, 23, 18, and 3 were found to have item *H*<.30 (Table 5) and were therefore excluded. Table 5 shows the results of the unidimensionality analysis of the remaining 19 items. All item *H* s and the scale *H* were >.30 (*H*=. 445, SE=0.030). The results confirmed that the 19-item VCCQ scale could be considered unidimensional. No significant violations of monotonicity were found (ie, all *crit* values <.40) Violations of IIO were found for items 15, 17, and 2, which were therefore excluded, resulting in a final 16-item VCCQ.

With regard to LI, in Study 1 items 11 and 14 and in Study 2 items 16, 20, and 21 were flagged as possibly violating LI – whereas these items were retained in the questionnaire in this study as LI is still under construction, these items may be considered for exclusion when developing a VCCQ short-form.

XSL•FO
RenderX

Table 5. Loevinger's scalability coefficients (*H*) for the 23-item and 21-item VCCQ (virtual care climate questionnaire) in Study 1 (N=230) and the 23-item and 19-item VCCQ in Study 2 (N=228).

No.	Description	Study 1		Study 2	
		23-item	21-item	23-item	19-item
		H_i(SE)[a]	H_i(SE)	H_i(SE)	H_i(SE)
1	VCCQ_choice	0.563 (0.043)	0.627 (0.045)	0.333 (0.037)	0.409 (0.042)[b]
2	VCCQ_understood	0.599 (0.034)	0.665 (0.035)	0.429 (0.028)	0.519 (0.031)[b]
3	VCCQ_honest	0.493 (0.047)	0.555 (0.050)	0.096 (0.045)	Excluded
4	VCCQ_confidence	0.604 (0.033)	0.659 (0.035)	0.366 (0.034)	0.431 (0.038)
5	VCCQ_judgment	0.596 (0.036)	0.667 (0.037)	0.345 (0.035)	0.434 (0.039)[b]
6	VCCQ_knowledge	0.620 (0.032)	0.688 (0.032)[b]	0.394 (0.035)	0.463 (0.040)
7	VCCQ_answers	0.607 (0.031)	0.671 (0.032)	0.402 (0.033)	0.462 (0.037)
8	VCCQ_trust	0.615 (0.035)	0.686 (0.035)	0.402 (0.031)	0.479 (0.035)
9	VCCQ_questions	0.643 (0.029)	0.713 (0.029)	0.412 (0.032)	0.488 (0.035)
10	VCCQ_input	0.622 (0.032)	0.681 (0.033)	0.392 (0.036)	0.471 (0.040)
11	VCCQ_emotions	0.594 (0.034)	0.662 (0.034)	0.412 (0.032)	0.492 (0.035)
12	VCCQ_care	0.638 (0.028)	0.707 (0.028)	0.399 (0.033)	0.480 (0.037)
13	VCCQ_communication	−0.011 (0.075)	Excluded	−0.043 (0.053)	Excluded
14	VCCQ_see	0.604 (0.034)	0.651 (0.037)	0.398 (0.034)	0.478 (0.038)
15	VCCQ_feelings	0.560 (0.036)	0.610 (0.038)	0.270 (0.039)	0.320 (0.045)
16	VCCQ_stimulant	0.596 (0.037)	0.649 (0.040)	0.321 (0.039)	0.381 (0.045)
17	VCCQ_feedback	0.626 (0.035)	0.685 (0.036)	0.360 (0.040)	0.406 (0.046)[b]
18	VCCQ_steering	0.511 (0.054)	0.548 (0.059)	0.151 (0.051)	Excluded
19	VCCQ_effective	0.638 (0.032)	0.691 (0.035)	0.406 (0.035)	0.473 (0.040)
20	VCCQ_way	0.654 (0.028)	0.719 (0.029)	0.402 (0.036)	0.486 (0.039)
21	VCCQ_idea	0.576 (0.040)	0.648 (0.041)	0.318 (0.041)	0.363 (0.045)
22	VCCQ_use	0.610 (0.042)	0.678 (0.045)	0.353 (0.038)	0.417 (0.043)
23	VCCQ_must	0.195 (0.077)	Excluded	−0.003 (0.054)	Excluded
	Scale	H (SE)	H (SE)	H (SE)	H (SE)
	VCCQ	0.547 (0.029)	0.660 (0.029)	0.321 (0.027)	0.445 (0.030)

[a]SE: standard error.

[b]Item was excluded based on violation of the IIO assumption.

Only 15 items remained in the VCCQ in both Study 1 and Study 2. For research as well as clinical purposes it is most desirable to keep the respondent burden of questionnaire completion as low as possible, and use the same items across interventions aimed at different behaviors. Therefore, we proceeded to the analyses of the questionnaire's reliability and construct validity with a 15-item version of the VCCQ. The remaining 15 items maintained the breadth of the theoretical concept. That is, an item was only excluded when its content showed overlap with other items (six instances), it was excluded based on item response theory (IRT) results in both studies (one instance), or when the item was retrospectively deemed inappropriate in the context of Web-based, computer-tailored interventions (one instance). Remaining items are presented in bold font in Table 1. Item step response functions of the 15 remaining items are presented in Multimedia Appendix 4.

Confirmatory Factor Analysis and Reliability

The CFA with data from Study 1 showed that the CFI of the one-factor model was 0.897, the TLI was 0.879, and the RMSEA was 0.128 (95% CI 0.115-0.140). In Study 2, results were similar: CFI=0.902; TLI=0.885; RMSEA=0.087 (95% CI 0.074-0.100). Chi-square tests were significant for both models (χ^2_{90}=426.6 and 244.8, P<.001). The scale showed good internal consistency both in Study 1 (Cronbach alpha=.97 and omega=.97, H=.66, mean 4.9 [SD 1.0]) and Study 2 (Cronbach alpha=.92 and omega=.92, H=.66, mean 4.9 [SD 1.0]). CFA diagrams are presented in Multimedia Appendix 5.

XSL•FO
RenderX

Construct Validity

Convergent Validity

In Study 1, the correlation between the final 15-item VCCQ and the autonomous motivation subscale of the TSRQ was strong and positive (r=.66, P<.001). That is, the extent to which participants perceived Drinktest to be autonomy-supportive—as measured by the VCCQ—was significantly and positively correlated with participants' autonomous motivation to reduce their alcohol intake. Furthermore, the correlation between the VCCQ and PCS for reducing alcohol intake was also strong and positive (r=.52, P<.001). The extent to which participants perceived Drinktest to be autonomy-supportive was significantly and positively correlated to participants' perceived competence for reducing their alcohol intake. These findings support the hypothesis of positive relationships between perceived autonomy-support in a virtual care setting and both autonomous motivations and perceived competence for reducing alcohol intake, indicating the convergent validity of the VCCQ.

In Study 2, the correlation between the final 15-item VCCQ and the autonomous motivation subscale of the TSRQ was moderately strong and positive (r=.37, P<.001), whereas the correlation between the VCCQ and the PCS for cannabis reduction was weak and not significant (r=.01, P=.12).

Divergent Validity

In line with expectations, results from Study 1 showed that the correlation between the VCCQ and the controlled motivation subscale of the TSRQ was moderate, though contrary to expectations this correlation was positive (r=.29, P<.001). The extent to which participants perceived Drinktest to be autonomy-supportive was significantly and positively correlated with participants' controlled motivations to reduce their alcohol intake—though the correlation was much weaker compared with the correlation between the VCCQ and autonomous motivation. Consequently, some support was found for the hypothesized divergent validity of the VCCQ.

In Study 2, similar results were found: the correlation between the VCCQ and the controlled motivation subscale of the TSRQ was moderately strong and positive (r=.37, P<.001)—this correlation had a similar strength as the correlation between the VCCQ and autonomous motivation. In line with expectations, the association between the VCCQ and the PCS for learning was weak and not significant (r=.05, P=.48), supporting the divergent validity of the VCCQ. Together, however, the results concerning the scale's divergent validity in this study were mixed.

Table 6 summarizes results for convergent and divergent validity in both studies.

Table 6. Pearson correlations between the virtual care climate questionnaire, the treatment self-regulation questionnaire autonomous and controlled motivation subscales, and the different perceived competence scales (N of Study 1=230; N of Study 2=228).

Results from correlation analyses	VCCQ[a] study 1	VCCQ study 2
Convergent validity		
TSRQ[b] autonomous motivation subscale	.66[c]	.37[c]
PCS[d] for reducing alcohol intake	.52[c]	N/A[e]
PCS for reducing cannabis consumption	N/A	.01(NS)[f]
Divergent validity		
TSRQ Controlled motivation subscale	.29[c]	.37[c]
PCS for learning	N/A	.05 (NS)

[a]VCCQ: virtual care climate questionnaire.

[b]TSRQ: treatment self-regulation questionnaire.

[c]P<.001.

[d]PCS: perceived competence scale.

[e]N/A: not applicable.

[f]NS: nonsignificant.

Discussion

Principal Findings

The objective of this study was to develop and validate the VCCQ: the first instrument specifically intended to measure perceived support for autonomy in a virtual care setting. On the basis of on NIRT analyses, we found that several items needed to be excluded from the questionnaire: three items in Study 1 and seven items in Study 2. The final 15-item version included items that were adjusted from the original HCCQ [7] and PASSES [8] items, as well as new Internet-specific items based on expert consultation. This suggests that, as expected, neither the HCCQ nor the PASSES would have sufficiently covered the concept of perceived autonomy-support in a virtual care setting and provide support for the approach that was taken for item development, that is, to use both existing questionnaires and expert consultation as input.

The final 15-item VCCQ was characterized by unidimensionality, monotonicity, and IIO, as required by NIRT. Therefore, the items can be used to compute a total score that reflects the respondents' relative positions on the latent construct of "perceived autonomy support," and these scores can also be

XSL•FO
RenderX

used for comparing groups. Most items also met the local independence criterion, though before any items can be excluded based on this new psychometric test, further research is needed. Yet the items flagged, based on the criterion of local independence in our studies, can be first candidates for exclusion from a short-form VCCQ. Furthermore, the CFA of the 15-item VCCQ showed an acceptable model fit consistent with MSA results. Fit values were slightly below the recommended thresholds [33,34], though there is a scientific debate on the use of and value that should be attached to these thresholds when determining a scale's model fit [36]. Reliability of the 15-item VCCQ was confirmed by two alternative tests (alpha and omega). Therefore, this version can be used confidently in future research, while continuing investigations into its measurement properties in different contexts and populations.

All three negatively worded items (ie, 13, 18, and 23) were excluded from the questionnaire due to low inter-item correlations in both studies. While including such items in questionnaires is recommended to reduce acquiescence bias, it may lead to response inconsistencies and may prove difficult to specify in measurement models. Therefore, some experts advise against their use [37,38]. Excluding these items, together with other misfitting items, resulted in good measurement properties of the current VCCQ version without limiting the construct breadth—an item was only excluded when its content showed overlap with other items, it was excluded from the scale in both studies, or was retrospectively deemed inappropriate in the context of Web-based, computer-tailored interventions. Future psychometric work on the VCCQ could investigate the risk of acquiescence in reporting on perceived autonomy support and possible solutions.

The scale's convergent validity was confirmed by positive associations with autonomous motivation for reducing alcohol intake and cannabis consumption, and with perceived competence for reducing alcohol intake. Divergent validity was confirmed by the nonsignificant association with perceived competence for learning, yet the expected negative or nonsignificant relationship with controlled motivation could not be confirmed. In fact, in both studies we found a significant positive association between perceived support for autonomy and controlled motivation, although weaker than the association with autonomous motivation in Study 1. Moreover, in both our studies, autonomous motivation showed a strong significant positive correlation with controlled motivation (r=.68 in Study 1; r=.61 in Study 2), supporting our idea that autonomous and controlled motivation may represent different motivational dimensions that can coexist within people. This has also been suggested in earlier SDT-based research into motivational profiles, suggesting not only a high quality (ie, high autonomous and low controlled motivation), but also a high quantity motivational profile (ie, high autonomous and high controlled motivation) that does not necessarily lead to inferior results in terms of learning and physical activity than the high-quality profile [39]. In addition, earlier research has shown that introjected motivation, a form of controlled motivation that is characterized by performance of a behavior to avoid feelings of guilt or anxiety or to attain ego enhancement such as pride [3], may have positive effects on health behavior change [40]. Thus, an increase in especially this type of controlled motivation from using an Web-based intervention may also be a desired outcome. In our studies, perceived support for autonomy as measured by the VCCQ showed a stronger association with the introjected motivation subscale (r=.35 in Study 1; r=.38 in Study 2) than with the extrinsic motivation subscale (r=.21 in Study 1; r=.34 in Study 2). Whether this can be the direct result of providing support for autonomy can, however, not be concluded from the present study due to its cross-sectional nature. Following this line of reasoning, an alternative explanation may be that both autonomously motivated and respondents with high levels of controlled motivation were included in the present studies, with both groups finding potentially different elements in the intervention as supportive of their autonomy. Longitudinal research that includes measurement of the different types of motivation both before and after participation in a virtual care intervention is needed to examine whether perceiving higher autonomy support as measured by the VCCQ indeed increases autonomous motivation and reduces or increases the different forms of controlled motivation, and to determine whether autonomy-supportive interventions differ in their effectiveness for people with different types of motivation—in terms of both motivational and behavioral outcome measures. In addition, longitudinal research may be used to obtain evidence related to the questionnaire's predictive validity.

Strengths and Limitations

An important strength of this study is that before proceeding to classical test theory analyses to determine the VCCQ's validity and reliability, we investigated the questionnaire's structure using NIRT analyses [26]. A known limitation of factor analytic methods is that they do not take into account differences in item difficulty, which are considered in IRT analyses, including MSA [40]. Moreover, as other parametric methods, they attempt to estimate quantitative differences between respondents, and are therefore unnecessarily restrictive for constructs that only refer to differences in degree [26]; perceived autonomy support is arguably among these constructs and more adequately investigated via MSA. So far, not many studies have used NIRT analyses in addition to classical test theory analyses and we are unaware of any previous studies that did so in the context of SDT-based questionnaires.

The VCCQ was developed to measure perceived support for autonomy in a variety of virtual care settings, targeting different kinds of health behavior. Yet, in the present study it was tested for its reliability and validity solely in the context of Web-based interventions aimed at two addictive behaviors, namely alcohol consumption and cannabis use. To increase the generalizability of the results to interventions targeting other nonaddictive health behaviors such as physical activity and healthy dietary behaviors, future research could consider testing the validity and reliability of the VCCQ in the context of other Internet-based interventions. Moreover, as both studies described were cross-sectional, test-retest reliability and predictive validity of the VCCQ could not be ascertained. To investigate these properties, longitudinal research designs should be considered. An additional limitation pertains to the use of a scenario (ie, to ensure that participants made proper use of the Web-based interventions, they were instructed to imagine they wanted to

XSL•FO
RenderX

change their behavior and that they used the Web-based intervention to this end). Whereas this may limit the ecological validity of the results presented, the aim of this specific study was to develop and validate a novel questionnaire, and not necessarily to actually help people to improve their health behavior. As a consequence, the exposure of people who truly wanted to change their behavior to interventions that were not necessarily developed to be autonomy-supportive was deemed undesirable. One of the uses of VCCQ is for guiding intervention development, and in such cases our choice of sample is actually recommended: people who are similar enough to the target group to be able to relate to the content, but not require the intervention as such (and thus be in a more vulnerable position). Due to the inclusion criteria set, the participants in the present studies can be considered very well-capable of imagining themselves wanting to change their behavior and of evaluating the interventions in terms of their autonomy-supportiveness. Therefore, the scenario used can be considered rather realistic. Another limitation that warrants attention concerns the use of an odd number of answering categories within the VCCQ, as also used in the HCCQ and PASSES, providing respondents with the possibility to equivocate when completing the questionnaire. In both studies, we prevented this from having a large influence on the psychometric analyses, by excluding respondents with a too high number of neutral answers (ie, respondents who had z-scores >3 for the number of neutral answers), though as is shown by the answering patterns presented in Multimedia Appendix 2 neutral answers were still rather common. The optimal number of response categories for questionnaires assessing support for autonomy in different contexts (virtual or face-to-face) is therefore a valuable research question for further studies. Finally, responding to the final 15 VCCQ-items within multimeasure surveys may still place a significant burden on respondents, potentially increasing nonresponse. Future research may therefore consider the development and validation of a VCCQ-short form, following examples from earlier research where the 15-item HCCQ was transformed into a valid and reliable six-item IOCQ (important other climate questionnaire) [41] and a six-item mHCCQ (modified HCCQ adapted for breast cancer patients) [42].

Conclusions

By developing the VCCQ, this study aimed to address the lack of validated tools that measure support for autonomy in the context of virtual care settings. Scientifically, the VCCQ enables further research into the role of autonomy-support in virtual care settings without virtual health care providers involved, a context in which evidence for its positive effects is currently lacking. Practically, using the VCCQ will help to identify how Web-based health behavior change interventions can most successfully support autonomy, subsequently increase autonomous forms of motivation, and ultimately promote a healthy lifestyle. Health behavior change interventions are increasingly provided via the Internet [9], and people tend to search for health-related information first and foremost through this medium [10]. While limited resources are available for the widespread implementation of effective lifestyle interventions and health care decision makers should select interventions based on their cost-effectiveness, Web-based health behavior change interventions are likely to be highly competitive when compared with other types of interventions [43,44]. Investigating the role of autonomy support in these settings is necessary, as it may lead to further improvements in the effectiveness of such interventions.

The VCCQ appeared to accurately assess participants' perceived autonomy-support offered by two Web-based health behavior change interventions. Overall, the scale showed the expected properties and relationships with relevant concepts and the studies presented suggest this first version of the VCCQ to be reasonably valid and reliable. As a result, the current version should be used cautiously in future research and practice to measure perceived support for autonomy within a virtual care climate. However, future research efforts are required that focus on further investigating the VCCQ's divergent validity, on determining the VCCQ's validity and reliability when used in the context of Web-based interventions aimed at improving nonaddictive or other health behaviors, and on developing and validating a VCCQ-short form. A Dutch version is available from the authors upon request.

Acknowledgments

We would like to thank all experts and members of the target group for their participation in pretesting the VCCQ. Furthermore, we would like to thank Andrea Perez Ortiz for her help in conducting the pretest and in the data collection and preliminary analyses for Study 1. This study was supported by the Amsterdam School of Communication Research (ASCoR) and the Innovational Research Incentives Scheme Veni from NWO-MaGW (Netherlands Organization for Scientific Research—Division for the Social Sciences) accredited to the first author (project number 451-15-028).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Flow of respondents who started participation.

[PDF File (Adobe PDF File), 97KB - jmir_v19i5e155_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2

Answering patterns for all 23 VCCQ (virtual care climate questionnaire) items.

[PDF File (Adobe PDF File), 358KB - jmir_v19i5e155_app2.pdf]

Multimedia Appendix 3

Inter-item correlations.

[PDF File (Adobe PDF File), 181KB - jmir_v19i5e155_app3.pdf]

Multimedia Appendix 4

Item Step Response Functions.

[PDF File (Adobe PDF File), 135KB - jmir_v19i5e155_app4.pdf]

Multimedia Appendix 5

Confirmatory factor analysis diagrams.

[PDF File (Adobe PDF File), 243KB - jmir_v19i5e155_app5.pdf]

References

1. World Health Organization. The World Health Report 2002: Reducing Risks, Promoting Healthy Life. Geneva: World Health Organization; 2002.
2. Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. J Public Health (Oxf) 2011 Dec;33(4):527-535 [FREE Full text] [doi: 10.1093/pubmed/fdr033] [Medline: 21562029]
3. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
4. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
5. Deci EL, Eghrari H, Patrick BC, Leone DR. Facilitating internalization: the self-determination theory perspective. J Pers 1994 Mar;62(1):119-142. [Medline: 8169757]
6. Williams GC, Cox EM, Kouides R, Deci EL. Presenting the facts about smoking to adolescents: effects of an autonomy-supportive style. Arch Pediatr Adolesc Med 1999 Sep;153(9):959-964. [Medline: 10482213]
7. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [Medline: 11846344]
8. Hagger MS, Chatzisarantis NL, Hein V, Pihu M, Soós I, Karsai I. The perceived autonomy support scale for exercise settings (PASSES): Development, validity, and cross-cultural invariance in young people. Psychology of Sport and Exercise 2007 Sep;8(5):632-653. [doi: 10.1016/j.psychsport.2006.09.001]
9. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? a systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
10. Van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
11. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
12. Williams GC, Niemiec CP, Elliot AJ, LaGuardia JG, Gorin AA, Rigby CS. Virtual Look AHEAD program: initial support for a partly virtualized intensive lifestyle intervention in type 2 diabetes. Diabetes Care 2014 Aug;37(8):e169-e170 [FREE Full text] [doi: 10.2337/dc14-0831] [Medline: 25061144]
13. Patrick H, Canevello A. Methodological overview of a self-determination theory-based computerized intervention to promote leisure-time physical activity. Psychol Sport Exerc 2011 Jan;12(1):13-19 [FREE Full text] [doi: 10.1016/j.psychsport.2010.04.011] [Medline: 21103069]
14. Smit ES, de Vries H, Hoving C. Effectiveness of a web-based multiple tailored smoking cessation program: a randomized controlled trial among Dutch adult smokers. J Med Internet Res 2012;14(3):e82 [FREE Full text] [doi: 10.2196/jmir.1812] [Medline: 22687887]

XSL•FO
RenderX

15. Boon B, Risselada A, Huiberts A, Riper H, Smit F. Curbing alcohol use in male adults through computer generated personalized advice: randomized controlled trial. J Med Internet Res 2011;13(2):e43 [FREE Full text] [doi: 10.2196/jmir.1695] [Medline: 21719412]
16. Schulz DN, Kremers SP, Vandelanotte C, van Adrichem MJ, Schneider F, Candel MJ, et al. Effects of a web-based tailored multiple-lifestyle intervention for adults: a two-year randomized controlled trial comparing sequential and simultaneous delivery modes. J Med Internet Res 2014;16(1):e26 [FREE Full text] [doi: 10.2196/jmir.3094] [Medline: 24472854]
17. Drinktest [Drinktest]. Utrecht, the Netherlands: Trimbos Institute URL: http://www.drinktest.nl [accessed 2016-06-15] [WebCite Cache ID 6iHJ5mMwy]
18. Wietcheck [Weed-check]. Utrecht, the Netherlands: Trimbos Institute URL: http://www.wietcheck.nl/index.cfm?act=home.tonen [accessed 2016-06-15] [WebCite Cache ID 6iHJFGCrb]
19. Panelclix. Amsterdam, The Netherlands: PanelClix/iFlavours BV URL: http://www.panelclix.nl [accessed 2016-06-15] [WebCite Cache ID 6iHJ2h3dX]
20. Levesque CS, Williams GC, Elliot D, Pickering MA, Bodenhamer B, Finley PJ. Validating the theoretical structure of the treatment self-regulation questionnaire (TSRQ) across three different health behaviors. Health Educ Res 2007 Oct;22(5):691-702 [FREE Full text] [doi: 10.1093/her/cyl148] [Medline: 17138613]
21. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998 Oct;21(10):1644-1651. [Medline: 9773724]
22. Williams GC, Deci EL. Internalization of biopsychosocial values by medical students: a test of self-determination theory. J Pers Soc Psychol 1996 Apr;70(4):767-779. [Medline: 8636897]
23. R Core Team. R-project. Vienna, Austria: R Foundation for Statistical Computing R: A language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2016-07-12] [WebCite Cache ID 6iwcBrW5U]
24. Linacre JM. Investigating rating scale category utility. J Outcome Meas 1999;3(2):103-122. [Medline: 10204322]
25. van Schuur WH. Mokken scale analysis: between the Guttman scale and parametric item response theory. Polit Anal 2003;11(2):139-163. [doi: 10.1093/pan/mpg002]
26. Van der Ark LA. Mokken scale analysis in R. J Stat Soft 2007;20(11):11-19. [doi: 10.18637/jss.v020.i11]
27. Sijtsma K, Meijer RR, van der Ark AL. Mokken scale analysis as time goes by: An update for scaling practitioners. Pers Individ Dif 2011 Jan;50(1):31-37. [doi: 10.1016/j.paid.2010.08.016]
28. Hemker BT, Sijtsma K, Molenaar IW. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken I RT model. Appl Psychol Meas 1995 Dec 01;19(4):337-352. [doi: 10.1177/014662169501900404]
29. Sijtsma K, Molenaar I. Introduction to nonparametric item response theory. Thousand Oaks, CA: SAGE Publications; 2002.
30. Sijtsma K, Straat J, van der Ark LA. Goodness-of-Fit Methods for Nonparametric IRT Models. In: Quantitative Psychology Research, Springer Proceedings in Mathematics & Statistics 140. Switzerland: Springer International Publishing; 2015:109-120.
31. Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. BMC Med Res Methodol 2012;12:74 [FREE Full text] [doi: 10.1186/1471-2288-12-74] [Medline: 22686586]
32. Rosseel Y. Lavaan: an R package for structural equation modeling. J Stat Soft 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
33. Jackson DL, Gillaspy JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: an overview and some recommendations. Psychol Methods 2009 Mar;14(1):6-23. [doi: 10.1037/a0014694] [Medline: 19271845]
34. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 2009;6(1):1-55. [doi: 10.1080/10705519909540118]
35. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014 Aug;105(3):399-412. [doi: 10.1111/bjop.12046] [Medline: 24844115]
36. Marsh H, Hau K, Wen Z. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Struct Equ Modeling 2004 Jul;11(3):320-341. [doi: 10.1207/s15328007sem1103_2]
37. Swain SD, Weathers D, Niedrich RW. Assessing three sources of misresponse to reversed likert items. J Mark Res 2008 Feb;45(1):116-131. [doi: 10.1509/jmkr.45.1.116]
38. Wong N, Rindfleisch A, Burroughs JE. Do reverse-worded items confound measures in cross-cultural consumer research? the case of the material values scale. J Consum Res 2003 Jun 01;30(1):72-91. [doi: 10.1086/374697]
39. Vansteenkiste M, Sierens E, Soenens B, Luyckx K, Lens W. Motivational profiles from a self-determination perspective: the quality of motivation matters. J Educ Psychol 2009;101(3):671-688. [doi: 10.1037/a0015083]
40. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
41. Williams GC, Lynch MF, McGregor HA, Ryan RM, Sharp D, Deci EL. Validation of the “Important Other” climate questionnaire: assessing autonomy support for health-related change. Fam Syst Health 2006;24(2):179-194. [doi: 10.1037/1091-7527.24.2.179]

XSL•FO
RenderX

42. Shumway D, Griffith KA, Jagsi R, Gabram SG, Williams GC, Resnicow K. Psychometric properties of a brief measure of autonomy support in breast cancer patients. BMC Med Inform Decis Mak 2015;15:51 [FREE Full text] [doi: 10.1186/s12911-015-0172-4] [Medline: 26155944]
43. Smit ES, 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;15(3):e57 [FREE Full text] [doi: 10.2196/jmir.2059] [Medline: 23491820]
44. Schulz DN, Smit ES, Stanczyk NE, Kremers SP, de Vries H, Evers SM. Economic evaluation of a web-based tailored lifestyle intervention for adults: findings regarding cost-effectiveness and cost-utility from a randomized controlled trial. J Med Internet Res 2014;16(3):e91 [FREE Full text] [doi: 10.2196/jmir.3159] [Medline: 24650860]

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
HCCQ: health care climate questionnaire
IIO: invariant item ordering
IOCQ: important other climate questionnaire
IRT: item response theory
ISRF: item step response function
LI: local independence
MSA: Mokken scale analysis
NIRT: nonparametric item response theory
PASSES: perceived autonomy support scale for exercise settings
PCS: perceived competence scale
RMSEA: root mean square error of approximation
SDT: self-determination theory
TLI: Tucker-Lewis index
TSRQ: treatment self-regulation questionnaire
VCCQ: virtual care climate questionnaire

Edited by G Eysenbach; submitted 29.09.16; peer-reviewed by K Zhao, L Vera, J Hensel, M Amato, G Papandonatos; comments to author 29.12.16; revised version received 19.02.17; accepted 20.03.17; published 08.05.17.

Please cite as:
Smit ES, Dima AL, Immerzeel SAM, van den Putte B, Williams GC
The Virtual Care Climate Questionnaire: Development and Validation of a Questionnaire Measuring Perceived Support for Autonomy in a Virtual Care Setting
J Med Internet Res 2017;19(5):e155
URL: http://www.jmir.org/2017/5/e155/
doi:10.2196/jmir.6714
PMID:28483743

©Eline Suzanne Smit, Alexandra Lelia Dima, Stephanie Annette Maria Immerzeel, Bas van den Putte, Geoffrey Colin Williams. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Development, Validation, and User Evaluation of Foodbook24: A Web-Based Dietary Assessment Tool Developed for the Irish Adult Population

Claire M Timon[1*], PhD; Richard J Blain[1*], MSc; Breige McNulty[1], PhD; Laura Kehoe[2], MSc; Katie Evans[2], MSc; Janette Walton[2], PhD; Albert Flynn[2], PhD; Eileen R Gibney[1], PhD

[1]UCD, Institute of Food and Health, University College Dublin, Dublin, Ireland

[2]UCC, School of Food and Nutritional Sciences, University College Cork, Cork, Ireland

[*]these authors contributed equally

Corresponding Author:
Eileen R Gibney, PhD
UCD
Institute of Food and Health
University College Dublin
Sceince Centre South
Befield
Dublin, Dublin 4
Ireland
Phone: 353 17162819
Fax: 353 17161147
Email: eileen.gibney@ucd.ie

Abstract

Background: The application of technology in the area of dietary assessment has resulted in the development of an array of tools, which are often specifically designed for a particular country or region.

Objective: The aim of this study was to describe the development, validation, and user evaluation of a Web-based dietary assessment tool "Foodbook24."

Methods: Foodbook24 is a Web-based, dietary assessment tool consisting of a 24-hour dietary recall (24HDR) and food frequency questionnaire (FFQ) alongside supplementary questionnaires. Validity of the 24HDR component was assessed by 40 participants, who completed 3 nonconsecutive, self-administered 24HDR using Foodbook24 and a 4-day semi-weighed food diary at separate time points. Participants also provided fasted blood samples and 24-hour urine collections for the identification of biomarkers of nutrient and food group intake during each recording period. Statistical analyses on the nutrient and food group intake data derived from each method were performed in SPSS version 20.0 (SPSS Inc). Mean nutrient intakes (and standard deviations) recorded using each method of dietary assessment were calculated. Spearman and Pearson correlations, Wilcoxon Signed Rank and Paired *t* test were used to investigate the agreement and differences between the nutritional output from Foodbook24 (test method) and the 4-day semi-weighed food diary (reference method). Urinary and plasma biomarkers of nutrient intake were used as an objective validation of Foodbook24. To investigate the user acceptability of Foodbook24, participants from different studies involved with Foodbook24 were asked to complete an evaluation questionnaire.

Results: For nutrient intake, correlations between the dietary assessment methods were acceptable to very good in strength and statistically significant (range r=.32 to .75). There were some significant differences between reported mean intakes of micronutrients recorded by both methods; however, with the exception of protein (P=.03), there were no significant differences in the reporting of energy or macronutrient intake. Of the 19 food groups investigated in this analysis, there were significant differences between 6 food groups reported by both methods. Spearman correlations for biomarkers of nutrient and food group intake and reported intake were similar for both methods. A total of 118 participants evaluated the acceptability of Foodbook24. The tool was well-received and the majority, 67.8% (80/118), opted for Foodbook24 as the preferred method for future dietary intake assessment when compared against a traditional interviewer led recall and semi-weighed food diary.

XSL•FO
RenderX

Conclusions: The results of this study demonstrate the validity and user acceptability of Foodbook24. The results also highlight the potential of Foodbook24, a Web-based dietary assessment method, and present a viable alternative to nutritional surveillance in Ireland.

(J Med Internet Res 2017;19(5):e158) doi:10.2196/jmir.6407

KEYWORDS

diet records; Internet; validity; biomarkers; method acceptability; adults

Introduction

Background

Dietary assessment methodologies are known to have both strengths and limitations [1]. Some of the methodological caveats among current dietary assessment methods include the participant burden, reliance on participant's honesty and ability to remember food and drinks consumed, and their individual portion sizes [2,3]. Cost, particularly when large-scale epidemiological studies and national nutrition surveys are concerned, can be another limiting factor [1]. The 24-hour dietary recall (24HDR) method is associated with low participant burden and can provide reliable intake data with minimal bias [3]. However, recalls can be expensive, time consuming to administer, and require skilled nutritionists or dietitians [4].

The application of technology in dietary assessment has made it possible to minimize the reliance on trained interviewers and instead facilitate automated self-administered 24HDR via Web-based platforms and mobile phone apps [5]. Web-based methodologies facilitate the collection of dietary intake across many geographic locations [6], from large cohorts [7], and are often preferred by participants compared with the traditional methods [8,9]. An example of a successful Web-based 24-hour recall tool is the ASA24 developed by the National Cancer Institute, USA. From its launch in 2009, more than 200 researchers have used ASA24 to carry out over 120,000 recalls [4].

Biomarker Analysis

A prerequisite for the acceptance and use of such Web-based dietary assessment tools is their validity. It is vital that new tools and methods measure what they are designed to measure. Assessing the relative validity of a new method or tool can be achieved by comparing intakes recorded by a new method to intakes derived from a method that is deemed more accurate [10]. The advent of biomarker analysis now also offers an objective measure of intake, which may overcome the bias associated with self-reported data [11]. Biomarkers of both nutrient and food intake can be analyzed in plasma, serum, and urine to indicate both short- and long-term intake and can provide an objective validation of dietary assessment tools as they reflect, but are independent of food intake. Although a feasible validation tool, biomarker analysis is not always included in the validation of new dietary assessment techniques, which is perhaps in part due to the invasive nature of sample collection and associated cost. Another common validation reference used is direct observation of participants during eating occasions that can then be compared with reported or recalled dietary intake data [12,13]. However, this too can be costly and can often take place in a laboratory setting, potentially influencing an individual's choices.

In a recent review of dietary assessment or tracker apps for mobile phones [14], the authors concluded that very few of the apps identified were based on scientifically valid nutrient composition databases and few had consulted nutrition professionals in the development process. With such unprecedented access to health and nutrition information the needs for scientifically validated, Web-based methods of dietary assessment are essential. The aim of this study was to describe the development, validation, and user evaluation of Ireland's first Web-based, self-administered 24HDR tool "Foodbook24."

Methods

Foodbook24

The Development of Foodbook24

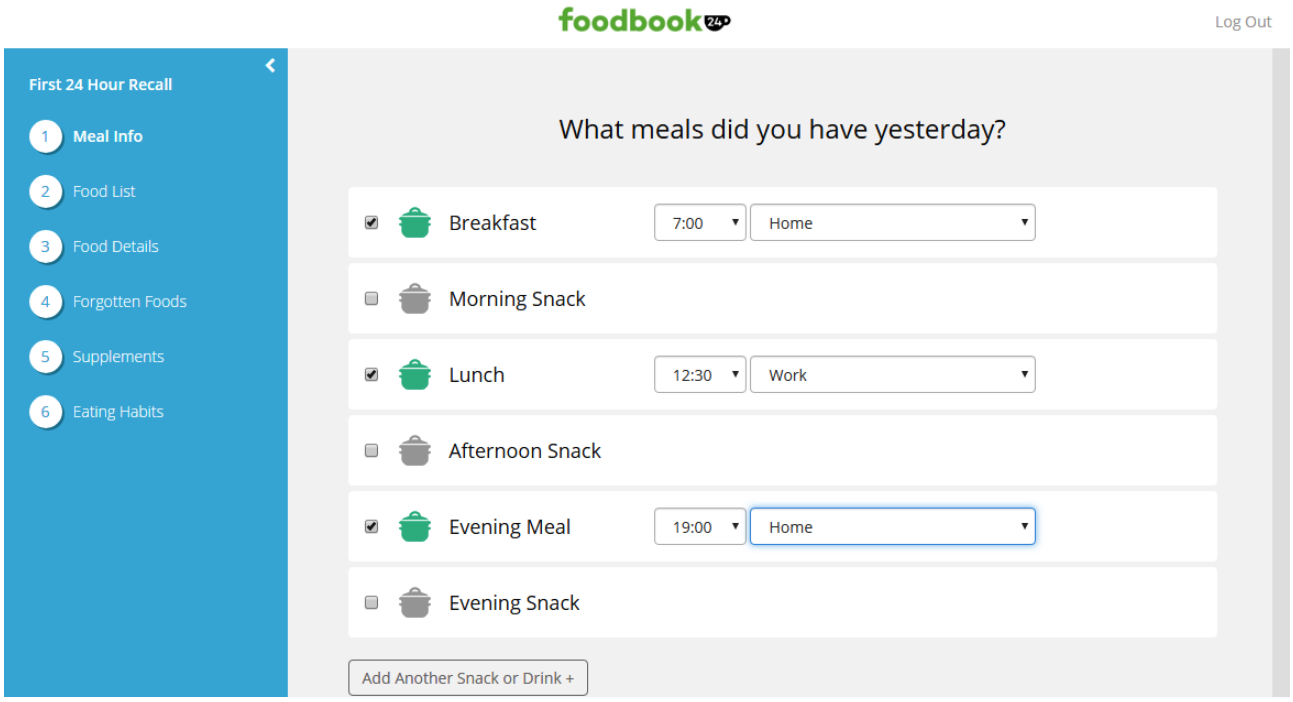
The design of the Foodbook24 tool was informed by guidelines issued on the collection of dietary information that can be used to estimate nutrient intake and to assess exposure to biological agents and chemical substances by the European Food Safety Authority in 2009 [15]. In addition, interviews with key stakeholder organizations or institutions in Ireland and an extensive review of the literature concerning Web-based dietary assessment platforms were conducted to further inform the design of Foodbook24. The final proposed design of Foodbook24 was a self-administered, Web-based tool consisting of different independent components that facilitate the collection of dietary intake data without direct interaction with a researcher. These components include a screening and consent stage, demographic questionnaire, 2x24-hour multiple pass recall (administered on nonconsecutive days), food frequency, and food choice questionnaires, and finally a tool evaluation questionnaire. All of these stages occur at predetermined time points and have been developed independently of each other, meaning different parts of the tool could potentially be activated or deactivated depending on the requirements of any given survey or study.

For the dietary recall component of Foodbook24, the user is required to complete multiple passes (as described by Moshfegh et al [16]) to report their dietary intake for the previous 24-hour period. Initially, the user lists the meals and snacks consumed the previous day, reports the times that these meals were consumed (as depicted in Figure 1), and also the location of food preparation. The user then adds individual food and drink items to each of the defined meals or snacks using a free text search function to select food and drink items from a predefined database. Further questions known as "completeness of

XSL•FO
RenderX

collection mechanisms" are presented to the user such as probe or linked food options and portion size information is then determined by selecting relevant amounts or portion size photographs. Finally, the user is presented with a review of selected items, a list of frequently forgotten foods, and queried about nutritional supplement intake and whether the reported intake was representative of usual intake. To populate the content of the Foodbook24 tool, various databases, completeness of collection mechanisms, and questionnaires were developed. These included a food list, nutrient composition, nutritional supplements and portion size databases, completeness of collection mechanisms for the 24-hour recall component and various supplementary questionnaires such as demographic and food choice questionnaires. The processes and considerations surrounding these components of the tool are described below.

Figure 1. The meal information stage of Foodbook24.



Food List and Nutrient Composition

The food and drink list used in Foodbook24 is a shortened list of food and drinks consumed in the Irish National Adult Nutrition Survey (NANS 2008-2010) [17]. The food composition data linked to the NANS dataset are derived from UK food composition tables [18] and the Irish Food Composition Database (IFCDB) [19]. The reduction process of the list involved the merging of food codes of a similar description and/or composition [20]. The aim of the reduction process was to reduce the food list that participants would have to search through to describe their dietary intake, thus reducing participant burden without compromising the nutrient composition output. This process significantly reduced the total number of food and drinks from 2552 to 751 individual items. An investigation into the agreement of the shortened food list to the original comprehensive list is reported elsewhere [20], but overall shows excellent agreement and was therefore deemed appropriate for inclusion in Foodbook24. The food and drink items were grouped into 58 different food groups and further categorized into 18 categories.

Completeness of Collection Mechanisms

On review of Web-based 24HDR tools, the use of "probe" and "linked" (as described by Foster [21]) food options are commonly used to ensure the complete capture of dietary intake data. Linked food options were added to 132 food and drink items within the Foodbook24 food and drink list (an example of a linked food option is highlighted in Figure 2). These options are linked to the primary selection and are a list of options known to be commonly consumed with the primary selection. The use of "probe questions," that is, questions posed to a respondent based on their primary food or drink selection provides more detail and further classifies that selection were implemented for 123 food and drink items (an example of a probe food question is depicted in Figure 3, where the user is asked to clarify whether the food item was homemade or retail). To improve the user experience of searching for food and drink items, "food tags" were applied to 484 of the 751 food and drink items. As the search function in this tool was based on the actual description of the food or drink item, "slang words" or brand names were tagged to the parent food to address common misspellings and multiple names of various food and drink items, for example, searching for "Houmus" would still retrieve "Hummus."

XSL•FO
RenderX

Figure 2. The "linked food" options available for the primary food selection of chips within Foodbook24.

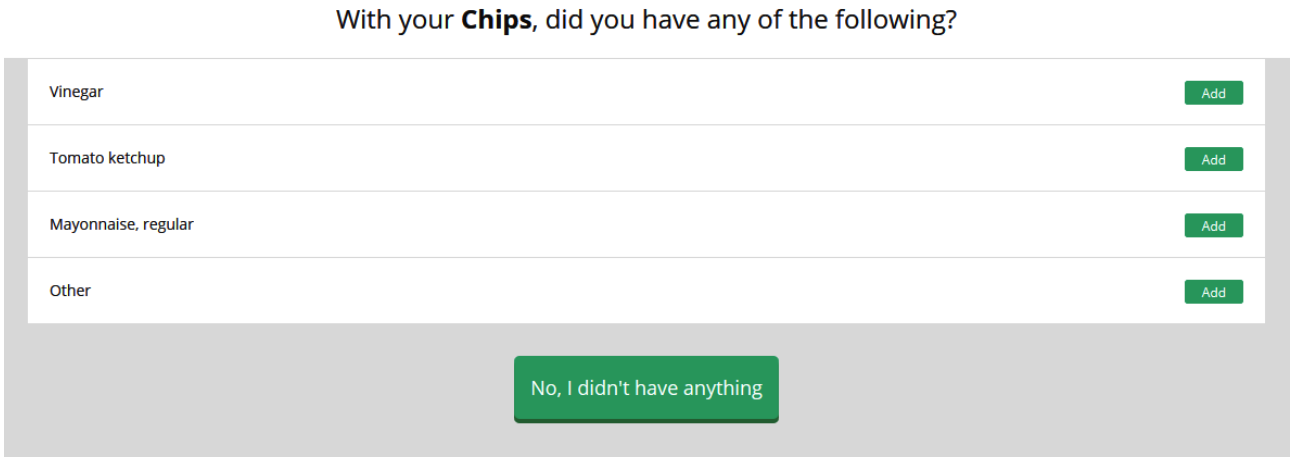
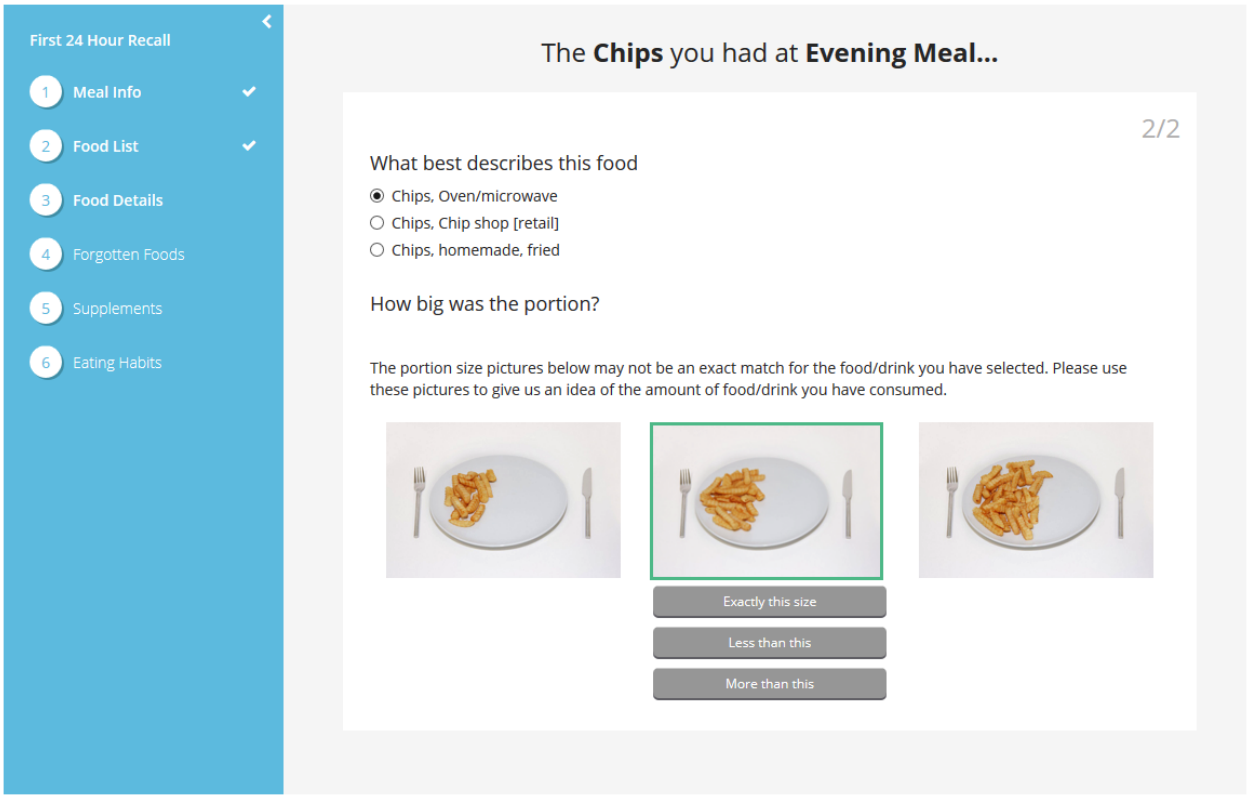


Figure 3. The food details stage of Foodbook24, presenting the user with a "probe question" and portion size images.



Portion Size

Portion size can be estimated in two different ways within Foodbook24. Building on an existing dataset of portion size images (created as part of the Food4me project) [6], 96 additional portion size images were created for Foodbook24. The range of food and drink weights for which the portion size images depict was based on ranges of weights consumed in NANS [17]. For the majority of food and drink items, there are a set of 3 portion size images representing small, medium, and large portions (although these terms were not alluded to in the tool) as shown in Figure 3. The respondent also had the option to select "less than this" for any of the 3 images in a set, "exactly this size" or "greater than this." Midpoint weights consumed in the NANS survey (eg, between small and medium portion size) underpin these options also. In Foodbook24, there are 174 sets of portion size images totaling 531 individual portion size images. For 195 food and drink items, average portion sizes from the Ministry for Agriculture, Fisheries and Food [22] were used. These are generally food items served in units such as cream crackers or biscuits; these were also followed by a question regarding how many serving of this item were consumed.

Nutritional Supplements

A database of 542 branded nutritional supplements with related nutrient composition was compiled to feature in the Foodbook24 tool. This nutritional supplement data consisted of supplements recorded as part of NANS [17] and those that were recorded as part of the Food4me study [6] by participants. These nutritional

XSL•FO
RenderX

supplements were grouped into 26 supplement categories, for example, zinc supplements contains several different brands of zinc supplements for participants to choose from. An unknown or generic supplement composition option was also created using the median nutrient composition of all those supplements in that category.

Supplementary Questionnaires

The following components feature in the overall tool design; however, these components were not utilized in the validation study. A food frequency questionnaire (FFQ) was compiled for inclusion in the Foodbook24 tool based on EU Menu guidelines [15] and the Food4me FFQ [6]. The final questionnaire included 81 items and used the same frequency responses as featured in EPIC FFQ [23]; however, the Foodbook24 FFQ did not collect information on portion size. The FFQ was included in the tool to capture the intake of food and drink items less commonly consumed rather than contribute to nutrient intake. The food choice questionnaire [24] was also included as a multidimensional measure of motives related to food choice. The user is presented with a statement related to food choice such as, "It is important to me that the food I eat on a typical day contains a lot of vitamins and minerals" and can then agree or disagree with this using a 7-point scale. Screening, demographic, and evaluation questionnaires were built into the tool alongside a study information sheet and consent form, the contents of which can be easily changed depending on the use of the tool.

Validation Study

Recruitment and Inclusion and Exclusion Criteria

Ethical approval for the study was obtained through University College Dublin (UCD) Human Research Ethics Committee (LS-15-27-Gibney-Timon). Participants were recruited via email using UCD mailing lists, University societies, and posters around campus. Individuals who expressed an interest in the study were contacted by phone and screened for eligibility. Subjects were eligible if the individuals were aged 18-64 years, fluent in English, had regular access to the Internet, were not pregnant, did not have any disease or condition that required chronic therapeutic nutritional or medical treatment, and had not been enrolled in or completed a degree, MSc or PhD in Human Nutrition. In total, 55 participants signed-up to take part in the study; however, 15 dropped out, which left a final sample size of 40.

Study Design

Participants were required to visit the Institute of Food and Health, UCD on three separate study visits during the study duration. At the first visit, informed consent, demographic information, and anthropometric measurements (including weight and body fat percentage and height) were collected. After the first visit, participants completed 3 nonconsecutive, unannounced, self-administered 24HDRs using the Web-based Foodbook24 tool. For this study, only the 24-h dietary component of the tool was used to record dietary intakes. Portion size photographs embedded in the tool depict a range of weights reported in NANS were used as portion size assessment aids. Emails were sent to participants on the morning and they were required to complete a recall using Foodbook24 without prior notice. In the middle of data collection using Foodbook24, participants attended study visit 2 and provided a fasting blood sample and a 24-hour urine collection. Following a 10-day wash out period, participants completed a 4-day semi-weighed food diary using a Tanita digital scale (KD-400) to weigh food and drink consumption as often as possible. On completion of this, participants attended the final study visit where they provided an additional fasting blood sample and 24-hour urine collection, completeness of check on their food diary, and completed a study evaluation questionnaire.

Collection of Biological Samples

Both blood and urine samples were collected from each participant in order to analyze specific biomarkers of nutrient intake. Blood samples (2x6 mL) were collected into lithium heparin tubes following a 12 h fasting period. Samples were spun for 15 min at relative centrifugal force (RCF) 1500 at 4°C. Plasma was transferred to labeled microtubes in 500 µL aliquots, two of which contained 10% meta-phosphoric acid (MPA) for the stabilization of ascorbic acid. All plasma samples were then frozen at −80°C. Plasma samples were analyzed by Vitas Analytical Services (Norway) for the determination of plasma ascorbic acid, carotenoids, and fatty acid content. Participants also provided a 24-hour urine collection and were instructed to collect the sample according to the protocol outlined in the National Diet and Nutrition Survey [25]. The urine samples were subsequently analyzed for urinary urea (as an indicator of protein intake [26]) using Daytona RX Clinical Analyzer (Randox, Nishinomiya, Japan) and urinary sodium, potassium, and creatinine were measured using the *Cobas Integra 700 Analyzer* (Roche Diagnostics) by the Department of Clinical Chemistry at St Vincent's University Hospital. The ratio of observed over expected urinary creatinine excretion (UCE; creatinine index) [27] alongside other criteria such as reported >1 missed void and samples with a total volume <0.5 L was used to exclude incomplete 24-hour urine collections from biomarker analysis [28].

Data and Statistical Analysis

Foodbook24 automatically generates a food and nutrient intake output for each user. The data from the semi-weighed food diaries was manually entered into WISP version 3 (Tinuviel Software, Anglesey, UK) by a single researcher in an attempt to maintain consistency and were reviewed independently by another researcher. Nutrient outputs for the semi-weighed food diaries were then generated. Mean daily nutrient intakes, standard deviations, and descriptive statistics (demographic data and evaluation questionnaire data) were computed in SPSS (version20) to determine the validity and user acceptability of Foodbook24. The normality of the data was assessed using the Shapiro-Wilk test before investigating the agreement between the dietary assessment methods, and parametric or nonparametric tests were used accordingly for subsequent analysis. Pearson and Spearman coefficient analyses was used to investigate the agreement between both methods in the reporting of nutrient intake, and to investigate the relationship between reported nutrient and food group intake and biological markers of intake. Correlation analysis was performed on energy

XSL•FO
RenderX

adjusted data (nutrient intakes were energy-adjusted, that is, the percentage of energy intake for macronutrients and gram per milligram per milligram (g/mg/mg) per 10 MJ energy intake for micronutrients). Deattenuated correlation coefficients were also computed by multiplying the initial coefficient by R_1, this was calculated as follows: $R_1=R_0\sqrt{(1((sw^2)/(sb^2))/n)}$, where $(sw^2)/(sb^2)$ is the ratio of the within- and between-person variances and n is the number of replicates per person for the given variable. The within- and between-person variances were obtained from an analysis of variance (ANOVA) model. Correlations coefficients were considered as follows: very good (0.7 and greater), good (0.5-0.69), acceptable (0.3-0.49), and poor (0.3 or less) [29].

The relative agreement between Foodbook24 and the food diary was assessed using cross-classification of nutrient and food group intakes to estimate the percentage of participants who were classified by the two methods into quartiles of "exact agreement," "exact agreement plus adjacent," "disagreement," and "extreme disagreement." Bland and Altman [30] analysis was performed to assess the limits of agreement in the reporting of macronutrient intake, considering the two methods of dietary assessment to be comparable if greater than 95% of the data plots were within the limits of agreement. Wilcoxon Signed Rank and Paired Student *t* test were used to identify the differences in the nutrient intake, and independent samples *t* test were used to compare daily food group intakes reported by both methods.

Evaluation of Foodbook24

Participants involved in two studies that investigated both the comparability of Foodbook24 (relative to an interviewer led 24-hour recall, results of which are not included in this publication) and the validity of Foodbook24 (compared with a 4-day semi-weighed food diary) were asked to complete an evaluation questionnaire once the study had concluded. In total, 118 participants (58 male and 60 female aged between 18 and 62 years) completed the optional questionnaire, 40 participants from the validation study and 79 from the comparison study. The design of the evaluation questionnaire was based around questionnaires used in similar studies that investigated the user acceptability of technology-based dietary assessment tools [31,32]. The questionnaire consisted of a 16-item evaluation questionnaire that was administered online. The focus of the questionnaire was to assess the participant's overall experience using the 24-hour recall component of the tool only (as participants did not use other components of the tool, eg, FFQ as part of these studies), and their acceptability of some of the software design features, method preference, and future use.

Results

Study Population

A total of 55 participants signed-up to complete the validation study, of which 15 participants withdrew and therefore did not complete the entire study (dropout rate of 27%), with N=40 completing. Of those that withdrew, 9 reported that the collection of biological samples was too burdensome, 3 could not attend the study visits due to prior commitments, and 3 did not disclose their reason for dropping out. This left a final sample size of 40 participants that completed the study; however, 1 participant was excluded from the analysis as he or she did not follow the study protocol correctly. Table 1 displays the demographic characteristics of participants (n=39). The mean of age of participants was 32 years (age range 18-62 years). Over half of participants were either employed as staff (46.2%) or enrolled as students (10.3%) in UCD. The remainder of participants were either employed locally to the university or heard about the study through an Irish volunteer website.

XSL•FO
RenderX

Table 1. Demographic characteristics of participants.

Demographic characteristics	Mean (SD) or n (%)
Age and BMI[a] , mean (SD)	
Age (years)	32.2 (13.4)
BMI[a] (kg/m²)	24.40 (3.75)
Gender, n (%)	
Female	20 (51)
Male	19 (49)
Occupation, n (%)	
Student	4 (10)
University staff	18 (46)
Employed outside of the University	16 (41)
Unemployed	1 (3)
Smoking habits, n (%)	
Smoker	4 (10)
Nonsmoker	29 (74)
Ex-smoker	6 (16)
Medical conditions, n (%)	
None	25 (65)
One or more	14 (35)

[a]BMI: body mass index.

[b]SD: standard deviation.

Comparison of Nutrient Intake Reported by Both Methods of Dietary Assessment

The unadjusted, mean daily intakes for energy, nutrients, and food groups recorded using Foodbook24 and a semi-weighed food diary are displayed in Multimedia Appendix 1 and Table 2. The energy adjusted correlations, deattenuated correlations, mean difference, and the limits of agreement (2 standard deviations of the mean) between the two methods for the reporting of nutrients are also displayed in Multimedia Appendix 1. For nutrient intake, the majority of correlations between the dietary assessment methods ranged from acceptable to very good, and are statistically significant (range r=.32 to .75).

XSL•FO
RenderX

Table 2. Food group intakes recorded by participants using the Foodbook24 tool and a 4-day semi-weighed food diary.

Food group	Foodbook24 (g), mean (SD[a])	Food diary (g), mean (SD)	*P* value
Grains, rice, pasta, and savories	207 (111.5)	171 (127.1)	.20
Bread and rolls	90.4 (54.54)	102 (46.29)	.31
Breakfast cereals	119 (83.9)	86.7 (80.99)	.09
Biscuits, cakes, and pastries	52.9 (36.67)	50.7 (44.68)	.82
Milk and yogurt	192 (228.4)	346 (223.4)	.05
Creams, ice creams, and desserts	67.0 (70.5)	73.7 (70.86)	.8
Cheeses	25.2 (13.40)	35 (16.1)	<.01[b]
Butter, spreading fats, and oils	14.93 (7.74)	10.42 (9.77)	<.05.[b]
Eggs and egg dishes	98.3 (69.61)	80.2 (52.68)	.26
Potatoes and potato dishes	135 (95.4)	147 (128.5)	.66
Veg and veg dishes	172 (107.5)	237 (152.5)	<.05[b]
Fruit and fruit dishes	372 (264.3)	252 (130.5)	<.01[b]
Fish and fish dishes	54.6 (58.23)	101 (102.0)	.05
Meat and meat products	244 (141.9)	249 (179.7)	.62
Alcoholic beverages	1314 (1208.2)	667 (761.5)	.09
Beverages other (sugar-sweetened)	1855 (1160.1)	1516 (854.7)	.14
Sugars, confectionary, preserves, and savory snacks	81.9 (50.81)	54.4 (47.63)	<.01[b]
Soups, sauces, and miscellaneous foods	91.8 (96.85)	130 (120.0)	.13
Nuts, seeds, herbs, and spices	16.8 (20.24)	37.4 (36.61)	<.001[b]

[a]SD: standard deviation.

[b]Significant difference in the reporting of food group intake between the two dietary assessment methodologies as defined by independent samples *t* test.

However, some correlation coefficients were not statistically significant including monounsaturated fat (r=.308, n=39, P=.05). Of the 34 nutrients investigated, there were significant differences between the reported mean intakes of 11 nutrients reported by the two methods; however, with the exception of protein (P=.02), there were no significant differences in the reporting of energy and macronutrient intake. Deattenuated correlation coefficients were higher, but the improvement was modest with the exception of intakes of fat (g/d).

Bland and Altman (Figures 4-8) analysis was used to further investigate the agreement between and the semi-weighed food diary. For macronutrients, Foodbook24 reported slightly lower intakes than the food diary; however, 95% or more of the data cases fell within the limits of agreement suggesting that the methods provide comparable intakes of these nutrients. The cross-classification of mean energy and nutrient intakes reported by the two methods are displayed in Table 3. The percentage of participants classified in “exact agreement” category of intake by both methods varied from 26% (% energy from saturated fat) to 74% (zinc). The majority of participants were classified in the “exact agreement and adjacent” category of intake by both methods with percentages varying from 69% (carotene) to 92% (zinc, potassium, and sodium). The percentages of participants that were classified into the “extreme disagreement” were low; and for some nutrients (protein, niacin, potassium, and sodium), no participants were classified into this category.

XSL•FO
RenderX

Table 3. Cross-classification of quartiles of mean energy and nutrient intake derived from Foodbook24 and a 4-day semi-weighed food diary.

Nutrient	Exact agreement (%)[a]	Exact agreement and adjacent (%)[b]	Disagreement (%)[c]	Extreme disagreement (%)[d]
Energy (kcal/d)	46.2	84.6	7.7	7.7
% Energy carbohydrate	30.8	76.9	20.5	2.6
% Energy protein	41.0	84.6	12.8	2.6
% Energy total fat	28.2	82.1	10.3	7.7
% Energy saturated fat	26.3	73.7	21.1	5.3
Protein (g/d)	56.4	84.6	7.7	0
Carbohydrate (g/d)	46.2	82.1	10.3	7.7
Sugars (g/d)	33.3	79.5	12.8	7.7
Starch (g/d)	43.6	84.6	10.3	5.1
Dietary fiber (g/d)	41.0	79.5	17.9	2.6
Fat (g/d)	38.5	74.4	17.9	7.7
Saturated fat (g/d)	43.6	76.9	17.9	5.1
Monounsaturated fat (g/d)	33.3	74.4	20.5	5.1
Polyunsaturated fat (g/d)	43.6	79.5	15.4	5.1
Retinol (µg/d)	30.8	71.8	20.5	7.7
Carotene (µg/d)	35.0	69.2	28.2	2.6
Vitamin D (µg/d)	33.3	87.2	10.3	2.6
Vitamin E (mg/d)	38.5	82.1	15.4	2.6
Riboflavin (mg/d)	43.6	83.7	12.8	2.6
Niacin (mg/d)	51.3	87.2	12.8	0
Vitamin B6 (mg/d)	41.0	84.6	12.8	2.6
Vitamin B12 (µg/d)	28.2	82.1	10.3	7.7
Folate (µg/d)	46.2	84.6	12.8	2.6
Vitamin C (mg/d)	41.0	79.5	12.8	7.7
Calcium (mg/d)	35.9	79.5	15.4	5.1
Magnesium (mg/d)	41.0	84.6	12.8	2.6
Iron (mg/d)	38.5	79.5	17.9	2.6
Copper (mg/d)	38.5	82.1	15.4	2.6
Zinc (mg/d)	74.4	92.3	5.1	2.6
Potassium (mg/d)	51.3	92.3	7.7	0
Sodium (mg/d)	46.2	92.3	7.7	0

[a]Exact agreement: percentage of cases cross-classified into the same quartile.

[b]Exact agreement and adjacent: percentage of cases cross-classified into the same or adjacent quartile.

[c]Disagreement: percentage of cases cross-classified 2 quartiles apart.

[d]Extreme disagreement: percentage of cases cross-classified 3 quartiles apart.

XSL•FO
RenderX

Figure 4. Bland and Altman plot examining the mean difference in reporting of energy intake by the two methods.

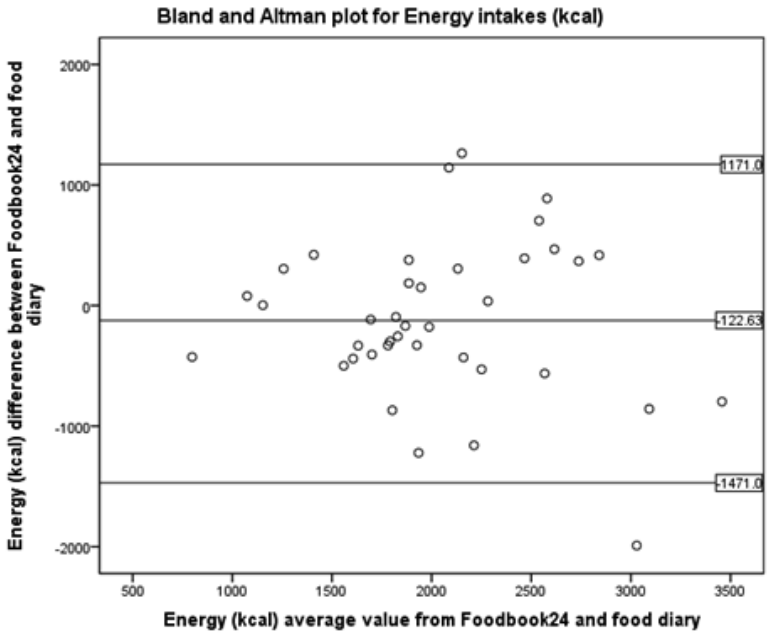
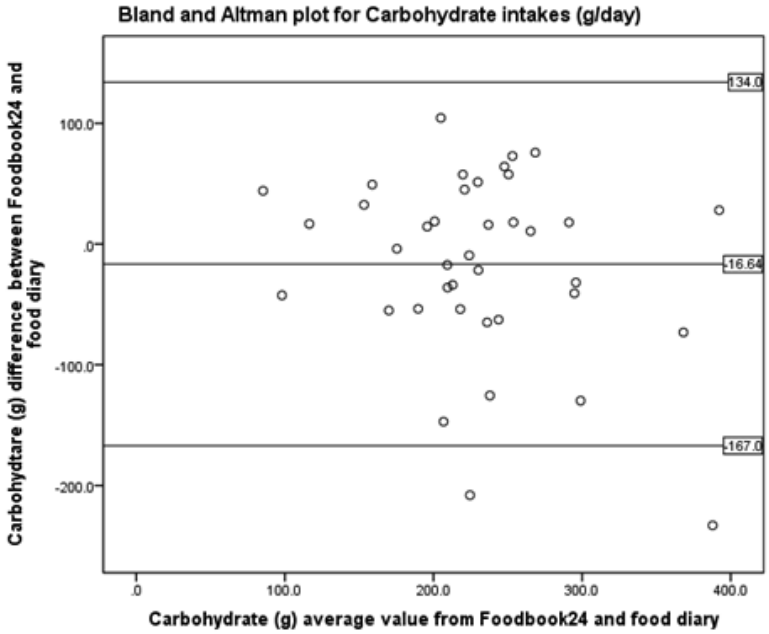


Figure 5. Bland and Altman plot examining the mean difference in reporting of carbohydrate intake by the two methods.



XSL•FO
RenderX

Figure 6. Bland and Altman plot examining the mean difference in reporting of protein intake by the two methods.

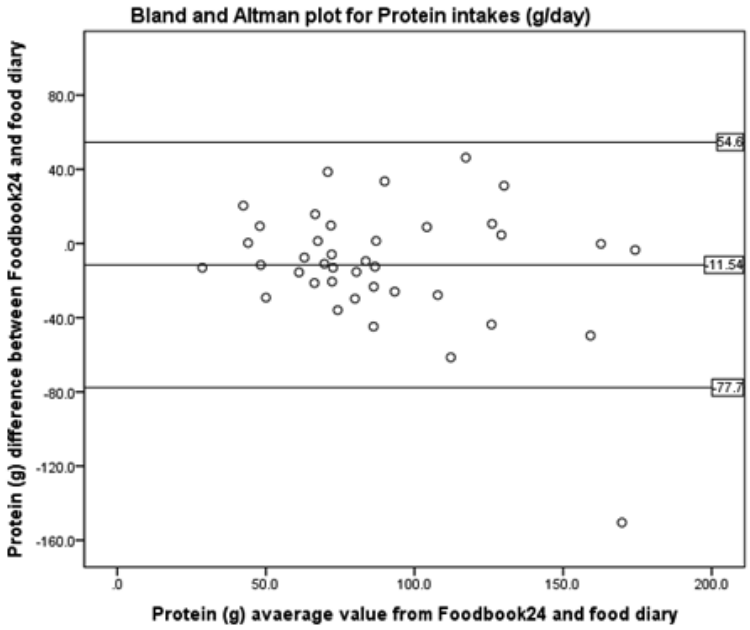
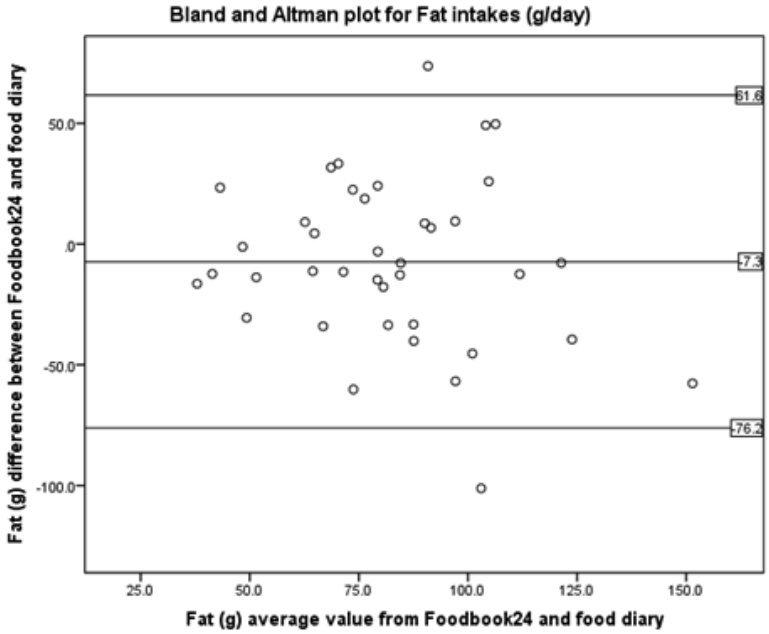
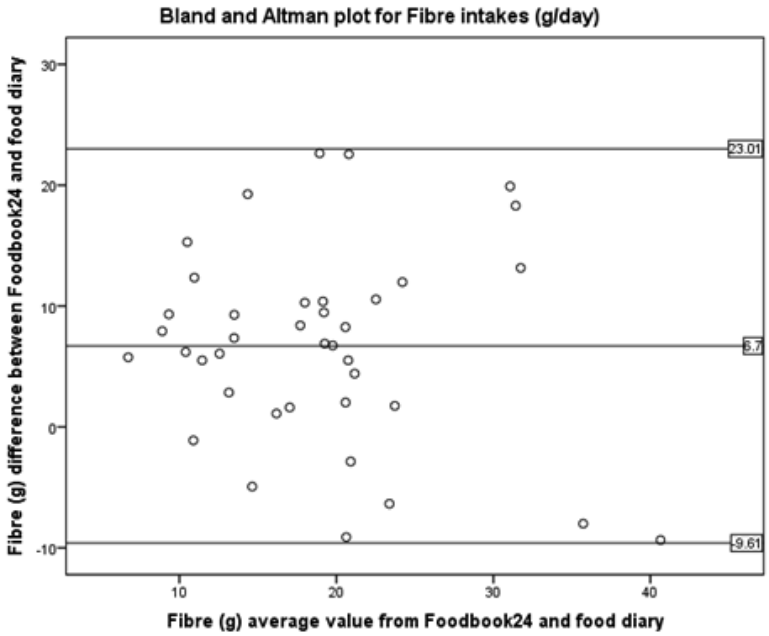


Figure 7. Bland and Altman plot examining the mean difference in reporting of fat intake by the two methods.



XSL•FO
RenderX

Figure 8. Bland and Altman plot examining the mean difference in reporting of fiber intake by the two methods.



Comparison of Food Group Intakes Reported by Both Methods of Dietary Assessment

The mean food group intakes reported by both methods and the significant difference in the reporting of food group intake between the methods is presented in Table 2. Of the 19 food groups investigated, there were significant differences between the reporting of 6 food groups including "fruit and fruit dishes," "alcoholic beverages," and "sugars and confectionary, preserves, and savory snacks." The cross-classification of mean food group intakes reported by the two methods is displayed in Table 4. The percentage of participants classified in "exact agreement" category of intake by both methods varied from 25.6% ("sugars, confectionary, preserves, and savory snacks" and "soups, sauces, and miscellaneous foods") to 79% ("creams, ice creams, and desserts"). The majority of participants were classified in the "exact agreement and adjacent" category of intake by both methods; however, in the "creams, ice creams, and desserts" food group, no participants were classified into this category. Similar to the results for nutrients intakes, the percentages of participants that were classified into the "extreme disagreement" were low; and for some food groups (eg, alcoholic beverages), no participants were classified into this category.

Comparison of Biological Markers of Intake Against Food and Nutrient Intake Reported by Methods of Dietary Assessment

The relationships between urinary (recovery) and plasma (concentration) biomarkers and nutrient and food group intake recorded by both methods are reported in Table 5. Despite the researcher's best efforts to recruit individuals that did not take supplements, this was not possible in every case. The researchers did ask participants not to take nutritional supplements during the study, where possible. As a result, participants who reported taking supplements, including protein, multivitamin, vitamin, mineral and/or fish oil, before the provision of blood and urine samples during this study were excluded from analysis. This excluded 15 participants from plasma biomarker analysis (resulting in n=24 for plasma biomarker analysis) and 11 participants from urinary biomarker analysis (resulting in n=28 for urinary biomarker analysis). With the exception of comparison of fruit and vegetable intakes (g/d) derived from Foodbook24 and total plasma carotenoids (r=.315, n=28, P=.10), there were good, significant correlations (r=.42 to .64) between food group and nutrient intakes reported by Foodbook24 and biomarkers of nutrient and food group intake from plasma and urine samples. Nutrient and food group intakes derived from the semi-weighed food diary compared with biomarkers from urine and plasma samples resulted in strong, significant correlations except in the case of urinary potassium.

XSL•FO
RenderX

Table 4. Cross-classification of quartiles of mean food group intake derived from Foodbook24 and a 4-day semi-weighed food diary.

Food group	Exact agreement (%)[a]	Exact agreement and adjacent (%)[b]	Disagreement (%)[c]	Extreme disagreement (%)[d]
Grains, rice, pasta, and savories	43.6	71.8	17.9	10.3
Bread and rolls	35.9	82.1	7.7	10.3
Breakfast cereals	28.2	71.8	17.9	10.3
Biscuits, cakes, and pastries	28.2	61.5	30.8	5.1
Milk and yogurt	35.9	79.5	17.9	2.6
Creams, ice creams, and desserts	79.5	0	20.5	0
Cheeses	48.7	74.4	12.8	12.8
Butter, spreading fats, and oils	38.5	66.7	25.6	7.7
Eggs and egg dishes	48.7	71.8	25.6	2.6
Potatoes and potato dishes	38.5	76.9	20.5	2.6
Veg and veg dishes	46.2	87.2	10.3	2.6
Fruit and fruit dishes	35.9	79.5	17.9	2.6
Fish and fish dishes	51.3	79.5	20.5	0
Meat and meat products	61.5	92.3	7.7	0
Alcoholic beverages	63.2	15.8	21.0	0
Beverages other (sugar-sweetened)	48.7	41.0	10.3	0
Sugars, confectionary, preserves, and savory snacks	25.6	71.8	17.9	10.3
Soups, sauces, and miscellaneous foods	25.6	76.9	12.8	10.3
Nuts, seeds, herbs, and spices	35.9	74.4	17.9	7.7

[a]Exact agreement: percentage of cases cross-classified into the same quartile.

[b]Exact agreement and adjacent: percentage of cases cross-classified into the same or adjacent quartile.

[c]Disagreement: percentage of cases cross-classified 2 quartiles apart.

[d]Extreme disagreement: percentage of cases cross-classified 3 quartiles apart.

Table 5. Biomarker and food group relationship (as derived from Foodbook24 and food diary).

Biomarkers[a]	Nutrient or food group[b]	Foodbook24 correlation coefficient (*r* value)[c]	Diary correlation coefficient (*r* value)[c]
Concentration biomarkers		(n=28)	(n=34)
Plasma ascorbic acid (µM)	Fruit and veg (g/d)	0.421	0.505
Plasma total carotenoids (µmol/L)	Fruit and veg (g/d)	0.315[e]	0.671
Plasma omega-3 index	Fish (g/d)	0.468	0.769
Plasma ascorbic acid (µM)	Vitamin C (mg/d)	0.518	0.605
Recovery biomarkers		(n=34)	(n=33)
Urinary urea (mmol/d)	Protein (g/d)	0.645	0.824
Urinary potassium (mmol/d)	Potassium (mg/d)	0.542	0.269[e]
Urinary sodium (mmol/d)[d]	Sodium (mg/d)	0.56	0.476

[a]Plasma and urinary biomarkers identified from fasted blood samples and 24-hour urine collections (respectively) collected from participants directly after recording intake using either dietary assessment methods.

[b]The data refer to nutrient and food group intakes recorded using Foodbook24 and the semi-weighed food diary.

[c]Spearman correlation coefficient.

[d]Sodium excretion measures were corrected to account for 90% excretion of all sodium consumed (28).

[e]Not statistically significant correlations (P<.05).

XSL•FO
RenderX

User Evaluation of Foodbook24

The main results of participants' evaluation of the 24-hour recall component of Foodbook24 are depicted in Table 6. The majority of respondents were very positive in their evaluation of Foodbook24. Overall, the majority found the Foodbook24 system user-friendly with 69.5% (82/118) reporting it easy or "OK" to use. When asked if participants felt that Foodbook24 changed what they ate and drank, a majority of 62.7% (74/118) felt it did not change at all, whereas 34.7% (41/118) felt it changed it a little, and 2.5% (3/118) felt it changed a lot. Importantly, when asked if there were any foods or drinks that participants did not want to record, a majority of 95.8% (113/118) stated "no."

Table 6. Participant acceptability of Foodbook24.

Question posed to participant	Participant responses			
Impact of Foodbook24 on diet	Changed a lot (%)	Changed a little (%)	Did not change at all (%)	
	2.5	34.7	62.7	
Completion time	Too long (%)	Okay (%)	Short (%)	Very short (%)
	6.8	63.6	22.9	6.8
User friendliness	Difficult (%)	Okay (%)	Easy (%)	Very easy (%)
	3.4	33.1	36.4	27.4
Remembering to use Foodbook24	Difficult (%)	Okay (%)	Easy (%)	Very easy (%)
	6.8	38.1	40.7	14.4
Preferred method	Foodbook24 (%)	Reference method[a] (%)	Other (%)	
	67.8	31.4	0.8	
Use Foodbook24 for longer	1 week (%)	1 month (%)	6 months (%)	No (%)
	27.1	30.5	24.6	17.8

[a]Two different reference methods for comparison and validation study. This equates to 38.5% (30/78) in favor of the traditional interview-led 24HDR and 17.5% (7/40) for semi-weighed food diary.

Participants were asked about using Foodbook24 for longer periods of time to gain insight into the potential long-term use of the tool. The results were favorable for the shorter time of a week (considering the completion of two 24-hour recalls per week) with 82.2% willing to use Foodbook24 for a week, persistence understandably decreased with 55.1% willing to use for a month, and 24.6% for 6 months. When asked to select which method participants would prefer to use in future (Foodbook24 vs the respective reference method), 67.8% (80/118) opted for Foodbook24 and 31.4% (27/118) for the reference method. The researchers were aware that these particular results may have been influenced by the different reference method used in either study (interviewer-led 24HDR in the comparison study and a 4-day semi-weighed food diary in the validation study). The participant burden associated with a 4-day semi-weighed food diary may have been greater than that of an interview-led 24HDR. As a result, responses were split by study involvement. A majority of 61.5% (48/78) opted for Foodbook24 when compared with 38.5% (30/78) in favor of the interview-led 24HDR. Foodbook24 was an even clearer preference when compared with the semi-weighed food diary with 80% (32/40) in favor of Foodbook24 as opposed to 17.5% (7/40) for semi-weighed food diary and 1 participant (2.5%) opted for "other."

Discussion

Principal Findings

As far as the researchers are aware, Foodbook24 is the first Web-based dietary assessment tool developed to estimate food and nutrient intakes specifically for Irish adults. The results of this investigation into the validity of Foodbook24 suggest that the tool provides nutrient and food group intake estimates comparable with that of a semi-weighed food diary. The use of an objective measure of validity; biological markers of nutrient intake in blood and urine samples further confirm this agreement between methods.

With regards to food group intake, the results of the paired *t* test and cross-classification analysis indicate that there is good agreement between the two methods in the reporting of the majority of food group intakes. Interestingly Foodbook24 reported higher intakes of food groups that are perceived unhealthy such as "alcoholic beverages" and "sugars and confectionary" compared with the food diary. This may be due to variation in diet as was the case for fiber intakes. However, the lack of face-to-face interaction between participant and researcher that is encountered on the completion of a food record (food record review with researcher) may encourage participants of Foodbook24 to report their intake with less inhibition or in a less inhibited manner [33,34]. This may highlight an advantage of Web-based dietary assessment in terms of attenuating the influence of social desirability when self-reporting dietary intake. A study conducted by Probst and Tapsell [35] found that

XSL•FO
RenderX

patients using a computer to self-report intake were more willing to report all foods eaten to the computer than to a dietitian. Intakes of "milk and yogurt" were lower with Foodbook24 compared with the food diary, although these differences were not statistically significant. Although milk was a linked food (to prompt the participant to record milk with items such as cereal and hot beverages), it may be the case that milk consumed as a beverage was frequently forgotten when recording dietary intake using Foodbook24.

Overall for nutrient intakes, correlations were acceptable to very good; however, there were few significant differences between nutrient intakes reported using Foodbook24 and the semi-weighed food diary. The correlation ranges observed were also comparable with other studies investigating the validity of Web-based 24-hour recall methods [36-39]. There were, however, nutrients that were not correlated such as the intakes of monounsaturated fat and intakes that were significantly different from the semi-weighed food diary, for example, protein and fiber. In an evaluation of the shortened food-list (n=751) for integration into Foodbook24 [20], it was observed that there was less agreement for mean daily intake of monounsaturated and saturated fat due to the changes in food composition data that resulted from merging similar food and drink items that had similar composition into single food or drink descriptors or codes. Expanding the number of food items within this category may improve the agreement between methods for these nutrients.

Foodbook24 reported lower intakes of protein compared with the semi-weighed food diary, potentially due to different portion size estimations (portion size photographs using Foodbook24 compared with free weight entry using semi-weighed food diary). The within-person variance (ie, day-to-day variation in diet) during the two different data collection time points also accounted for the differences of intakes reported by both methods for fiber. Food items with high fiber content, for example, baked beans in tomato sauce, fruit smoothies, and breakfast cereals were more frequently reported by participants recording dietary intake using Foodbook24 than with the food diary. The challenges of the variation of diet during dietary assessment validation whereby two separate methods (the test and references measure) assess dietary intakes over two different time points have been noted by others [40,41]. However, despite the differences in fiber intake recorded, none of the cases fell outside of the limits of agreement in the Bland and Altman plot for fiber intake (Figure 8) suggesting that there may be an acceptable level of agreement between the two methods.

Biological markers of nutrient intake can serve as an objective validation of dietary assessment methods as they reflect nutritional status, metabolism, and recent dietary intake, but the error associated with biological markers is independent of dietary intake assessment error [42]. Urinary urea excretion was used as an independent marker of protein intake in this study as it can be assumed that urinary urea is excreted in constant proportion to urinary nitrogen for individuals in energy balance and consuming a westernized diet [26]. Overall, a slightly stronger correlation was observed for intakes derived from the food diary compared with Foodbook24, but this was to be expected considering more accurate portion size assessment observed with semi-weighed food records [43]. Dietary intakes recorded by both methods correlated significantly with recovery biomarkers (urinary urea, potassium, and sodium) with the exception of urinary potassium, which did not significantly correlate with potassium intakes reported from the food diary but did with dietary intakes reported from Foodbook24. This was an unexpected, but promising finding considering potassium is present in a large variety of food groups and is considered a reliable recovery biomarker in dietary studies [44]. Concentration biomarkers can be used to assess which assessment method yielded the most reliable estimates of intakes [45]. The food diary provided more reliable intakes of fish intake and total carotenoids, but estimates were similar for fruit and vegetable intake and ascorbic acid suggesting that both methods are valid and comparable in the reporting of these dietary components. The correlation coefficients between nutrient intakes and biomarkers of nutrient intakes reported in this study are comparable with those reported in other validation studies [46-48], although the correlations between protein intake and urinary urea observed in this study were stronger than those reported in the pooled results from 5 validation studies of dietary self-report instruments [49]. The pooled results study reported an average correlation coefficient for reported protein intakes versus true protein intakes of r=.29 when assessed using an FFQ and r=.48 when assessed using the average of three 24-hour recalls. However, urinary nitrogen was used as a biomarker of protein intake for these studies so that a direct comparison cannot be made. Overall, these results indicate that self-administered 24HDR via Foodbook24 provide estimates of certain nutrients and fruit and vegetable intakes similar to that of a 4-day semi-weighed food diary.

The majority of participants who used Foodbook24 were enthusiastic in their evaluation of the Web-based tool, and a large proportion of respondents claimed that they would be willing to use Foodbook24 for a week. Freese et al [50] reported similar positive responses where 95% of 370 adult participants would be willing to repeat the Web-based 24HDR after completing 3 recalls. In contrast, Maes et al [31] reported that adolescent participants involved in the HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) "Food-O-Meter" project were not eager to use their computer-based FFQ more than once. Most importantly, Foodbook24 was the preferred method of dietary intake assessment for the majority of participants. Vereecken et al [51] reported a similar preference for an online method with 73% of parents in the Children's and Adolescents' Nutrition Assessment and Advice on the Web (CANAA-W) study stating that they preferred a 3-day computerized food record over the paper and pencil 3-day food record (12%), whereas 10% selected a computerized FFQ and 6% selected a paper and pencil FFQ. Similarly, Monnerie et al [52] reported a 77% (77/100) preference for online assessment versus a traditional diary over 7 days. Thompson et al [53] also observed a clear participant preference for ASA24 when compared with the traditional interviewer administered method across a range of age groups (20-70 years) and education levels. This may highlight a willingness among Irish adults to record their dietary intake with the aid of technology and as such offers hope that Web-based methods can act as a viable alternative or

XSL•FO
RenderX

accompaniment to nutritional surveillance in Ireland. Future research is required to ascertain the actual potential for Web-based innovations to work in tandem with current methods in nutritional surveillance.

Strengths and Limitations

Although this study has many strengths including the use of biomarkers of intake in the tool validation and the inclusion of nationally representative food intake data in the tool design, it is also important to consider the study limitations. The small sample size recruited was a limitation of this study and the exclusion of participants that took nutritional supplements further reduced numbers for certain aspects of analysis. Unfortunately, high dropout rates as observed with this study are commonly reported in studies that require participants to provide biological samples on more than one occasion, particularly 24-hour urine collections. With regards to the analysis of urine samples, the use of urinary nitrogen and para-aminobenzoic acid (*PABA*) would have been preferable as an objective measure of protein intake and as a check for completeness of collection, respectively; however, these measurements were not possible for this study. Finally, the majority of participants recruited as part of the validation study and in other studies evaluating Foodbook24 were young, healthy, and motivated individuals, and therefore may not represent the opinions of the general adult population with respects to their ability to use Foodbook24 and their preference of dietary assessment methods. To further evaluate Foodbook24, a proof-of-principle (PoP) study that involves Foodbook24 being made freely available to the general Irish adult population is currently underway. The PoP study will provide insight into the acceptability of Foodbook24 with a more representative sample of the general Irish adult population.

Conclusions

In this paper, we investigated the relative validity of a Web-based 24HDR tool, Foodbook24. Although this study only investigates the validity of the 24HDR component of the Foodbook24 tool, the tool itself incorporates the use of blended assessment methods that has the potential to yield more accurate data on habitual intake [5]. This study describes the robust validation of Foodbook24 against a semi-weighed food diary and biological markers of nutrient and food group intake. The results from this study demonstrate that Foodbook24 performs well when compared with a semi-weighed food diary and provides comparable estimates of food and nutrient intakes. A major advantage of Foodbook24 and of similar Web-based dietary assessment tools is the reduced cost associated with the collection of dietary intake data compared with traditional methods. More importantly, Web-based methodologies facilitate the collection of data in a neutral environment, in the absence of a researcher with less burden for the participant which may encourage participants to report intake more honestly. Participant acceptability data gathered so far suggests Foodbook24 was well received by the majority of participants in this study sample which indicates the potential of Foodbook24 for use in nutrition related research or as a means of intermittent data collection between national nutrition surveys in Ireland.

Acknowledgments

This research was completed as part of a Food Institutional Research Measure (FIRM) project, DietIreland, funded by the Department of Agriculture, Food and Marine (13/F/424). The researchers wish to acknowledge the participants who took part in the study and also Creme Global for developing the software for Foodbook24.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Daily energy and nutrient intakes recorded by participants using the Foodbook24 tool and a 4-day semi-weighed food diary.

[PDF File (Adobe PDF File), 46KB - jmir_v19i5e158_app1.pdf]

References

1. Thompson F, Subar A. Nutrition in the Prevention and Treatment of Disease. London: Elsevier; 2013.
2. MacIntyre U. Measuring Food Intake. In: Introduction to Human Nutrition - The Nutrition Society Textbook. Oxford: Wiley-Blackwell; 2002.
3. Walton J. Dietary assessment methodology for nutritional assessment. Top Clin Nutr 2015;30(1):33-46. [doi: 10.1097/TIN.0000000000000018]
4. Subar AF, Crafts J, Zimmerman TP, Wilson M, Mittl B, Islam NG, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. J Am Diet Assoc 2010 Jan;110(1):55-64 [FREE Full text] [doi: 10.1016/j.jada.2009.10.007] [Medline: 20102828]
5. Illner A, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. Int J Epidemiol 2012 Aug;41(4):1187-1203 [FREE Full text] [doi: 10.1093/ije/dys105] [Medline: 22933652]

XSL•FO
RenderX

6. Fallaize R, Forster H, Macready AL, Walsh MC, Mathers JC, Brennan L, et al. Online dietary intake estimation: reproducibility and validity of the Food4Me food frequency questionnaire against a 4-day weighed food record. J Med Internet Res 2014 Aug;16(8):e190 [FREE Full text] [doi: 10.2196/jmir.3355] [Medline: 25113936]
7. Slimani N, Casagrande C, Nicolas G, Freisling H, Huybrechts I, Ocké MC, et al. The standardized computerized 24-h dietary recall method EPIC-Soft adapted for pan-European dietary monitoring. Eur J Clin Nutr 2011 Jul;65 Suppl 1:S5-15. [doi: 10.1038/ejcn.2011.83] [Medline: 21731006]
8. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. Eur J Clin Nutr 2009 Feb;63:S50-S57 [FREE Full text] [doi: 10.1038/ejcn.2008.65]
9. Touvier M, Kesse-Guyot E, Méjean C, Pollet C, Malon A, Castetbon K, et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. Br J Nutr 2011 Apr;105(7):1055-1064. [doi: 10.1017/S0007114510004617] [Medline: 21080983]
10. Margetts B, Nelson M. Design concepts in nutritional epidemiology. Oxford: Oxford University Press; 1997.
11. McCabe-Sellers B. Advancing the art and science of dietary assessment through technology. J Am Diet Assoc 2010 Jan;110(1):52-54. [doi: 10.1016/j.jada.2009.10.014] [Medline: 20102827]
12. Diep C, Hingle M, Chen T, Dadabhoy H, Beltran A, Baranowski J. The automated self-administered 24-hour dietary recall for children, 2012 version, for youth aged 9 to 11 years: a validation study. J Acad Nutr Diet 2012;115(10):1591-1598. [doi: 10.1016/j.jand.2015.02.021]
13. Kirkpatrick SI, Subar AF, Douglass D, Zimmerman TP, Thompson FE, Kahle LL, et al. Performance of the automated aelf-administered 24-hour recall relative to a measure of true intakes and to an interviewer-administered 24-h recall. Am J Clin Nutr 2014 Jul;100(1):233-240 [FREE Full text] [doi: 10.3945/ajcn.114.083238] [Medline: 24787491]
14. Burrows TL, Khambalia AZ, Perry R, Carty D, Hendrie GA, Allman-Farinelli MA, et al. Great 'app-eal' but not there yet: a review of iPhone nutrition applications relevant to child weight management. Nutr Diet 2015 Apr 16;72(4):363-367. [doi: 10.1111/1747-0080.12184]
15. European Food Safety Authority. General principles for the collection of national food consumption data in the view of a pan-European dietary survey. EFSA Journal 2009 Dec 18;7(12):1435. [doi: 10.2903/j.efsa.2009.1435]
16. Moshfegh A, Raper N, Ingwersen L. An improved approach to 24-hour dietary recall methodology. Ann Nutr Metab (suppl) 2001;45:156.
17. Irish Universities Nutrition Alliance. 2010. National adult nutrition survey: summary report URL: http://www.iuna.net/ [accessed 2016-07-25] [WebCite Cache ID 6jGbgac8E]
18. McCance RA, Widdowson EM. McCance and Widdowsons The Composition of Foods. Cambridge: Royal Society of Chemistry; 2002.
19. Black L, Ireland J, Møller A, Roe M, Walton J, Flynn A, et al. Development of an on-line Irish food composition database for nutrients. J Food Comp Anal 2011 Nov;24(7):1017-1023. [doi: 10.1016/j.jfca.2011.01.015]
20. Evans K, Hennessy A, Timon C, Walton J, Gibney E, Flynn A. Development of a short food list for use in a web-based dietary assessment tool. In: Ann Nutr Metab. 2015 Presented at: 12th FENS European Nutrition Conference; October 2015; Berlin p. 214.
21. Foster E. Final Report for Food Standards Agency. Comparison study: INTAKE24 vs Interviewer led recall URL: http://webarchive.nationalarchives.gov.uk/+/http://www.food.gov.uk/scotland/news-updates/news/2014/13135/intake24 [accessed 2016-11-04] [WebCite Cache ID 6llphUnFa]
22. Mills A, Patel S. Food Standards Agency. In: Food Portion Sizes. London: TSO; 1993.
23. Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, et al. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. Int J Epidemiol 1997;26 Suppl 1:S137-S151 [FREE Full text] [Medline: 9126542]
24. Steptoe A, Pollard TM, Wardle J. Development of a measure of the motives underlying the selection of food: the food choice questionnaire. Appetite 1995 Dec;25(3):267-284. [doi: 10.1006/appe.1995.0061] [Medline: 8746966]
25. Nicholson S, Cox L, Mistry P. Food. NDNS appendix T: the 24-hour urine sample: collecting and processing the urine and assessment of completeness of collection URL: https://www.food.gov.uk/sites/default/files/ndns-appendix-t-ni.pdf [accessed 2016-11-17] [WebCite Cache ID 6m5ofdQif]
26. Bingham SA, Williams R, Cole TJ, Price CP, Cummings JH. Reference values for analytes of 24-h urine collections known to be complete. Ann Clin Biochem 1988 Nov;25:610-619. [Medline: 3267076]
27. Joossens J, Geboers J. Surveillance of dietary habits of the population with regard to cardiovascular diseases: validity and reproducibility of methods. In: Euro-Nut Report 2: A concerted action project on nutrition in the European Community. Wageningen, The Netherlands: Department of Human Nutrition, Agricultural University; 1984.
28. Mercado CI, Cogswell ME, Valderrama AL, Wang C, Loria CM, Moshfegh AJ, et al. Difference between 24-h diet recall and urine excretion for assessing population sodium and potassium intake in adults aged 18-39 y. Am J Clin Nutr 2015 Feb;101(2):376-386 [FREE Full text] [doi: 10.3945/ajcn.113.081604] [Medline: 25646336]
29. Bland M. An introduction to medical statistics. Oxford: Oxford University Press; 2000.
30. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]

XSL•FO
RenderX

31. Maes L, Cook TL, Ottovaere C, Matthijs C, Moreno LA, Kersting M, et al. Pilot evaluation of the HELENA (healthy lifestyle in europe by nutrition in adolescence) food-o-meter, a computer-tailored nutrition advice for adolescents: a study in six European cities. Public Health Nutr 2011 Jul;14(7):1292-1302. [doi: 10.1017/S1368980010003563] [Medline: 21281542]
32. Vereecken C, Covents M, Maes L, Moyson T. Formative evaluation of the dietary assessment component of children's and adolescents' nutrition assessment and advice on the web (CANAA-W). J Hum Nutr Diet 2014 Jan;27 Suppl 1:54-65. [doi: 10.1111/j.1365-277X.2012.01290.x] [Medline: 23980931]
33. van Gelder Marleen MH, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 1;172(11):1292-1298 [FREE Full text] [doi: 10.1093/aje/kwq291] [Medline: 20880962]
34. Mathieu E, Barratt A, Carter SM, Jamtvedt G. Internet trials: participant experiences and perspectives. BMC Med Res Methodol 2012 Oct 23;12:162 [FREE Full text] [doi: 10.1186/1471-2288-12-162] [Medline: 23092116]
35. Probst Y, Tapsell L. Over- and underreporting of energy intake by patients with metabolic syndrome using an automated dietary assessment website. Nutr Diet 2007 Dec;64(4):280-284. [doi: 10.1111/j.1747-0080.2007.00220.x]
36. De KW, Huybrechts I, De VV, Vandevijvere S, Slimani N, Van OH, et al. Repeated 24-hour recalls versus dietary records for estimating nutrient intakes in a national food consumption survey. Food Nutr Res 2011 Nov;55:7307 [FREE Full text] [doi: 10.3402/fnr.v55i0.7307] [Medline: 22084625]
37. Comrie F, Masson LF, McNeill G. A novel online Food Recall Checklist for use in an undergraduate student population: a comparison with diet diaries. Nutr J 2009 Feb 19;8:13 [FREE Full text] [doi: 10.1186/1475-2891-8-13] [Medline: 19228392]
38. Liu B, Young H, Crowe FL, Benson VS, Spencer EA, Key TJ, et al. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. Public Health Nutr 2011 Nov;14(11):1998-2005. [doi: 10.1017/S1368980011000942] [Medline: 21729481]
39. Zoellner J, Anderson J, Gould SM. Comparative validation of a bilingual interactive multimedia dietary assessment tool. J Am Diet Assoc 2005 Aug;105(8):1206-1214. [doi: 10.1016/j.jada.2005.05.011] [Medline: 16182635]
40. Labonté M, Cyr A, Baril-Gravel L, Royer M, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. Eur J Clin Nutr 2012 Feb;66(2):166-173. [doi: 10.1038/ejcn.2011.163] [Medline: 21934698]
41. Chan SG, Ho SC, Kreiger N, Darlington G, Adlaf EM, So KF, et al. Validation of a food frequency questionnaire for assessing dietary soy isoflavone intake among midlife Chinese women in Hong Kong. J Nutr 2008 Mar;138(3):567-573 [FREE Full text] [Medline: 18287368]
42. McKeown NM, Day NE, Welch AA, Runswick SA, Luben RN, Mulligan AA, et al. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. Am J Clin Nutr 2001 Aug;74(2):188-196 [FREE Full text] [Medline: 11470719]
43. Bingham SA, Gill C, Welch A, Day K, Cassidy A, Khaw KT, et al. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. Br J Nutr 1994 Oct;72(4):619-643. [Medline: 7986792]
44. Tasevska N, Runswick SA, Bingham SA. Urinary potassium is as reliable as urinary nitrogen for use as a recovery biomarker in dietary studies of free living individuals. J Nutr 2006 May;136(5):1334-1340 [FREE Full text] [Medline: 16614426]
45. Bingham S, Luben R, Welch A, Low YL, Khaw KT, Wareham N, et al. Associations between dietary methods and biomarkers, and between fruits and vegetables and risk of ischaemic heart disease, in the EPIC Norfolk Cohort Study. Int J Epidemiol 2008 Oct;37(5):978-987 [FREE Full text] [doi: 10.1093/ije/dyn111] [Medline: 18579574]
46. Lassale C, Castetbon K, Laporte F, Camilleri GM, Deschamps V, Vernay M, et al. Validation of a Web-based, self-administered, non-consecutive-day dietary record tool against urinary biomarkers. Br J Nutr 2015 Mar 28;113(6):953-962. [doi: 10.1017/S0007114515000057] [Medline: 25772032]
47. Timon CM, Astell AJ, Hwang F, Adlam TD, Smith T, Maclean L, et al. The validation of a computer-based food record for older adults: the novel assessment of nutrition and ageing (NANA) method. Br J Nutr 2015 Feb 28;113(4):654-664. [doi: 10.1017/S0007114514003808] [Medline: 25630436]
48. Patterson AC, Hogg RC, Kishi DM, Stark KD. Biomarker and dietary validation of a Canadian food frequency questionnaire to measure eicosapentaenoic and docosahexaenoic acid intakes from whole food, functional food, and nutraceutical sources. J Acad Nutr Diet 2012 Jul;112(7):1005-1014. [doi: 10.1016/j.jand.2012.03.030] [Medline: 22583924]
49. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. Am J Epidemiol 2014 Jul 15;180(2):172-188. [doi: 10.1093/aje/kwu116] [Medline: 24918187]
50. Freese J, Feller S, Harttig U, Kleiser C, Linseisen J, Fischer B, et al. Development and evaluation of a short 24-h food list as part of a blended dietary assessment strategy in large-scale cohort studies. Eur J Clin Nutr 2014 Mar;68(3):324-329. [doi: 10.1038/ejcn.2013.274] [Medline: 24398637]
51. Vereecken CA, Covents M, Haynie D, Maes L. Feasibility of the young children's nutrition assessment on the web. J Am Diet Assoc 2009 Nov;109(11):1896-1902. [doi: 10.1016/j.jada.2009.08.013] [Medline: 19857631]

XSL•FO
RenderX

52. Monnerie B, Tavoularis LG, Guelinckx I, Hebel P, Boisvieux T, Cousin A, et al. A cross-over study comparing an online versus a paper 7-day food record: focus on total water intake data and participant's perception of the records. Eur J Nutr 2015 Jun;54 Suppl 2:27-34 [FREE Full text] [doi: 10.1007/s00394-015-0945-7] [Medline: 26066355]
53. Thompson FE, Dixit-Joshi S, Potischman N, Dodd KW, Kirkpatrick SI, Kushi LH, et al. Comparison of interviewer-administered and automated self-administered 24-hour dietary recalls in 3 diverse integrated health systems. Am J Epidemiol 2015 Jun 15;181(12):970-978 [FREE Full text] [doi: 10.1093/aje/kwu467] [Medline: 25964261]

Abbreviations

ANOVA: analysis of variance
FFQ: food frequency questionnaire
24HDR: 24-hour dietary recall
RCF: relative centrifugal force
UCE: urinary creatinine excretion

Edited by G Eysenbach; submitted 26.07.16; peer-reviewed by A Subar, F Thompson, T Baranowski, J Sheats; comments to author 27.09.16; revised version received 18.11.16; accepted 11.02.17; published 11.05.17.

Please cite as:
Timon CM, Blain RJ, McNulty B, Kehoe L, Evans K, Walton J, Flynn A, Gibney ER
The Development, Validation, and User Evaluation of Foodbook24: A Web-Based Dietary Assessment Tool Developed for the Irish Adult Population
J Med Internet Res 2017;19(5):e158
URL: http://www.jmir.org/2017/5/e158/
doi:10.2196/jmir.6407
PMID:28495662

©Claire M Timon, Richard J Blain, Breige McNulty, Laura Kehoe, Katie Evans, Janette Walton, Albert Flynn, Eileen R Gibney. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 11.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Design and Evaluation of a Computer-Based 24-Hour Physical Activity Recall (cpar24) Instrument

Simone Kohler[1*], MPH; Gundula Behrens[1*], PhD; Matthias Olden[1], Dr oec; Sebastian E Baumeister[1,2], Dr rer med; Alexander Horsch[3], Dr rer nat; Beate Fischer[1], Dr oec troph; Michael F Leitzmann[1], MD

[1]Department of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany

[2]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

[3]Department of Computer Science, UiT – The Arctic university of Norway, Breivika, Tromsø, Norway

[*]these authors contributed equally

Corresponding Author:
Gundula Behrens, PhD
Department of Epidemiology and Preventive Medicine
University of Regensburg
Franz-Josef-Strauss-Allee 11
Regensburg, 93053
Germany
Phone: 49 941 944 ext 5217
Fax: 49 941 944 5202
Email: Gundula.Behrens@klinik.uni-regensburg.de

Abstract

Background: Widespread access to the Internet and an increasing number of Internet users offers the opportunity of using Web-based recalls to collect detailed physical activity data in epidemiologic studies.

Objective: The aim of this investigation was to evaluate the validity and reliability of a computer-based 24-hour physical activity recall (cpar24) instrument with respect to the recalled 24-h period.

Methods: A random sample of 67 German residents aged 22 to 70 years was instructed to wear an ActiGraph GT3X+ accelerometer for 3 days. Accelerometer counts per min were used to classify activities as sedentary (<100 counts per min), light (100-1951 counts per min), and moderate to vigorous (≥1952 counts per min). On day 3, participants were also requested to specify the type, intensity, timing, and context of all activities performed during day 2 using the cpar24. Using metabolic equivalent of task (MET), the cpar24 activities were classified as sedentary (<1.5 MET), light (1.5-2.9 MET), and moderate to vigorous (≥3.0 MET). The cpar24 was administered twice at a 3-h interval. The Spearman correlation coefficient (r) was used as primary measure of concurrent validity and test-retest reliability.

Results: As compared with accelerometry, the cpar24 underestimated light activity by −123 min (median difference, *P* difference <.001) and overestimated moderate to vigorous activity by 89 min (*P* difference <.001). By comparison, time spent sedentary assessed by the 2 methods was similar (median difference=+7 min, *P* difference=.39). There was modest agreement between the cpar24 and accelerometry regarding sedentary (*r*=.54), light (*r*=.46), and moderate to vigorous (*r*=.50) activities. Reliability analyses revealed modest to high intraclass correlation coefficients for sedentary (*r*=.75), light (*r*=.65), and moderate to vigorous (*r*=.92) activities and no statistically significant differences between replicate cpar24 measurements (median difference for sedentary activities=+10 min, for light activities=−5 min, for moderate to vigorous activities=0 min, all *P* difference ≥.60).

Conclusion: These data show that the cpar24 is a valid and reproducible Web-based measure of physical activity in adults.

(J Med Internet Res 2017;19(5):e186) doi:10.2196/jmir.7620

KEYWORDS

web-based method; validity; reliability; usability; lifestyle behavior; physical activity; sedentary behavior

XSL•FO
RenderX

Introduction

Physical activity is associated with decreased risk of numerous chronic diseases, including type 2 diabetes [1], cardiovascular disease [2], and certain types of cancer [3]. However, information regarding which frequencies, intensities, and durations of specific activities or combinations of activities are relevant to reducing disease risk is sparse. Thus, comprehensive assessments of physical activity are required to better characterize the relation of physical activity to risk of chronic disease.

A variety of methods to assess physical activity in epidemiologic studies exist, and each measurement technique has particular advantages and limitations [4-6]. Increased availability of the Internet with rising numbers of Internet users and recent progress in computer technology provide the opportunity to use Internet-based instruments to assess physical activity in large populations with enhanced accuracy, minimal logistic burden, and reduced time and costs. Although a wide range of physical activity questionnaires is established, few instruments are Web-based or provide information about the type, frequency, and duration of physical activity across the entire day. Previous Web-based [7], computer-based [8-11], and cell-phone-based [12] 24-h physical activity recalls were developed in English [8-12] and Japanese [7]. Those instruments showed high validity correlation coefficients of .87 to .91 for total energy expenditure estimates when compared against doubly-labeled water [7] and multi-sensors [8], low to moderate accelerometer-based validation correlation coefficients of .36 to .72 for total energy expenditure [9,10], and of .26 to .59 for total time spent in sedentary and moderate to vigorous intensity activities [10-12].

The purpose of this study was to develop a computer-based 24-h physical activity recall (cpar24) instrument and to evaluate its concurrent validity, test-retest reliability, and usability with respect to the recalled 24-h period based on a population-based sample from a pilot study of the German National Cohort [13]. The cpar24 represents part of the physical activity assessment in the German National Cohort [13], a population-based prospective study of 200,000 men and women aged 20-69 years, which was initiated in 2014.

Methods

Study Protocol and Participants

The study was conducted from July to August, 2011 as part of a pilot study of the German National Cohort and included a random sample of 67 healthy participants (34 women and 33 men) aged 22 to 70 years from Regensburg, Germany. Exclusion criteria were lack of language skills, no Internet access, no computer experience, and unwillingness to wear an accelerometer. Sixty-seven participants took part in the study by completing the cpar24 twice during their visit at the study center. Of those, 49 subjects (73%, 49/67) wore the GT3X+ accelerometer for 3 days and subsequently completed the cpar24 a third time at home. Fifty-three subjects (79%, 53/67) responded to the usability survey. The study protocol was approved by the ethics committee of Regensburg University, and all participants provided written informed consent.

Description of the Computer-Based 24-hour Physical Activity Recall (cpar24) Instrument

The cpar24 is a self-administered, computer-based, Web-based-accessible 24-h physical activity recall instrument designed to assess detailed information about the specific types, durations, and intensities of active and sedentary behaviors on the previous day (midnight to midnight). It was developed to be easy to administer, with minimal user training, and a completion time of 30 min or less for the majority of participants. Specifically, the cpar24 guides a participant to select, in chronological order, specific activities performed throughout the previous 24-h period from a list of 262 activities that are divided into the 13 following broad categories: (1) sleeping and reclining; (2) personal care; (3) food preparation and eating; (4) walking, transportation, and traveling; (5) household chores; (6) occupational activity; (7) shopping, errands, and appointments; (8) leisure and hobbies; (9) sports; (10) family life and social activities; (11) outdoor activities; (12) lawn and garden; and (13) miscellaneous activities. In addition, the respondent may refer to an alphabetical list of activities using a search function or select a specific activity using a search box. The response categories and follow-up probes were designed to allow the respondent to select broad activity classifications (eg, sports) followed by questions regarding more specific aspects of the activity within the category reported (eg, soccer). The participant can view his or her responses through an interactive calendar that allows response editing by dragging or dropping response items.

Once an activity is selected, the respondent is asked to indicate the start and end times of the activity in durations of 5 min or more. A minimal bout length of 5 min was chosen to facilitate reporting of activities of short duration. The respondent is able to enter 2 activities during the same 5 min time period, in line with the recommendation that physical activity diaries should include main activities as well as activities performed in parallel [14]. For activities that require a ranking of intensity (eg, cycling and Nordic walking), the respondent is asked to indicate the level of effort using categories of light, medium, and hard intensities. For activities that can be performed either standing or sitting or a combination of standing and sitting, the respondent is requested to specify the ratio of standing to sitting time using a scale from 0% to 100%. Each activity reported is assigned a metabolic equivalent of task (MET) value based on the most recent compendium of physical activities published by Ainsworth et al [15].

Respondents are asked to fully complete the recall before ending the session. To ensure complete data entry, a review of all items reported is provided, and the respondent is informed about missing or incomplete activity entries (ie, time gaps) with the option of adding new activity items in order to arrive at the desired total amount of 1440 min (=24 h) of logged activities per day. At the end of the recall, a brief survey on respondent burden and usability is administered. Specifically, the respondent is asked to report the time needed to complete the recall and to respond to the following 6 questions, with response options ranging from 1 (excellent) to 6 (unsatisfactory): (1) “How well were you able to recall activities performed yesterday?” (2) “How helpful was the user’s manual?” (3) “How helpful were

XSL•FO
RenderX

the broad activity categories (eg, household chores, outdoor activities) to find a specific activity?" (4) "How would you rate the overall ease of using the cpar24?" (5) "How well were you able to navigate the cpar24 interface?" and (6) "Do you like the design of the cpar24?"

Criterion Measure of Physical Activity

Accelerometry is an established simple, noninvasive, and cost-efficient method for assessing physical activity in a detailed and objective manner [16,17] and was therefore selected as criterion measure. We used the GT3X+ accelerometer (ActiGraph, LLC, Pensacola, FL, USA). This device measures motion in 3 axes with a sampling rate of 100 Hz and the output is expressed as counts per epoch. Participants wore the GT3X+ accelerometer over a 3-day period and subsequently completed the cpar24 at home on the third day, recalling their previous day's activity, that is, their activities on the second day of accelerometer measurement. Accelerometers were fitted by skilled personnel at the study center and worn on a belt at the natural waistline on the right hip in line with the right axilla. Participants were instructed to wear the monitor at all times (day and night) except during swimming, sauna, and martial arts and to report the number and reasons of wear interruptions in a specific document. Accelerometer data were downloaded using the ActiLife v5.6.4 Firmware v2.1.0 software (ActiGraph, LLC, Pensacola, FL, USA) and were subsequently checked to ensure that the device had been functioning properly. Accelerometer data with less than 12 h (50%) of wear time were excluded from analysis. Since the second day of accelerometer monitoring covered the cpar24 recall period, only data referring to that 24-h time period were included.

Statistical Methods

To examine the validity of the cpar24 in relation to accelerometry, we compared cpar24 data with accelerometer data among participants with complete data from both assessment methods. For cpar24 data, activity intensities were classified as sedentary (<1.5 MET), light (1.5-2.9 MET), and moderate to vigorous (≥3.0 MET). For accelerometer data, the activity intensity classification was based on the Freedson formula [18] in combination with the 100 counts per minute cut-off for sedentary activities as suggested by Matthews et al [19], classifying the intensity of activities as sedentary (less than 100 counts per min), light (100 to 1951 counts per min), and moderate to vigorous (1952 or more counts per min). In subanalyses, we assessed the validity of the cpar24 stratified by age (<60 years, ≥60 years), gender (men, women), and body mass index (BMI: <25 kg/m^2, ≥25 kg/m^2).

We assessed the reliability of the cpar24 instrument based on two cpar24 recalls from the same 24-h period, the completions of which were separated by approximately 3 h. We used the first cpar24 recall as criterion measure to assess the reliability of the instrument in the entire sample and in subgroups defined by age, gender, and body mass index (BMI). To assess the usability of the cpar24, we evaluated the 6-item usability questionnaire stratified by age, gender, and BMI.

All statistical analyses were conducted using nonparametric methods, including Spearman correlations, median, and rank comparisons. In particular, we tested if the median total time spent in sedentary, light or moderate to vigorous activities varied according to the assessment method (accelerometer vs cpar24) using the Wilcoxon signed rank test. In addition, we computed the difference in the total time spent in a specific physical activity intensity level between the two assessments (accelerometry vs cpar24) for each participant, and we tested if that difference varied across strata defined by the participants' age, gender, and BMI using the Wilcoxon rank sum test. We also generated Bland-Altman plots [20] to examine the agreement between the activity variables. We conducted 2-sided statistical tests at a significance level of 5%. All analyses were performed using the statistical software R, version 3.2.3 [21].

Results

Participants' Characteristics

The study sample showed a nearly equal gender distribution (34 women and 33 men, Table 1). The mean age of the participants was 52 years (range=22-70 years), and their mean BMI was 26.1 kg/m^2 (range=18.1-41.2 kg/m^2).

Validity of the Computer-Based 24-Hour Physical Activity Recall (cpar24) Instrument Estimates

The cpar24 and accelerometer estimates of the total activity duration were modestly positively correlated, showing Spearman correlations of .54 for sedentary activity, .46 for light activity, and .50 for moderate to vigorous activity (Table 2). However, the cpar24 underestimated the time spent in light activities by −123 min (corresponding to −28%, *P* difference <.001), and it overestimated moderate to vigorous activity by 89 min (corresponding to +353%, *P* difference <.001) when compared with accelerometer measurements. In contrast, the 2 assessment methods agreed with respect to time spent sedentary (*P* difference=.39). The pattern of agreement of total time spent in sedentary, light, and moderate to vigorous activities was not affected by age, gender, and BMI of participants (all *P* difference ≥.23).

XSL•FO
RenderX

Table 1. Characteristics of the participants included in the reliability, validity, and usability studies of the computer-assisted 24-hour physical activity recall (cpar24) instrument.

Variable	Reliability study	Validity study	Usability study
Participants, n (%)			
Total	67 (100)	49 (100)	53 (100)
Men	33 (49)	24 (49)	26 (49)
Women	34 (51)	25 (51)	27 (51)
Age, years			
Minimum	22	22	22
Maximum	70	69	69
Mean	52	50	53
Standard deviation	13	13	13
Body mass index, kg/m^2			
Minimum	18.1	18.2	18.2
Maximum	41.2	41.2	41.2
Mean	26.1	26.1	26
Standard deviation	4.4	4.7	4.1

Bland-Altman plots illustrated the previously described bias regarding the assessments of light activity and moderate to vigorous activity (Figure 1). The difference between the estimates increased with the magnitude of the estimates. This also held true for sedentary behavior (Figure 1) despite the previously observed comparability of the corresponding median values (Table 2). According to the Bland-Altman analyses, the mean bias and limits of agreement (LoA) were −31 min (LoA=−380 to +319 min) for sedentary time, −98 min (LoA=−399 to +204 min) for light intensity physical activity, and +128 min (LoA=−151 to +407 min) for moderate to vigorous intensity physical activity.

Reliability of the cpar24

Reliability analyses (Table 3) yielded moderate to strong Spearman correlations for time spent sedentary (r=.75), light (r=.65), and moderate to vigorous activities (r=.92). In the reliability analyses, no systematic bias was observed between the two cpar24 assessments of the total durations of sedentary, light, and moderate to vigorous activities (all P difference ≥.60). In general, age, gender, and BMI of participants did not influence the results (all P difference ≥.09). However, for moderate to vigorous physical activity, the median difference between the two assessments varied statistically significantly across age groups even though the absolute difference was not substantial. Specifically, the median difference between the two 24-h physical activity recalls with respect to total duration of moderate to vigorous physical activities was null among people aged less than 60 years, and it was 8 min among people aged 60 years or more; P difference=.04). Similarly, the average MET values were comparable across the two 24-h recalls, yielding median values of 1.71 and 1.69 for the first and second 24-h recall, respectively (P difference=.34 as assessed by the Wilcoxon signed rank test; Spearman correlation=.91).

XSL•FO
RenderX

Table 2. Comparison of total time spent in sedentary, light, and moderate to vigorous activity during the 24-h period as assessed by accelerometery and by computer-based 24-h physical activity recall (cpar24) instrument.

Stratum and variable	Sedentary activity[a]	Light activity[a]	Moderate to vigorous activity[a]
Total time during 24-h period			
All participants			
Median total time based on accelerometer data, in min	1004	377	30
Median total time based on cpar24[e] data, in min	980	265	145
Median difference between cpar24 and accelerometer total time[b], in min (and in %)	7 (+1)	−123 (−28)	89 (+353)
P difference[c]	.39	<.001	<.001
Spearman correlation	.54	.46	.50
Participants aged < 60 years			
Median total time based on accelerometer data, in min	978	391	30
Median total time based on cpar24 data, in min	980	265	120
Median difference between cpar24 and accelerometer total time[b], in min (and in %)	7 (+1)	−130 (−31)	85 (+353)
P difference[c]	.66	<.001	<.001
Spearman correlation	.56	.48	.46
Participants aged ≥ 60 years			
Median total time based on accelerometer data, in min	1022	361	42
Median total time based on cpar24 data, in min	968	255	150
Median difference between cpar24 and accelerometer total time[b], in min (and in %)	−36 (−3)	−102 (−21)	96 (+391)
P difference[c]	.26	.03	.003
Spearman correlation	.37	.33	.73
P value for the influence of age on the difference between cpar24 and accelerometer data[d]	.46	.55	.38
Men			
Median total time based on accelerometer data, in min	1014	361	39
Median total time based on cpar24 data, in min	985	182	148
Median difference between cpar24 and accelerometer total time[b], in min (and in %)	8 (+1)	−149 (−49)	92 (+350)
P difference[c]	.82	<.001	<.001
Spearman correlation	.65	.47	.62
Women			
Median total time based on accelerometer data, in min	978	400	30
Median total time based on cpar24 data, in min	930	305	125
Median difference between cpar24 and accelerometer total time[b], in min (and in %)	−24 (−2)	−83 (−21)	85 (+400)
P difference[c]	.20	.01	<.001
Spearman correlation	.50	.40	.40
P value for the influence of gender on the difference between cpar24 and accelerometer data[d]	.27	.26	.99
Participants with a BMI[f]<25.0 kg/m^2			

XSL•FO
RenderX

Stratum and variable	Sedentary activity[a]	Light activity[a]	Moderate to vigorous activity[a]
Median total time based on accelerometer data, in min	972	381	38
Median total time based on cpar24 data, in min	992	265	110
Median difference between cpar24 and accelerometer total time[b], in min (and in %)	17 (+2)	−118 (−28)	70 (+192)
P difference[c]	.84	.005	<.001
Spearman correlation	.56	.54	.59
Participants with a BMI≥25.0 kg/m^2			
Median total time based on accelerometer data, in min	1017	367	30
Median total time based on cpar24 data, in min	950	255	145
Median difference between cpar24 and accelerometer total time[b], in min (and in %)	−53 (−5)	−135 (−31)	122 (+600)
P difference[c]	.23	.001	<.001
Spearman correlation	.48	.24	.49
P value for the influence of the BMI on the difference between cpar24 and accelerometer data[d]	.36	.95	.23

[a]For accelerometer counts, we classified the physical activity intensity according to the Freedson formula combined with the 100 counts per min cut-off suggested by Matthews: sedentary activity (counts per min<100), light physical activity (100≤counts per min<1952), moderate to vigorous physical activity (1952≤counts per min); for self-reported physical activity (cpar24), we classified the physical activity intensity according to the corresponding metabolic equivalent of task (MET) value from the Ainsworth compendium: sedentary activity (MET≤1.5), light physical activity (1.5<MET<3.0), moderate to vigorous physical activity (3.0≤MET).

[b]Please note that the median of the difference between 2 variables does not necessarily correspond to the difference between the medians of the 2 variables.

[c]We tested if the median total time spent in sedentary, light or moderate to vigorous activities varied according to the assessment method (accelerometer vs cpar24) using the Wilcoxon signed rank test.

[d]We computed the difference in the total time spent in a specific physical activity intensity level between the two assessments (accelerometry vs cpar24) for each participant, and we tested if that difference varied across the two strata of participants using the Wilcoxon rank sum test.

[e]cpar24: computer-based 24-h physical activity recall.

[f]BMI: body mass index.

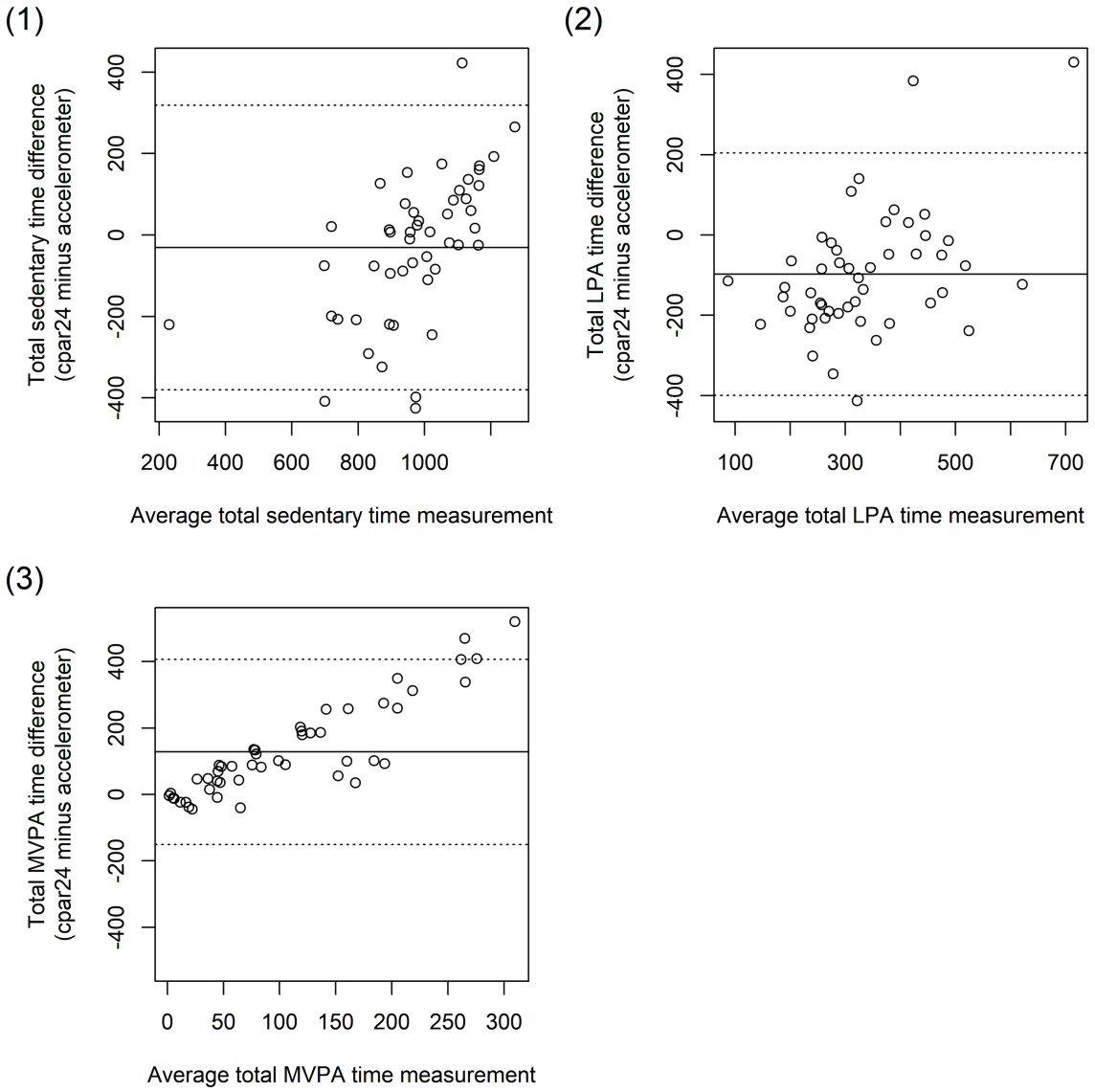
In agreement with the reliability analyses from Table 3, Bland-Altman plots did not indicate any systematic bias for total time spent in sedentary, light, and moderate to vigorous activities and for the average MET value for the entire 24-h period (Figure 2). According to the Bland-Altman analyses, the mean bias and limits of agreement were −17 min (LoA=−292 to +259 min) for sedentary time, +20 min (LoA=−256 to +296 min) for light intensity physical activity, −3 min (LoA=−109 to +102 min) for moderate to vigorous intensity physical activity, and 0.0 METs (LoA=−0.3 to +0.3 METs) for the average MET value.

Usability of the cpar24

The usability of the cpar24 varied according to age. When considering participants aged less than 60 years, 82% to 91% rated the usability of the cpar24 as “excellent” or “good” with regards to their ability to recall activities performed during the previous 24 h, to find specific activities within broad activity categories, to rate the overall ease of using the cpar24, and to navigate the cpar24 interface. By comparison, when considering participants aged 60 years or more, only 58% to 74% rated the cpar24 as “excellent” or “good” (*P* difference by age <.05 for all of the aforementioned usability ratings, Figure 3). In contrast, no statistically significant difference was observed between the ratings of participants aged less than 60 years and the rating of participants aged 60 years or more with respect to the usefulness of the user’s manual and the appeal of the cpar24 design, which received “excellent” or “good” ratings from 70% to 85% of participants aged less than 60 years and 68% to 89% of participants aged 60 years or more. In contrast, gender (all *P* difference ≥.07) and BMI (*P* difference ≥ .08) did not affect the ratings for any of the usability survey items after stratification by age. Participants completed the cpar24 within an average of 25 min (median, range=10-53 min, interquartile range=20-30 min).

XSL•FO
RenderX

Figure 1. Bland Altman plots comparing computer-based 24-hour physical activity recall (cpar24) instrument data against accelerometry data of the 49 participants of the validity study with respect to (1) the total time spent in sedentary activities, (2) the total time spent in light physical activities, and (3) the total time spent in moderate to vigorous physical activities. LPA=light physical activity; MVPA=moderate to vigorous physical activity.



XSL•FO
RenderX

Table 3. Comparison of total time spent in sedentary, light, and moderate to vigorous activity during the 24-hour period across the two 24-hour physical activity recalls (cpar24).

Stratum and variable	Sedentary activity[a]	Light activity[a]	Moderate to vigorous activity[a]
Total time during 24-h period			
All participants			
Median total time based on 1st 24-h recall, in min	1010	265	120
Median total time based on 2nd 24-h recall, in min	990	300	115
Median difference between 1st and 2nd 24-h recall[b], in min (and in %)	10 (+1)	−5 (−1)	0 (+0)
P difference[c]	.60	.89	.68
Spearman correlation	.75	.65	.92
Participants aged < 60 years			
Median total time based on 1st 24-h recall, in min	1035	250	105
Median total time based on 2nd 24-h recall, in min	1025	295	100
Median difference between 1st and 2nd 24-h recall[b], in min (and in %)	10 (+1%)	0 (+0%)	0 (+0%)
P difference[c]	.88	.55	.10
Spearman correlation	.75	.58	.93
Participants aged ≥60 years			
Median total time based on 1st 24-h recall, in min	1002	288	120
Median total time based on 2nd 24-h recall, in min	990	305	130
Median difference between 1st and 2nd 24-h recall[b], in min (and in %)	25 (+3%)	−18 (−9%)	8 (+0%)
P difference[c]	.55	.24	.22
Spearman correlation	.76	.87	.83
P value for the influence of age on the difference between cpar24 and accelerometer data[d]	.45	.21	.04
Men			
Median total time based on 1st 24-h recall, in min	1010	240	160
Median total time based on 2nd 24-h recall, in min	1025	255	120
Median difference between 1st and 2nd 24-h recall[b], in min (and in %)	25	−5	0
P difference[c]	.15	.61	.23
Spearman correlation	.75	.64	.94
Women			
Median total time based on 1st 24-h recall, in min	1012	282	85
Median total time based on 2nd 24-h recall, in min	970	320	82
Median difference between 1st and 2nd 24-h recall[b], in min (and in %)	−10 (−1)	0 (+0)	0 (+0)
P difference[c]	.49	.80	.37
Spearman correlation	.75	.66	.92
P value for the influence of gender on the difference between 1st and 2nd 24-h recall[d]	.17	.57	.09
Participants with a BMI<25.0 kg/m^2			
Median total time based on 1st 24-h recall, in min	1035	232	128
Median total time based on 2nd 24-h recall, in min	990	298	128

XSL•FO
RenderX

Stratum and variable	Sedentary activity[a]	Light activity[a]	Moderate to vigorous activity[a]
Median difference between 1st and 2nd 24-h recall[b], in min (and in percent)	5	0	0
P difference[c]	.73	.32	.20
Spearman correlation	.82	.65	.96
Participants with a BMI≥25.0 kg/m^2			
Median total time based on 1st 24-h recall, in min	1000	290	105
Median total time based on 2nd 24-h recall, in min	990	300	70
Median difference between 1st and 2nd 24-h recall[b], in min (and in %)	15 (+1)	−15 (−7)	0 (+0)
P difference[c]	.29	.25	.73
Spearman correlation	.69	.71	.89
P value for the influence of the BMI on the difference between 1st and 2nd 24-h recall[d]	.18	.10	.21

[a]For self-reported physical activity (cpar24), we classified the physical activity intensity according to the corresponding metabolic equivalent of the MET value from the Ainsworth compendium: sedentary activity (MET≤1.5), light physical activity (1.5<MET<3.0), moderate to vigorous physical activity (3.0≤MET).

[b]Please note that the median of the difference between 2 variables does not necessarily correspond to the difference between the medians of the 2 variables.

[c]We tested if the median total time spent in sedentary, light or moderate to vigorous activities varied between the 1st and 2nd 24-h recall using the Wilcoxon signed rank test.

[d]We computed the difference in the total time spent in a specific physical activity intensity level between the 1st and 2nd 24-h recall for each participant, and we tested if that difference varied across the two strata of participants using the Wilcoxon rank sum test.

[e]cpar24: computer-based 24-h physical activity recall.

Discussion

Principal Findings

We assessed the validity, reliability, and usability of the cpar24. Information from the cpar24 was modestly positively correlated with information from accelerometry regarding estimates of the total time spent in sedentary, light, and moderate to vigorous activities. However, as compared with accelerometry, the cpar24 tended to overestimate time spent in moderate to vigorous activities while underestimating time spent in light activities. In contrast, we observed strong positive correlations and no systematic bias between repeated cpar24 assessments. Participants assigned high rankings to the usability of the cpar24, particularly those younger than age 60 years.

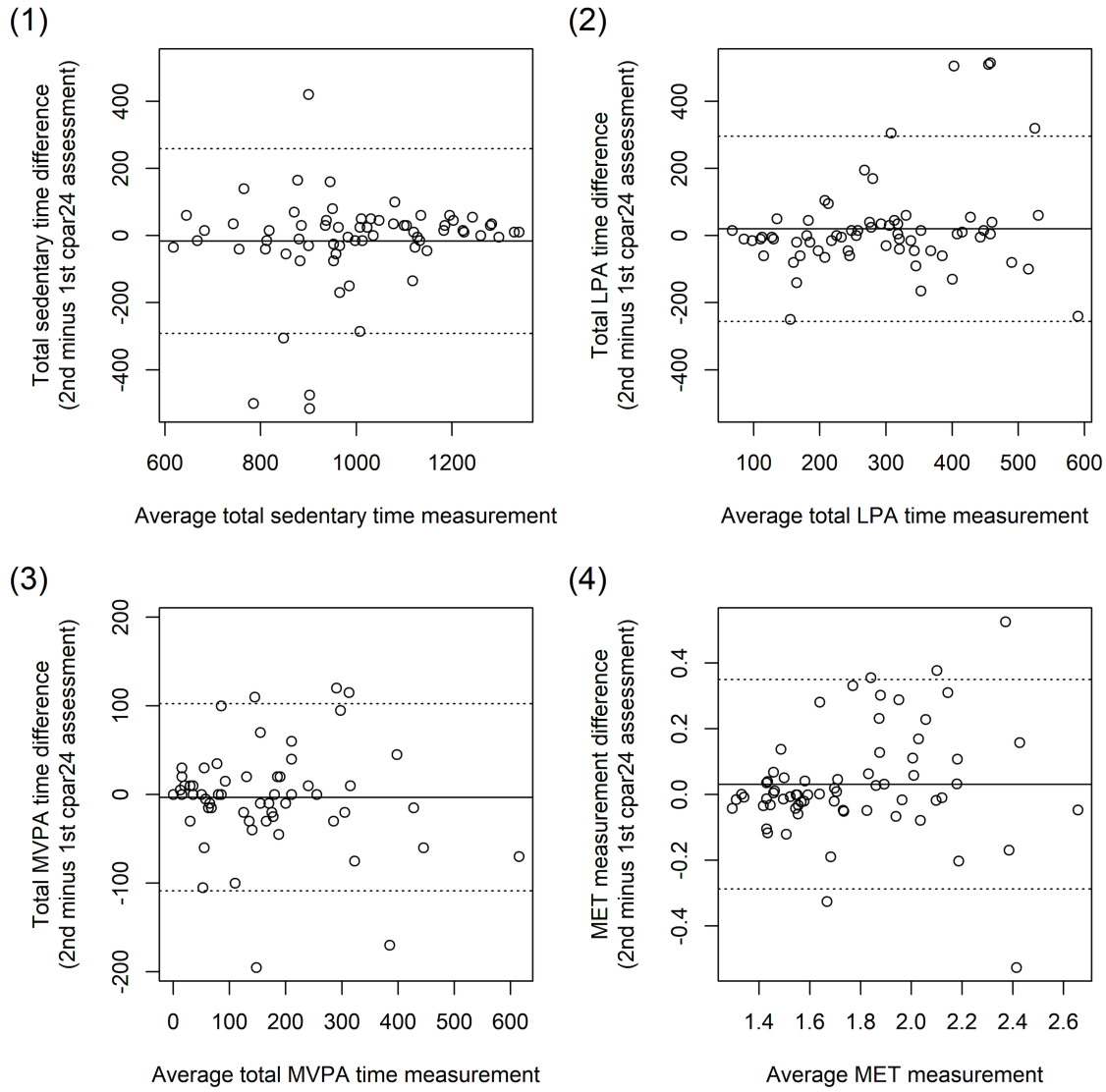
Relevance of Short-Term Physical Activity Recalls to Assess Physical Activity in Epidemiologic Studies

Most available physical activity questionnaires assess the intensity, frequency, and duration of common physical activities performed during the past week, past month, or past year [22]. In the past 20 years, assessments of physical activities of the previous week have become the most prevalent form [23], most likely because estimates of recent past activity patterns (past 24 h, past 7 days) are more accurate than estimates of average physical activity levels representative of longer time periods (eg, past month, past year) [4], leading to an average accelerometer based validity correlation coefficient of .41 for previous week questionnaires as compared with an average correlation coefficient of .30 for previous year assessments [22]. To improve the accuracy of estimates, there have been recommendations to administer multiple short-term physical activity recalls (past 24 h, past 7 days) and to average activity levels over those recalls when using self-reports in large-scale epidemiologic studies [4]. In addition, thanks to recent technologic advances, accelerometers can now be employed to measure short-term physical activity (24 h to 7 days) in large studies [24]. The objective nature of accelerometer measurements represents a potential advantage over self-reported physical activity because the latter may be prone to recall bias and to measurement error resulting from the difficulty of classifying physical activity intensity and from reporting socially desirable physical activity patterns. However, as compared with questionnaires, accelerometers come at the costs of greater logistic burden, increased data complexity, and lower acceptance among participants [24]. In addition, accelerometry has difficulty in recognizing resistance components of activities [25-27], such as carrying heavy objects or ascending stairs. Furthermore, in accelerometry, low and high pass filters are used in an attempt to distinguish human acceleration from noise and from motorized acceleration, implying that accelerometers cannot detect very fast human motion [28]. For example, the ActiGraph digital filter rejects frequencies below 0.25 Hz because those frequencies are mainly associated with gravity acceleration [29] and it rejects frequencies above 2.5 Hz because those frequencies are mainly associated with motorized acceleration (eg, when traveling by car or train). The remaining frequency range of 0.25-2.5 Hz is thought to reflect human body acceleration but it can only

XSL•FO
RenderX

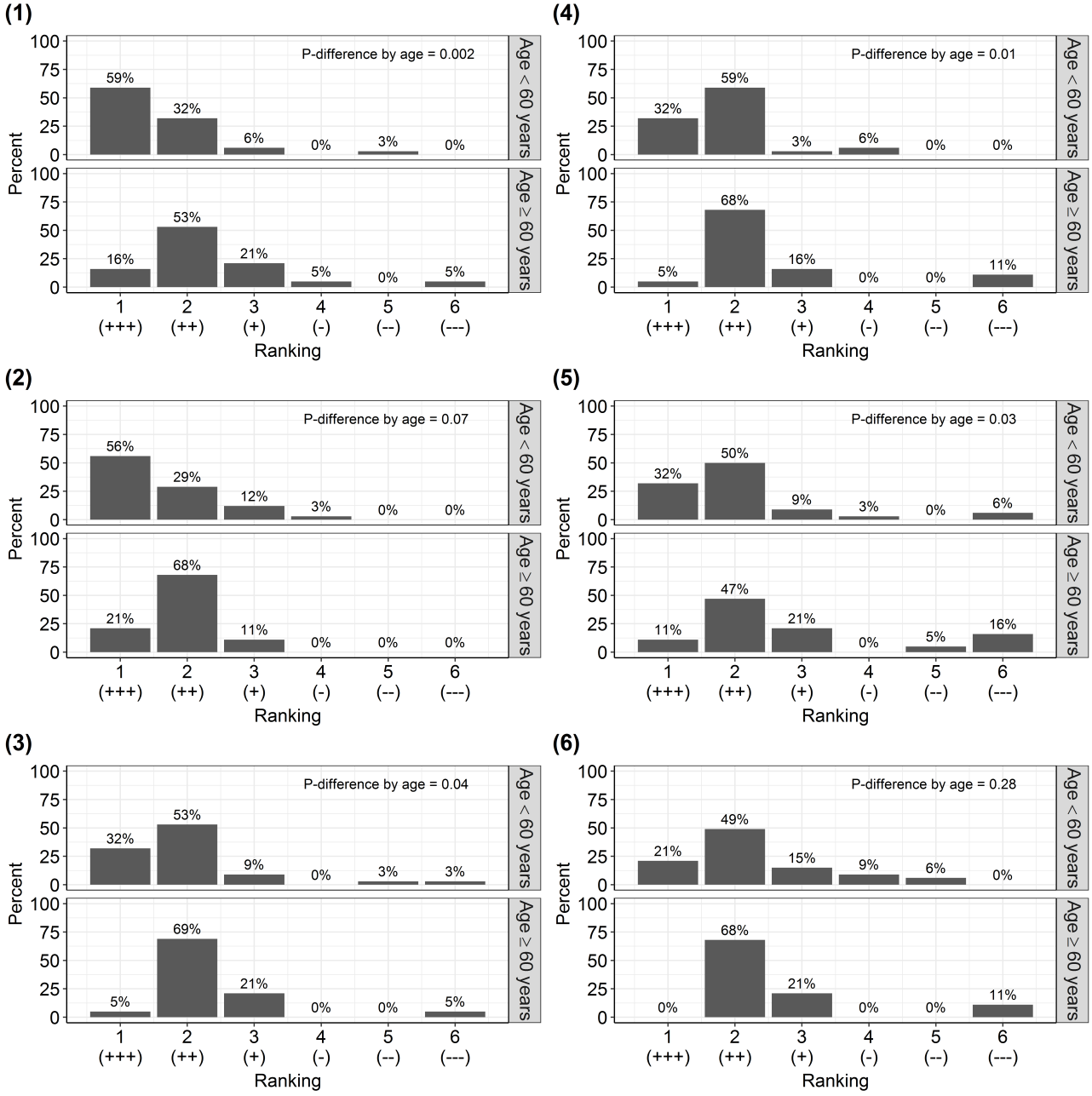
identify gait speeds up to 12 km per h [28]. For higher gait speeds, there is an inverse relation between the true gait speed and the gait speed derived when only considering frequencies in the range of 0.25-2.5 Hz, leading to a circumstance in which frequencies from running at 16, 18, and 20 km per h resemble gait speeds of 10, 8, and 6 km per h, respectively [30,31]. However, few people achieve high gait speeds. In addition, any potential misclassification of the exact speed will not affect estimates for time spent in moderate to vigorous physical activity because even a gait speed of 6 km per h is classified as moderate to vigorous physical activity.

Figure 2. Bland Altman plots comparing data from the second computer-based 24-hour physical activity recall (cpar24) against data from the first cpar24 recall among the 67 participants of the reliability study with respect to (1) the total time spent in sedentary activities, (2) the total time spent in light physical activities, (3) the total time spent in moderate to vigorous physical activities, and (4) the average metabolic equivalent of task (MET) value. LPA=light physical activity; MVPA=moderate to vigorous physical activity; MET=metabolic equivalent of task.



XSL•FO
RenderX

Figure 3. Proportion of the 53 participants of the usability study awarding the ranks 1 (excellent) to 6 (unsatisfactory) to the six items: (1) "How well were you able to recall activities performed yesterday?", (2) "How helpful was the user's manual?", (3) "How helpful were the broad activity categories (eg, household chores, outdoor activities) to find a specific activity?", (4) "How would you rate the overall ease of using the cpar24?", (5) "How well were you able to navigate the cpar24 interface?", and (6) "Do you like the design of the cpar24?", stratified by age group. The heterogeneity across age was assessed using the Wilcoxon rank sum test. Please note that the result of the Wilcoxon rank sum test was not similar for items (3) and (6) in spite of comparable patterns between the age-specific distributions of rankings for items (3) and (6). The reason was that the Wilcoxon rank sum test assessed the difference between medians and not between distributions. If the difference in distributions across age groups had been tested using Fisher exact test, statistically significant difference in the distributions of rankings by age group would have been observed for all items except for item (5).



Validity of Previous Short-Term Physical Activity Recalls as Compared With the Gold Standard of Doubly-Labeled Water

Studies using doubly-labeled water measurements as the gold standard to validate energy expenditure estimates obtained from short-term physical activity recalls (24 h to 7 days) and from accelerometry revealed similar validation correlation coefficients for both methods although, within each method, the validation correlation coefficients of total energy expenditure tended to be greater than the validation correlation coefficients of physical activity energy expenditure (total energy expenditure estimates from short-term physical activity recalls [7,32-39]: average correlation=.57, range=.32-.88; total energy expenditure from accelerometry [40]: average correlation=.52, range=.18-.83; physical activity energy expenditure from short-term physical activity recalls [7,38,39,41,42]: average correlation=.21, range=−.07-.68; physical activity energy expenditure from accelerometry [40]: average correlation=.39, range=−.30-.83). Similarly, there appeared to be less variation across accuracy estimates of total energy expenditure than physical activity energy expenditure for both methods when using doubly-labeled

XSL•FO
RenderX

water as the gold standard (total energy expenditure from short-term physical activity recalls [7,32-35,37-39,42-48]: mean percent difference=7%, range=−27%-37%; total energy expenditure from accelerometry [40]: mean percent difference=−12%, range=−22%-1%; physical activity energy expenditure from short-term physical activity recalls [7,38,39,41,48,49]: mean percent difference=20%, range=−20%-113%; physical activity energy expenditure from accelerometry [40]: mean percent difference=−24%, range=−59%-40%). However, accelerometry tended to underestimate energy expenditure, whereas short-term physical activity recalls tended to overestimate energy expenditure.

Validity of Previous Short-Term Physical Activity Recalls as Compared With Accelerometry

Studies validating physical activity recalls using accelerometry as the gold standard reported stronger average correlation coefficients between energy expenditure estimates (expressed as total energy expenditure, physical activity energy expenditure, average MET per hour, or physical activity MET per week) and accelerometer counts per minute from 24-h recalls than from 7-day recalls (24-h recalls [9,10,41,50,51]: average correlation=.48, range=.23-.82; 7-day recalls [41,51-55]: average correlation=.36, range=−.02-.90). In contrast, average correlations between self-report and accelerometer-based estimates for time spent in sedentary and light activities were greater for 7-day recalls than for 24-h recalls, whereas those for time spent in moderate to vigorous activities agreed across 24-h recalls and 7-day recalls (time spent in sedentary activities among 24-h recalls [10,12]: average correlation=.19, range=−.05-.59; time spent in sedentary activities among 7-day recalls [52,54-57]: average correlation=.37, range=.20-.65; time spent in light activities among 24-h recalls [11,12,58]: average correlation=.18, range=−.16-.45; time spent in light activities among a single 7-day recall [58]: correlation=.37; time spent in moderate to vigorous activities among 24-h recalls [10,11,59,60]: average correlation=.19, range=.05-.26; time spent in moderate to vigorous activities among 7-day recalls [39,55-57,60,61]: average correlation=.26, range=.06-.51). Studies comparing short-term physical activity recalls (24 h to 7 days) with accelerometry tended to report greater estimates of total energy expenditure, light activities, and moderate to vigorous activities (percent difference for total energy expenditure [50,51]: mean=+19%, range=+12%-+31%; for physical activity energy expenditure [49,53]: mean=+87%, range=+80%-+95%; for light activities [11,12,58]: mean=+36%, range=−8%-+107%; for moderate to vigorous activities [11,55,56,59-61]: mean=+260%, range=+29%-+778%). In contrast, there were as many studies overestimating sedentary activities as there were studies underestimating sedentary activities (percent difference for sedentary activities among studies overestimating sedentary activities [11,12,54]: mean=+17%, range=+11%-+27%; among studies underestimating sedentary activities [55-57]: mean=−32%, range=−44% to −13%; among all studies estimating sedentary activities [11,12,54-57]: mean=−4%, range=−44%-+27%).

Reliability of Previous Short-Term Physical Activity Recalls

The reliability correlation coefficients of short term physical activity recalls (24 h to 7 days) appear to decrease with increasing time between replicate measurements. Replicate measurements of a specific 24-h physical activity recall separated by a time lag of 4 hours yielded a positive correlation of .99 [9,10], whereas correlation coefficients for another 24-h physical activity recall varied between .55 and .63 for a time lag of 6 months [12]. Similarly, the smaller the time lag between replicate measurements, the greater the reliability coefficient for 7-day physical activity recalls. Specifically, the reliability coefficient for a time lag of less than a week is .79 (range=.45-.99), for a time lag of 1-4 weeks it is .63 (range=.22-.91), and for a time lag of 2-12 months it is .50 (range=.33-.65) [23]. In contrast, reliability coefficients of 12-month physical activity recalls appear to be less sensitive to the length of the period between measurements. The reliability coefficient for a time lag of less than 1 month is .68 (range=.17-.99) and for a time lag of 2-12 months it is .72 (range=.65-.78) [23]. Reliability coefficients of 7-day physical activity recalls administered less than 1 week apart appeared to be greater for sedentary (mean reliability coefficient=.81, range=.71-.91) [23] than for moderate to vigorous physical activity (mean reliability coefficient=.76, range=.45-.99) [23] and for total energy expenditure (mean reliability coefficient=.73, range=.54-.93) [23]. Two previous studies [9,10] investigated the reliability of a single 24-h physical activity recall with measurements taken 4 hours apart and reported reliability coefficients of .99 each for time spent in moderate to vigorous physical activity and for total energy expenditure. Those studies [9,10] did not report reliability coefficients for total sedentary activity but provided data for sleep (r=.99), screen time (r=.99), and the complement of sedentary time (nonsedentary time, r=.99).

Reliability of the cpar24 in Comparison With Previous Short-Term Physical Activity Recalls

In our study, the reliability correlation coefficients for the total time spent in sedentary (r=.75), light (r=.65), and moderate to vigorous (r=.92) activities, and the reliability correlation coefficient for total energy expenditure (r=.91, assessed as average MET per h) were in the top range of reliability coefficients observed previously for 7-day physical activity recalls administered less than 1 week apart (average correlation=.76, range=.45-.99) [23]. However, the reliability correlation coefficients for our 24-h physical activity recall ranging between .65 and .92 were smaller than those reported for a previous 24-h physical activity recall (all r=.99) [9,10]. To our knowledge, reliability correlation coefficients for additional previous 24-h physical activity recalls are currently not available for further comparison. In line with a previous 24-h recall [9], no statistically significant differences emerged between estimates of average MET and time spent in specific activity intensities obtained from two 24-h physical activity recalls, the second of which was completed 3 hours after completion of the first recall.

XSL•FO
RenderX

Validity of the cpar24 in Comparison With Previous Short-Term Physical Activity Recalls

In our validation study, we deliberately refrained from comparing MET values derived from accelerometer counts with MET values derived from the 24-h physical activity recall because neither method provides accurate MET estimates. In particular, the derivation of MET values from accelerometer counts is challenging, and no conversion rule has been proven universally valid, not even with respect to treadmill walking or running, the discipline for which most formulae were derived [18,25-27,62-69]. Similarly, divergences of measured MET values from the Ainsworth MET values in either direction have been reported for a wide range of activities, including walking or running, ascending or descending stairs, and moving heavy objects [15,70], suggesting that representing a specific activity by a single MET value is challenging.

We found that the validity correlation coefficients for our 24-h physical activity recall for the total time spent in sedentary (r=.54), light (r=.46), and moderate to vigorous activity (r=.50) were superior to the average validity correlation coefficients reported for previous 24-h physical activity recalls evaluated against accelerometry (validity correlation coefficient for sedentary activity [10,12]: mean=.19, range=−.05-.59; for light physical activity [11,12,58]: mean=.18, range=−.16-.45; for moderate to vigorous physical activity [10,11,59,60]: mean=.19, range=.05-.26). The validity correlation coefficients of our 24-h physical activity recall were also in the top range when compared with previous 7-day physical activity recalls evaluated against accelerometry (validity correlation coefficient for sedentary activity [52,54-57]: mean=.37, range=.20-.65; for light physical activity [58]: mean=.37; for moderate to vigorous physical activity [39,55-57,60,61]: mean=.26, range=.06-.51).

When comparing cpar24 data with accelerometer data, we found that the cpar24 statistically significantly overestimated moderate to vigorous physical activity time by +353%, which was greater than the average overestimation of +260% (range=+29%-+778%) reported in 6 previous studies [11,55,56,59-61] evaluating short-term recalls (24 h to 7 days) against accelerometry. In our study, the statistically significant overestimation of moderate to vigorous physical activity corresponded to an absolute difference of 89 min, and it was compensated by a statistically significant underestimation of time spent in light activities by −123 minutes (−28%). By comparison, previous short-term physical activity recalls (24 h to 7 days) [11,12,58] tended to overestimate light activities by an average of 36% (range −8%-107%) as compared with accelerometry. In contrast to previous statistically significant over-reporting of sedentary time [11,12,54] (by an average of +17%, range=+11%-+27%) and in contrast to previous statistically significant under-reporting of sedentary time [55-57] (by an average of −32%, range=−44% to −13%), we observed a small, statistically nonsignificant overall difference of +1% between cpar24 and accelerometer estimates of sedentary time in our study. Yet, Bland Altman plots for our study revealed that the overall difference of +1% between cpar24 and accelerometer estimates of sedentary time resulted from an averaging out of over-reporting of sedentary time among sedentary participants and under-reporting of sedentary time among physically active participants. Similar observations were made in previous studies [10,54-56]. In addition, over-reporting of moderate to vigorous physical activities was stronger among physically active than sedentary participants in our and other studies [54-56,59,61].

In stratified analyses, we observed no statistically significant differences between cpar24 and accelerometer data regarding estimates of activities of various intensities across strata defined by age (aged < 60 years, aged ≥60 years), gender (men, women), and BMI (BMI<25 kg/m^2, BMI≥25 kg/m^2) (all P difference≥.23) as did several previous studies [11,12,57,59-61]. In contrast, one previous 7-day physical activity recall evaluation study [54] found that over-reporting of sedentary activities was statistically significantly greater among men as compared to women and among participants aged 18-34 years as compared with participants aged 50 years or more, whereas under-reporting of moderate to vigorous physical activities was greater among normal weight participants than among overweight and obese participants. In contrast to that study [54], another previous 7-day physical activity recall evaluation study [55] reported less over-reporting of moderate to vigorous physical activities among participants aged 18-39 years as compared with participants aged 65 years or more, whereas no statistically significant differences were seen for moderate to vigorous activities across gender and for sedentary activities across age and gender.

Strengths and Limitations

An important strength of our study is the use of accelerometry as objective comparison criterion, which enabled us to validate our estimates of total time spent in sedentary, light, and moderate to vigorous activities. Furthermore, the inclusion of a random sample of men and women aged 22 to 70 years from the general population allowed us to demonstrate the applicability of our 24-h physical activity recall to the general population. In addition, we conducted extensive comparisons between the validity and reliability correlation coefficients observed for our 24-h recall with those reported for a wide range of existing 24-h to 7-day physical activity recalls. We found that the validity and reliability correlation coefficients of our 24-h physical activity recall were in the top range of those reported for previous 24-h to 7-day physical activity recalls.

One limitation of our study is that we were not able to validate resistance-based activities (eg, stair climbing or carrying heavy loads) and vehicle-based activities (driving a car or cycling) due to the technologic limitations of our accelerometer. To close that gap, behavior recognition methods based on simultaneous monitoring of heart rate, body heat, body motion and position, limb motion and position, foot pressure, global positioning system, and barometric pressure are currently being evaluated [71]. Furthermore, we were not able to evaluate the absolute validity of total energy expenditure estimates, which should be done in future studies. In addition, we did not investigate the within-person variation in accelerometer and cpar24 measurements across different days and different seasons, and we can therefore not comment on how many days of measurements are required to obtain reliable physical activity estimates for a specific study period. Further studies are required

XSL•FO
RenderX

to investigate the validity and reliability of cpar24 recalls to estimate average physical activity levels for longer study periods, to examine the influences of season and day of the week on the validity and reliability of those estimates, and to compare those estimates against estimates obtained from physical activity questionnaires covering the same study period.

Conclusions

In conclusion, our cpar24 is a feasible, valid, reliable and user-friendly assessment of physical activity in adults. It provides estimates of total energy expenditure and time spent in sedentary, light, and moderate to vigorous activities with above-average validity correlation coefficients of .46 to .54 as compared with previous 24-h recall instruments. While we were able to establish the relative validity of our instrument as compared with accelerometer measurements, future studies are needed to verify the absolute validity of our cpar24.

Conflicts of Interest

None declared.

References

1. Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. Eur J Epidemiol 2015 Jul;30(7):529-542. [doi: 10.1007/s10654-015-0056-z] [Medline: 26092138]
2. Li J, Loerbroks A, Angerer P. Physical activity and risk of cardiovascular disease: what does the new epidemiological evidence show? Curr Opin Cardiol 2013 Sep;28(5):575-583. [doi: 10.1097/HCO.0b013e328364289c] [Medline: 23928923]
3. Liu L, Shi Y, Li T, Qin Q, Yin J, Pang S, et al. Leisure time physical activity and cancer risk: evaluation of the WHO's recommendation based on 126 high-quality epidemiological studies. Br J Sports Med 2016 Mar;50(6):372-378. [doi: 10.1136/bjsports-2015-094728] [Medline: 26500336]
4. Matthews CE, Moore SC, George SM, Sampson J, Bowles HR. Improving self-reports of active and sedentary behaviors in large epidemiologic studies. Exerc Sport Sci Rev 2012 Jul;40(3):118-126. [doi: 10.1097/JES.0b013e31825b34a0] [Medline: 22653275]
5. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L, Experts Panel. Assessment of physical activity - a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. Eur J Cardiovasc Prev Rehabil 2010 Apr;17(2):127-139. [doi: 10.1097/HJR.0b013e32832ed875] [Medline: 20215971]
6. Lagerros YT, Lagiou P. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. Eur J Epidemiol 2007;22(6):353-362. [doi: 10.1007/s10654-007-9154-x] [Medline: 17599242]
7. Namba H, Yamaguchi Y, Yamada Y, Tokushima S, Hatamoto Y, Sagayama H, et al. Validation of Web-based physical activity measurement systems using doubly labeled water. J Med Internet Res 2012;14(5):e123 [FREE Full text] [doi: 10.2196/jmir.2253] [Medline: 23010345]
8. Calabro MA, Welk GJ, Carriquiry AL, Nusser SM, Beyler NK, Mathews CE. Validation of a computerized 24-hour physical activity recall (24PAR) instrument with pattern-recognition activity monitors. J Phys Act Health 2009 Mar;6(2):211-220. [Medline: 19420399]
9. Gomersall SR, Olds TS, Ridley K. Development and evaluation of an adult use-of-time instrument with an energy expenditure focus. J Sci Med Sport 2011 Mar;14(2):143-148. [doi: 10.1016/j.jsams.2010.08.006] [Medline: 20932797]
10. Mace CJ, Maddison R, Olds T, Kerse N. Validation of a computerized use of time recall for activity measurement in advanced-aged adults. J Aging Phys Act 2014 Apr;22(2):245-254. [doi: 10.1123/japa.2012-0280] [Medline: 23752221]
11. Matthews CE, Keadle SK, Sampson J, Lyden K, Bowles HR, Moore SC, et al. Validation of a previous-day recall measure of active and sedentary behaviors. Med Sci Sports Exerc 2013 Aug;45(8):1629-1638 [FREE Full text] [doi: 10.1249/MSS.0b013e3182897690] [Medline: 23863547]
12. Sternfeld B, Jiang S, Picchi T, Chasan-Taber L, Ainsworth B, Quesenberry CP. Evaluation of a cell phone-based physical activity diary. Med Sci Sports Exerc 2012 Mar;44(3):487-495. [doi: 10.1249/MSS.0b013e3182325f45] [Medline: 21857369]
13. German National Cohort (GNC) Consortium. The German National Cohort: aims, study design and organization. Eur J Epidemiol 2014 May;29(5):371-382 [FREE Full text] [doi: 10.1007/s10654-014-9890-7] [Medline: 24840228]
14. Kitterød RH. Does the recording of parallel activities in time use diaries affect the way people report their main activities? Soc Indic Res 2001;56(2):145-178. [doi: 10.1023/A:1012289811886]
15. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 Compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
16. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis 2015;57(4):387-395. [doi: 10.1016/j.pcad.2014.10.005] [Medline: 25446555]

XSL•FO
RenderX

17. Pedišić Ž, Bauman A. Accelerometer-based measures in physical activity surveillance: current practices and issues. Br J Sports Med 2015 Feb;49(4):219-223. [doi: 10.1136/bjsports-2013-093407] [Medline: 25370153]
18. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
19. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
21. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
22. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. Sports Med 2010 Jul 1;40(7):565-600. [doi: 10.2165/11531930-000000000-00000] [Medline: 20545381]
23. Helmerhorst HJF, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act 2012;9:103 [FREE Full text] [doi: 10.1186/1479-5868-9-103] [Medline: 22938557]
24. Lee I, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. Br J Sports Med 2014 Feb;48(3):197-201 [FREE Full text] [doi: 10.1136/bjsports-2013-093154] [Medline: 24297837]
25. Crouter SE, Churilla JR, Bassett DR. Estimating energy expenditure using accelerometers. Eur J Appl Physiol 2006 Dec;98(6):601-612. [doi: 10.1007/s00421-006-0307-5] [Medline: 17058102]
26. Crouter SE, Clowers KG, Bassett DR. A novel method for using accelerometer data to predict energy expenditure. J Appl Physiol (1985) 2006 Apr;100(4):1324-1331 [FREE Full text] [doi: 10.1152/japplphysiol.00818.2005] [Medline: 16322367]
27. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. Eur J Appl Physiol 2011 Feb;111(2):187-201 [FREE Full text] [doi: 10.1007/s00421-010-1639-8] [Medline: 20842375]
28. John D, Miller R, Kozey-Keadle S, Caldwell G, Freedson P. Biomechanical examination of the 'plateau phenomenon' in ActiGraph vertical activity counts. Physiol Meas 2012 Feb;33(2):219-230 [FREE Full text] [doi: 10.1088/0967-3334/33/2/219] [Medline: 22260902]
29. Mathie MJ, Coster AC, Lovell NH, Celler BG. Detection of daily physical activities using a triaxial accelerometer. Med Biol Eng Comput 2003 May;41(3):296-301. [Medline: 12803294]
30. Rowlands AV, Stone MR, Eston RG. Influence of speed and step frequency during walking and running on motion sensor output. Med Sci Sports Exerc 2007 Apr;39(4):716-727. [doi: 10.1249/mss.0b013e318031126c] [Medline: 17414811]
31. John D, Tyo B, Bassett DR. Comparison of four ActiGraph accelerometers during walking and running. Med Sci Sports Exerc 2010 Feb;42(2):368-374 [FREE Full text] [doi: 10.1249/MSS.0b013e3181b3af49] [Medline: 19927022]
32. Bonnefoy M, Normand S, Pachiaudi C, Lacour JR, Laville M, Kostka T. Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study. J Am Geriatr Soc 2001 Jan;49(1):28-35. [Medline: 11207839]
33. Koebnick C, Wagner K, Thielecke F, Moeseneder J, Hoehne A, Franke A, Zunft H J F. Validation of a simplified physical activity record by doubly labeled water technique. Int J Obes (Lond) 2005 Mar;29(3):302-309. [doi: 10.1038/sj.ijo.0802882] [Medline: 15672111]
34. Paul DR, Rhodes DG, Kramer M, Baer DJ, Rumpler WV. Validation of a food frequency questionnaire by direct measurement of habitual ad libitum food intake. Am J Epidemiol 2005 Oct 15;162(8):806-814. [doi: 10.1093/aje/kwi279] [Medline: 16120695]
35. Rothenberg E, Bosaeus I, Lernfelt B, Landahl S, Steen B. Energy intake and expenditure: validation of a diet history by heart rate monitoring, activity diary and doubly labeled water. Eur J Clin Nutr 1998 Nov;52(11):832-838. [Medline: 9846597]
36. Rush EC, Valencia ME, Plank LD. Validation of a 7-day physical activity diary against doubly-labelled water. Ann Hum Biol 2008;35(4):416-421. [doi: 10.1080/03014460802089825] [Medline: 18608111]
37. Schulz S, Westerterp KR, Brück K. Comparison of energy expenditure by the doubly labeled water technique with energy intake, heart rate, and activity recording in man. Am J Clin Nutr 1989 Jun;49(6):1146-1154. [Medline: 2729154]
38. Washburn RA, Jacobsen DJ, Sonko BJ, Hill JO, Donnelly JE. The validity of the stanford seven-day physical activity recall in young adults. Med Sci Sports Exerc 2003 Aug;35(8):1374-1380. [doi: 10.1249/01.MSS.0000079081.08476.EA] [Medline: 12900693]
39. Mâsse LC, Fulton JE, Watson KB, Tortolero S, Kohl HW, Meyers MC, et al. Comparing the validity of 2 physical activity questionnaire formats in African-American and Hispanic women. J Phys Act Health 2012 Feb;9(2):237-248. [Medline: 22368223]
40. Van Remoortel H, Giavedoni S, Raste Y, Burtin C, Louvaris Z, Gimeno-Santos E, PROactive consortium. Validity of activity monitors in health and chronic disease: a systematic review. Int J Behav Nutr Phys Act 2012;9:84 [FREE Full text] [doi: 10.1186/1479-5868-9-84] [Medline: 22776399]

XSL•FO
RenderX

41. Adams SA, Matthews CE, Ebbeling CB, Moore CG, Cunningham JE, Fulton J, et al. The effect of social desirability and social approval on self-reports of physical activity. Am J Epidemiol 2005 Feb 15;161(4):389-398 [FREE Full text] [doi: 10.1093/aje/kwi054] [Medline: 15692083]
42. Maddison R, Mhurchu CN, Jiang Y, Vander Hoorn S, Rodgers A, Lawes CMM, et al. International physical activity questionnaire (IPAQ) and New Zealand physical activity questionnaire (NZPAQ): a doubly labelled water validation. Int J Behav Nutr Phys Act 2007 Dec 03;4:62 [FREE Full text] [doi: 10.1186/1479-5868-4-62] [Medline: 18053188]
43. Clark D, Tomas F, Withers RT, Chandler C, Brinkman M, Phillips J, et al. Energy metabolism in free-living, 'large-eating' and 'small-eating' women: studies using 2H2(18)O. Br J Nutr 1994 Jul;72(1):21-31. [Medline: 7918325]
44. Conway JM, Irwin ML, Ainsworth BE. Estimating energy expenditure from the Minnesota leisure time physical activity and Tecumseh occupational activity questionnaires - a doubly labeled water validation. J Clin Epidemiol 2002 Apr;55(4):392-399. [Medline: 11927208]
45. Irwin ML, Ainsworth BE, Conway JM. Estimation of energy expenditure from physical activity measures: determinants of accuracy. Obes Res 2001 Sep;9(9):517-525 [FREE Full text] [doi: 10.1038/oby.2001.68] [Medline: 11557832]
46. Racette SB, Schoeller DA, Kushner RF. Comparison of heart rate and physical activity recall with doubly labeled water in obese women. Med Sci Sports Exerc 1995 Jan;27(1):126-133. [Medline: 7898328]
47. Seale JL, Klein G, Friedmann J, Jensen GL, Mitchell DC, Smiciklas-Wright H. Energy expenditure measured by doubly labeled water, activity recall, and diet records in the rural elderly. Nutrition 2002;18(7-8):568-573. [Medline: 12093431]
48. Staten LK, Taren DL, Howell WH, Tobar M, Poehlman ET, Hill A, et al. Validation of the Arizona activity frequency questionnaire using doubly labeled water. Med Sci Sports Exerc 2001 Nov;33(11):1959-1967. [Medline: 11689750]
49. Leenders NY, Sherman WM, Nagaraja HN, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. Med Sci Sports Exerc 2001 Jul;33(7):1233-1240. [Medline: 11445774]
50. Bharathi AV, Kuriyan R, Kurpad AV, Thomas T, Ebrahim S, Kinra S, Indian Migrants Study Group. Assessment of physical activity using accelerometry, an activity diary, the heart rate method and the Indian migration study questionnaire in south Indian adults. Public Health Nutr 2010 Jan;13(1):47-53. [doi: 10.1017/S1368980009005850] [Medline: 19656418]
51. Matthews CE, Freedson PS. Field trial of a three-dimensional activity monitor: comparison with self report. Med Sci Sports Exerc 1995 Jul;27(7):1071-1078. [Medline: 7564975]
52. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
53. Leenders N, Sherman WM, Nagaraja HN. Comparisons of four methods of estimating physical activity in adult women. Med Sci Sports Exerc 2000 Jul;32(7):1320-1326. [Medline: 10912900]
54. Hagstromer M, Ainsworth BE, Oja P, Sjostrom M. Comparison of a subjective and an objective measure of physical activity in a population sample. J Phys Act Health 2010 Jul;7(4):541-550. [Medline: 20683097]
55. Wanner M, Probst-Hensch N, Kriemler S, Meier F, Autenrieth C, Martin BW. Validation of the long international physical activity questionnaire: influence of age and language region. Prev Med Rep 2016 Jun;3:250-256 [FREE Full text] [doi: 10.1016/j.pmedr.2016.03.003] [Medline: 27419023]
56. Celis-Morales CA, Perez-Bravo F, Ibañez L, Salas C, Bailey ME, Gill JM. Objective vs. self-reported physical activity and sedentary time: effects of measurement method on relationships with risk biomarkers. PLoS One 2012;7(5):e36345 [FREE Full text] [doi: 10.1371/journal.pone.0036345] [Medline: 22590532]
57. Grimm EK, Swartz AM, Hart T, Miller NE, Strath SJ. Comparison of the IPAQ-Short Form and accelerometry predictions of physical activity in older adults. J Aging Phys Act 2012 Jan;20(1):64-79. [Medline: 22190120]
58. Macfarlane DJ, Lee CCY, Ho EYK, Chan KL, Chan D. Convergent validity of six methods to assess physical activity in daily life. J Appl Physiol (1985) 2006 Nov;101(5):1328-1334 [FREE Full text] [doi: 10.1152/japplphysiol.00336.2006] [Medline: 16825525]
59. Schmidt MD, Freedson PS, Chasan-Taber L. Estimating physical activity using the CSA accelerometer and a physical activity log. Med Sci Sports Exerc 2003 Sep;35(9):1605-1611. [doi: 10.1249/01.MSS.0000084421.97661.17] [Medline: 12972884]
60. Lee RE, Mama SK, Medina AV, Reese-Smith JY, Banda JA, Layne CS, et al. Multiple measures of physical activity, dietary habits and weight status in African American and Hispanic or Latina women. J Community Health 2011 Dec;36(6):1011-1023 [FREE Full text] [doi: 10.1007/s10900-011-9403-5] [Medline: 21519867]
61. Kwak L, Hagströmer M, Sjostrom M. Can the IPAQ-long be used to assess occupational physical activity? J Phys Act Health 2012 Nov;9(8):1130-1137. [Medline: 22207150]
62. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S450-S456. [Medline: 10993414]
63. Brooks AG, Gunn SM, Withers RT, Gore CJ, Plummer JL. Predicting walking METs and energy expenditure from speed or accelerometry. Med Sci Sports Exerc 2005 Jul;37(7):1216-1223. [Medline: 16015141]
64. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S442-S449. [Medline: 10993413]

XSL•FO
RenderX

65. Heil DP, Higginson BK, Keller CP, Juergens CA. Body size as a determinant of activity monitor output during overground walking. JEPonline 2003;6:1-11.
66. Leenders NYJM, Nelson TE, Sherman WM. Ability of different physical activity monitors to detect movement during treadmill walking. Int J Sports Med 2003 Jan;24(1):43-50. [doi: 10.1055/s-2003-37196] [Medline: 12582951]
67. Yngve A, Nilsson A, Sjostrom M, Ekelund U. Effect of monitor placement and of activity setting on the MTI accelerometer output. Med Sci Sports Exerc 2003 Feb;35(2):320-326. [doi: 10.1249/01.MSS.0000048829.75758.A0] [Medline: 12569223]
68. Rothney MP, Brychta RJ, Meade NN, Chen KY, Buchowski MS. Validation of the ActiGraph two-regression model for predicting energy expenditure. Med Sci Sports Exerc 2010 Sep;42(9):1785-1792 [FREE Full text] [doi: 10.1249/MSS.0b013e3181d5a984] [Medline: 20142778]
69. Leenders NY, Sherman WM, Nagaraja HN. Energy expenditure estimated by accelerometry and doubly labeled water: do they agree? Med Sci Sports Exerc 2006 Dec;38(12):2165-2172. [doi: 10.1249/01.mss.0000235883.94357.95] [Medline: 17146325]
70. Kozey SL, Lyden K, Howe CA, Staudenmayer JW, Freedson PS. Accelerometer output and MET values of common physical activities. Med Sci Sports Exerc 2010 Sep;42(9):1776-1784 [FREE Full text] [doi: 10.1249/MSS.0b013e3181d479f2] [Medline: 20142781]
71. Intille SS, Lester J, Sallis JF, Duncan G. New horizons in sensor development. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S24-S31 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399c7d] [Medline: 22157771]

Abbreviations

BMI: body mass index
cpar24: computer-based 24-hour physical activity recall instrument
MET: metabolic equivalent of task

Edited by G Eysenbach; submitted 01.03.17; peer-reviewed by P Bergman, Y Yamada; comments to author 23.03.17; revised version received 03.04.17; accepted 03.04.17; published 30.05.17.

Please cite as:
Kohler S, Behrens G, Olden M, Baumeister SE, Horsch A, Fischer B, Leitzmann MF
Design and Evaluation of a Computer-Based 24-Hour Physical Activity Recall (cpar24) Instrument
J Med Internet Res 2017;19(5):e186
URL: http://www.jmir.org/2017/5/e186/
doi:10.2196/jmir.7620
PMID:28559229

©Simone Kohler, Gundula Behrens, Matthias Olden, Sebastian E Baumeister, Alexander Horsch, Beate Fischer, Michael F Leitzmann. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Improving Primary Health Care Workers' Knowledge About Public Health Services in Rural China: A Comparative Study of Blended Learning and Pure E-Learning

Xingxin Zhan[1], PhD; Zhixia Zhang[1], PhD; Fang Sun[1], MPH; Qian Liu[1], MMSc; Weijun Peng[1], MMSc; Heng Zhang[1], MMSc; Weirong Yan[1], PhD

Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, China

Corresponding Author:
Weirong Yan, PhD
Department of Epidemiology and Biostatistics
School of Public Health
Tongji Medical College, Huazhong University of Science & Technology
No.13 Hangkong Road
Qiaokou District
Wuhan, 430030
China
Phone: 86 (027)83650713
Fax: 86 (027)83693763
Email: weirong.yan@hust.edu.cn

Abstract

Background: Primary health care workers (PHCWs) are a major force in delivering basic public health services (BPHS) in rural China. It is necessary to take effective training approaches to improve PHCWs' competency on BPHS. Both electronic learning (e-learning) and blended learning have been widely used in the health workers' education. However, there is limited evidence on the effects of blended learning in comparison with pure e-learning.

Objective: The aim of this study was to evaluate the effects of a blended-learning approach for rural PHCWs in improving their knowledge about BPHS as well as training satisfaction in comparison with a pure e-learning approach.

Methods: The study was conducted among PHCWs in 6 rural counties of Hubei Province, China, between August 2013 and April 2014. Three counties were randomly allocated blended-learning courses (29 township centers or 612 PHCWs—the experimental group), and three counties were allocated pure e-learning courses (31 township centers or 625 PHCWs—the control group). Three course modules were administered for 5 weeks, with assessments at baseline and postcourse. Primary outcomes were score changes in courses' knowledge. Secondary outcome was participant satisfaction (5-point Likert scale anchored between 1 [strongly agree] and 5 [strongly disagree]).

Results: The experimental group had higher mean scores than the control group in knowledge achievement in three course modules: (1) module 1: 93.21 (95% CI 92.49-93.93) in experimental group versus 88.29 (95% CI 87.19-89.40) in the control group; adjusted difference, 4.92 (95% CI 2.61-7.24; *P*<.001); (2) module 2: 94.05 (95% CI 93.37-94.73) in the experimental group vs 90.22 (95% CI 89.12-91.31) in the control group; adjusted difference, 3.67 (95% CI 1.17-6.18; *P*=.004); (3) module 3: 93.88 (95% CI 93.08-94.68) in the experimental group versus 89.09 (95% CI 87.89-90.30) in control group; adjusted difference, 4.63 (95% CI 2.12-7.14; *P*<.001). The participants in the experimental learning group gave more positive responses with the four issues than control group participants: (1) the increase of interest in learning, 1.85 (95% CI 1.22-2.80; *P*=.003); (2) the increase of interaction with others, 1.77 (95% CI 1.20-2.60; *P*=.004); (3) the satisfaction with learning experience, 1.78 (95% CI 1.11-2.88; *P*=.02); and (4) achievement of learning objectives, 1.63 (95% CI 1.08-2.48; *P*=.02).

Conclusions: Among PHCWs in rural China, a blended-learning approach to BPHS training could result in a higher knowledge achievement and satisfaction level compared with a pure e-learning approach. The findings of the study will contribute knowledge to improve the competency of PHCWs in similar settings.

(J Med Internet Res 2017;19(5):e116) doi:10.2196/jmir.6453

XSL•FO
RenderX

KEYWORDS

blended learning; e-learning; primary health care workers; public health services

Introduction

Background

In rural China, health services were delivered by a 3-tiered system consisting of county-level health care facilities, township hospitals, and village health clinics [1]. With the goal of providing affordable and equitable basic health care for all residents by 2020, the Chinese government launched a health care reform plan in April 2009. One of the main measures of this plan is the provision of a package of basic public health services (BPHS) for all residents. In 2015, the BPHS package included 13 kinds of services: health records management for residents; health education; vaccination; health management for children under 6 years of age; maternal health care; health care for the elderly; health care management of patients with hypertension, type 2 diabetes, severe mental illness, or tuberculosis (TB); reporting of infectious diseases and public health emergencies; health administrative oversight; and health management of Chinese traditional medicine [2]. Primary health care workers (PHCWs), especially those from village clinics and township hospitals, are at the bottom tier in terms of delivering most BPHS to rural residents.

Human resources is the crucial core of a health system, especially with regard to quantity and quality [3]. The competency of PHCWs can affect the delivery of BPHS in rural China, particularly the service quality [4]. Previous studies have revealed that most PHCWs, especially village doctors, have poor education and insufficient competency to provide high quality service [4,5]. One important strategy to improve the competency of PHCWs is training [6]. At present, the main training mode for PHCWs is the traditional face-to-face training [7], but its inflexibility, time constraints, travel costs, and limited training opportunities have negative effects on training [8,9]. Our previous qualitative study showed that the BPHS training was inadequate and ineffective in rural China [9]. Thus, there is a need for more effective solutions for training rural PHCWs on BPHS.

The increased popularity of the Internet and the growth of computer processing power during the past decade have provided opportunities for innovation and new approaches for training [10]. Alternatives to the traditional face-to-face training delivery, electronic learning (e-learning), and blended learning (a combination of e-learning and face-to-face learning) have been widely used in the health workers education [11-14]. Cook et al's [15] systematic review reported that Internet-based learning had more positive effects when compared with no intervention in health professions, but more comparisons of different Internet-based interventions need to be conducted. To our knowledge, there is limited evidence on the effects of blended learning in comparison with pure e-learning [16-19]. In another systematic review in 2016, Liu et al [20] showed that blended learning is more effective or at least as effective as pure e-learning or pure traditional face-to-face learning among health professions and suggested that the more evaluation studies of blended learning, especially with e-learning should be conducted in future research.

Aim of This Study

On the basis of the fact that most PHCWs in rural China need more effective training modes to improve their knowledge on BPHS, our study aimed to evaluate the effects of a blended-learning approach in improving BPHS knowledge among PHCWs in comparison with a pure e-learning approach.

Methods

Study Design, Setting, and Participants

A comparative study was conducted in 3 cities (Yichang, Ezhou, and Xianning) in Hubei Province between August 2013 and April 2014. A multistage clustering sampling method was used to select participants in this study. In the first stage, according to their gross domestic product (GDP) rank in 2013 in Hubei Province, the cities of Yichang, Xianning, and Ezhou city were selected (low: Ezhou; medium: Xianning; high: Yichang). In the second stage, 2 counties with similar background characteristics in each city were selected; a total of 6 counties (Yiling and Zhijiang from Yichang city, Xianan and Chibi from Xianning city, and Huarong and Liangzihu District from Ezhou city) with 60 township centers were approached. In the third stage, the 2 counties in each city were randomly allocated to 2 groups, and therefore 3 counties, including 29 township centers were included in the blended-learning group (Zhijiang, Xianan, and Huarong counties; experimental intervention, 612 participants), and the other 3 counties, including 31 township centers, were in the pure e-learning group (Yiling, Chibi, and Liangzihu counties; control intervention, 625 participants). The selected counties in each city were at an average distance of more than 43 km.

Included participants were PHCWs, either from township centers or village clinics within the administrative prefecture of each selected township, who are currently providing BPHS to rural residents. Exclusion criteria were refusal to provide informed consent, lack of space to attend the training, lack the basic computer skills, or lack of an Internet connection.

Intervention and Data Collection

Three course modules were developed based on the BPHS contents [21]: Course module 1: health management of patients with hypertension; course module 2: health records management for residents; and course module 3: vaccination. Each course module consisted of 2 parts: theoretical learning and case studies. Both the theoretical and case materials were piloted in township centers and modified according to the feedback from interviews with experts and PHCWs in primary health institutions. The experimental and control groups had the same course materials. The public health services Web-based training platform based on Moodle was created for the study from August to October 2013 [22]. In addition, PHCWs outside the study area were invited to test the ease of use and stability of the training platform during the development period to ensure

XSL•FO
RenderX

normal use of the platform. The experimental group received theoretical knowledge on the training platform and the cases delivered through the face-to-face method. In the control group, both theoretical knowledge and cases were delivered by the training platform.

All participants were enrolled in the study for an overall period of 5 weeks (1 week for trainees to familiarize themselves with training platform; 3 weeks for the theoretical learning; and 1 week for the case study). Before theoretical learning, all trainees could have access to the manual about training platform for 1 week and receive training or guidance for using the platform. I For the sake of consistency between the two groups, all study subjects were required to complete the theoretical learning of the three course modules first before starting the case studies. During the intervention implementation period, there was no regular meeting held at the county CDC (Center for Disease Control and Prevention) to reduce contaminations between the two intervention groups. Two facilitators were present during the training sessions of both groups for assistance and to answer questions. The details for the interventions are as follows.

Control Group

The pure e-learning group received Internet training on the training platform. Theoretical learning was presented in the format of Microsoft PowerPoint with 5-6 questions inserted into the slides, and a synchronous audio explanation was attached in each slide. Case studies consisted of 3 video sessions in which “real-world” examples or cases were delivered by a lecturer. Each case-study video, consisting of 4-5 cases, was about 30 min in length. All learning activities had to be completed independently at a self-paced rate. Two discussion forums were developed on the training platform, for the theoretical learning and case studies respectively. The discussion forum for the theoretical learning was set to separate groups, meaning only the same group learners could discuss and talk to each other, to reduce contaminations between the two intervention groups [23]. Another discussion forum for case studies was only available to pure e-learning trainees, and it encouraged them to discuss cases and ask questions.

Experimental Group

Participants in the blended group studied the same PowerPoint-based theoretical materials (available at the same training platform) during the same period. After that, participants received the handouts of all case-study materials for self-studying 4-5 days and attended 1-day (8-h) face-to-face case-study training. All cases were administered on the day by the same lecturers as in the videos in the meeting room at county CDC. PHCWs were encouraged to discuss the cases with educators and other physicians during the face-to-face training.

Assessments

Assessment instruments consisted of two parts: the same pre- and posttest multiple-choice questions (MCQ) test in a different order to evaluate knowledge achievement, and a questionnaire to evaluate trainees’ satisfaction. Each trainee at the start answered the pretest questionnaire to gain access to the three training course modules for 4 weeks. After 4 weeks of learning, trainees were asked to complete the posttest MCQ for three course modules. Due to the various dropouts from each course module, there were different numbers of participants in each course training. After the completion of the three course modules, all participants were asked to fill out an online evaluation questionnaire during the following week.

MCQ Test to Knowledge

A similar pre- and a posttest questionnaire was developed to measure trainees' knowledge achievement in each course module. A total of 3 knowledge MCQ tests were developed, consisting of a 10-item MCQ test in course module 1, a 15-item MCQ test in course module 2, and a 20-item MCQ test in course module 3. Both groups finished the precourse MCQ tests online within 60 min (each MCQ test under 20 min). Experimental group learners finished the post-course MCQ tests onsite, and control group learners finished them online—both within 60 minutes. All questions were scored as one point per correct response and zero points for an incorrect response. Scores were changed as a percentage of questions answered correctly.

Questionnaire to Evaluate Trainee’s Satisfaction With the Course and Training Methods

An additional 8-item questionnaire was administered to all participants to evaluate their experience with the courses and training methods on a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) after finishing the three course modules (both the theoretical learning and case studies). The questionnaire was piloted with 52 PHCWs and revised accordingly to ensure that the questions could be understood and answered well by all respondents. Cronbach alpha for the questionnaire was .975 according to the pilot study. Subjects who participated in the pilot test were excluded from the final analysis.

Outcomes

The primary outcome was the difference between the control and experimental intervention group in knowledge achievement (measured by baseline and postcourse MCQ tests). The secondary outcome was the difference in trainees’ satisfaction with the courses and training methods between the control and experimental intervention groups (measured by an 8-item evaluation questionnaire)..

Sample Size

The information regarding baseline knowledge, possible gains, and intracluster (intraclass) correlation coefficient was obtained from our pilot study work to calculate the sample size and power calculation. A total of 56 clusters (township centers) are needed to detect a knowledge gain of 5% in the experimental intervention compared with the control intervention using a 2-sided test, an alpha level of 5%, 80% power, assuming a standard deviation of 20, an intracluster (intraclass) correlation coefficient of .06, and expecting a mean cluster size of around 20.

Statistical Analysis

Data was presented as mean with 95% CI. Responses to the baseline and postcourse assessments were scored, and comparisons between the 2 groups were made. The MCQ postscores were compared between the two groups using a

XSL•FO
RenderX

multilevel linear mixed model, with intervention group, time of assessment (baseline or postcourse), and intervention × time interaction as fixed effects and township centers and participants as random effects. For evaluating participants' satisfaction with the training modalities, the responses were computed on a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree). Because very few participants chose scores of 3, 4, or 5, in the analysis, we combined responses with scores of 3, 4, and 5 into a single category "neutral or disagree." Univariate logistic regression analysis was used to calculate the odds ratios (ORs) and 95% CI for comparing the difference between the two groups on each item of the questionnaire.

All comparisons were 2-sided and were considered statistically significant at $P<.05$. On the basis of Cohen guidelines [24], an overall between-group effect size for outcome variable was calculated by dividing the between-group difference by the within-cluster standard deviation, with effect sizes of 0.8 considered large, 0.5 considered medium, and 0.2 considered small. SAS version 9.1 (SAS Institute) was used for all analyses.

Ethics and Consent

This study was approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. Written informed content was obtained from all study subjects before the study.

Results

Participants Characteristics and Study Participation

A total of 1237 PHCWs were recruited (Figure 1); 3 counties (Zhijiang, Xianan, and Huarong) with 29 township centers including 612 participants were assigned to the blended-learning group and 3 counties (Yiling, Chibi, and Liangzihu) with 31 towns including 625 participants were assigned to the pure e-learning group. A total of 43 participants in the blended-learning group and 62 in the pure e-learning group withdrew after the allocation due to refusal to participate or absence from baseline assessment. In total, 105 participants were lost to follow-up in course module 1; 95 in course module 2; 124 in course module 3 in the experimental group; and 87, 84, and 78 participants were lost to follow-up in course module 1, module 2, and module 3, respectively, in the control group.

Table 1 summarizes baseline characteristics of participants. Most participants had a technical secondary education background level or below and majored in western medicine. An analysis of baseline characteristics showed no statistically significant difference between the two groups.

Table 1. Demographic characteristics of the participants.

Variable	Blended-learning group (N=569)	Pure e-learning group (N=563)	DF	χ^2/ t[c]	*P*
Mean age (SD[a]), years	41.67 (11.08)	41.98 (9.58)	1130	0.5	.59
Age category, n (%)					
≤29 years	72 (12.7)	55 (9.8)	3	7.4	.06
30-39 years	184 (32.3)	182 (32.3)			
40-49 years	165 (29.0)	200 (35.5)			
≥50 years	148 (26.0)	126 (22.4)			
Gender, n (%)					
Male	291 (51.1)	320 (56.8)	1	3.7	.06
Female	278 (48.9)	243 (43.2)			
Educational level, n (%)					
Technical secondary school or below[b]	453 (79.6)	435 (77.3)	2	1.1	.57
Junior college	101 (17.8)	109 (19.4)			
Undergraduate or above	15 (2.6)	19 (3.4)			
Major, n (%)					
Western medicine	308 (54.9)	346 (61.5)	4	6.2	.18
Nursing	129 (23.0)	113 (20.1)			
Preventive medicine	52 (9.3)	40 (7.1)			
Traditional Chinese medicine	29 (5.2)	31 (5.5)			
Other	43 (7.7)	33 (5.9)			

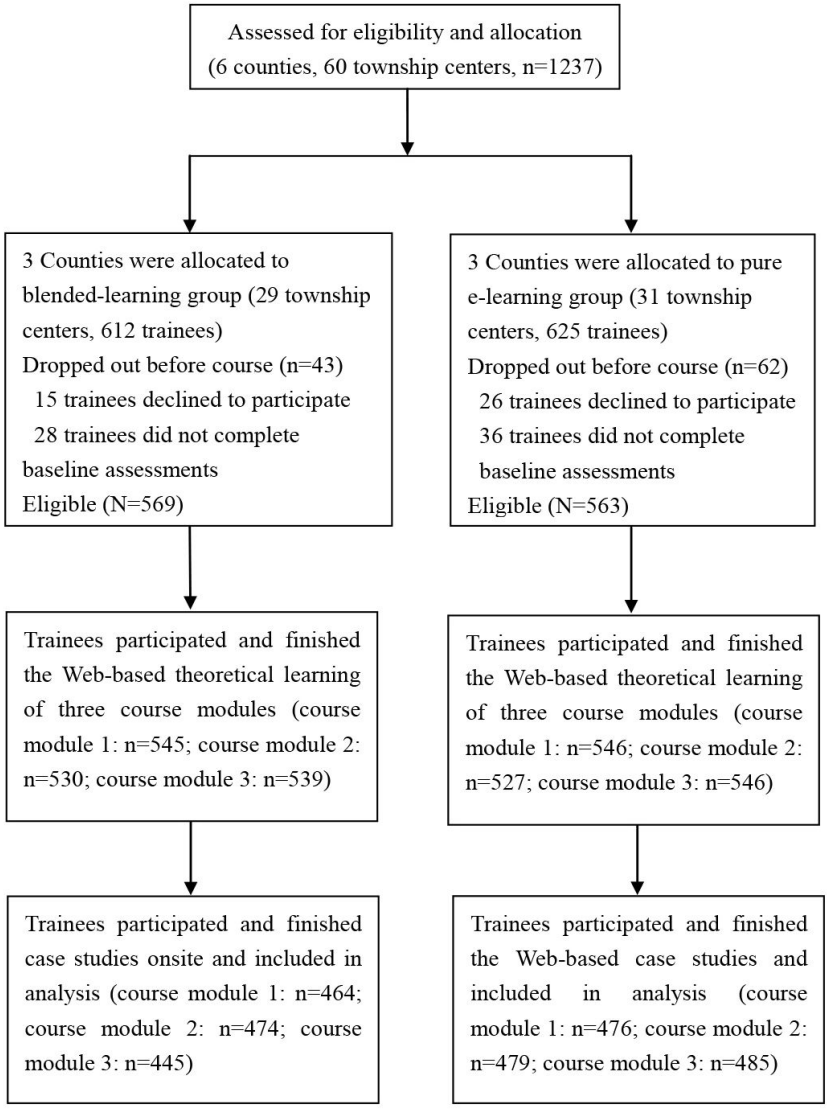
[a]SD: standard deviation.

[b]Technical secondary school or below: illiterate or primary school, middle school, high school, or technical secondary school.

[c] χ^2/ t: t test was used to compare the "mean age (SD)" between two groups with χ^2test comparing the differences between two groups in other variables such as "age category," "gender," "educational level," and "major."

XSL•FO
RenderX

Figure 1. Study flow diagram.



Knowledge Achievement

Baseline knowledge scores of the three course modules between experimental and control group were similar. After the interventions, there were more gains in the experimental group than in the control group: (1) Course module 1: postcourse mean, 93.21 (95% CI 92.49-93.93) in the experimental group versus 88.29 (95% CI 87.19-89.40) in the control group; adjusted mean difference, 4.92 (95% CI 2.61-7.24; *P*<.001). (2) Course module 2: postcourse mean, 94.05 (95% CI 93.37-94.73) in the experimental group versus 90.22 (95% CI 89.12-91.31) in the control group; adjusted mean difference, 3.67 (95% CI 1.17-6.18; *P*=.004). (3) Course module 3: postcourse mean, 93.88 (95% CI 93.08-94.68) in the experimental group versus 89.09 (95% CI 87.89-90.30) in the control group; adjusted mean difference, 4.63 (95% CI 2.12-7.14; *P*<.001). See Table 2. These gains represented moderate effect sizes for knowledge in these course modules (0.40, 0.34, and 0.40, respectively).

XSL•FO
RenderX

Table 2. Changes in knowledge using scores obtained with multiple-choice questions between blended-learning group and pure e-learning group.

Knowledge MCQ[a] scores (%), mean (95% CI)	Blended-learning group N=464 for course module 1; N=474 for course module 2; N=445 for course module 3		Pure e-learning group N=476 for course module 1; N=479 for course module 2; N=485 for course module 3		Comparisons between two groups	
	Baseline	Postcourse	Baseline	Postcourse	Adjusted difference[b], Mean (95% CI)	*P* value
Course module 1[c]	69.69 (68.10-71.27)	93.21 (92.49-93.93)	69.63 (68.16-71.1)	88.29 (87.19-89.40)	4.92 (2.61-7.24)	<.001
Course module 2[d]	71.20 (69.75-72.65)	94.05 (93.37-94.73)	72.71 (71.38-74.05)	90.22 (89.12-91.31)	3.67 (1.17-6.18)	.004
Course module 3[e]	74.12 (72.45-75.79)	93.88 (93.08-94.68)	73.85 (72.37-75.34)	89.09 (87.89-90.30)	4.63 (2.12-7.14)	<.001

[a]MCQ: multiple-choice questions.

[b]Adjusted difference is the mean difference between groups (intervention-control) adjusted for time of assessment and intervention × time interaction in a multilevel model with township center and participants as a random effect.

[c]Course module 1: health management of patients with hypertension.

[d]Course module 2: health records management for residents.

[e]Course module 3: vaccination.

Trainee's Satisfaction With the Interventions Methods

A questionnaire response rate of 71.9% (409/569) was achieved in the blended-learning group compared with 80.3% (452/563) in the pure e-learning group. Trainees' subjective opinions toward the interventions were investigated, including training benefits (confidence increase, aim realization, and knowledge improvement), changes in learning interest, and satisfaction with the training mode and the interaction. A majority of PHCWs agreed that the contents were well relevant to their work (93.9% in experimental group vs 94.5% in control group, *P*=.70) and that they would like to try the training mode again (92.4% in experimental group vs 90.8% in control group, *P*=.37; Table 3). The blended-learning trainee was found to be more in agreement than the pure e-learning trainee due to the following four issues: (1) "Participation in the training had increased my interest in learning" OR 1.85 (95% CI 1.22-2.80; *P*=.003); (2) "Participation in the training increased the interaction with others" OR 1.77 (95% CI 1.20-2.60; *P*=.004); (3) "Overall, I was satisfied with learning experience" OR 1.78 (95% CI 1.11-2.88; *P*=.02); and (4) "I achieved the objectives of each course" OR 1.63 (95% CI 1.08-2.48; *P*=.02). Concerning other questions in the evaluation questionnaire, there were no significant differences found between the experimental and control groups (Table 3).

XSL•FO
RenderX

Table 3. Questionnaire evaluation of the training between the blended-learning and pure e-learning group.

Courses evaluation questions	Blended-learning group (N=409) n (%)[a]					Pure e-learning group (N=452) n (%)[a]					OR[b] (95% CI)[c]	*P* value
	1	2	3	4	5	1	2	3	4	5		
1. The courses are relevant to the daily work.	157 (38.7)	224 (55.2)	15 (3.7)	5 (1.2)	5 (1.2)	169 (37.4)	258 (57.1)	24 (5.3)	1 (0.2)	0 (0)	0.89 (0.50-1.58)	.70
2. I achieved the objectives of each course.	141 (34.5)	228 (55.8)	30 (7.3)	9 (2.2)	1 (0.3)	139 (30.8)	245 (54.2)	64 (14.2)	4 (0.9)	0 (0)	1.63 (1.08-2.48)	.02
3. Participation in the training had increased my interest in learning.	153 (37.5)	216 (53.0)	32 (7.8)	5 (1.2)	3 (0.5)	147 (32.5)	231 (51.1)	61 (13.5)	12 (2.7)	1 (0.2)	1.85 (1.22-2.80)	.003
4. Participation in the training had increased my confidence.	150 (36.9)	211 (51.8)	31 (7.6)	14 (3.4)	1 (0.3)	159 (35.2)	230 (50.8)	43 (9.5)	19 (4.2)	1 (0.2)	1.27 (0.85-1.91)	.25
5. Participation in this training had improved my knowledge.	183 (45.5)	199 (49.5)	12 (3.0)	8 (2.0)	0 (0)	191 (42.3)	243 (53.8)	16 (3.5)	2 (0.4)	0 (0)	0.79 (0.42-1.52)	.48
6. Participation in the training increased the interaction with others.	155 (38.1)	206 (50.6)	35 (8.6)	11 (2.7)	0 (0)	144 (31.9)	226 (49.8)	70 (15.7)	12 (2.6)	0 (0)	1.77 (1.20-2.60)	.004
7. I would like to try the training mode again.	190 (46.6)	187 (45.8)	20 (4.9)	4 (1.0)	7 (1.7)	205 (45.4)	205 (45.4)	36 (8.0)	5 (1.1)	1 (0.2)	1.25 (0.77-2.02)	.37
8. Overall, I was satisfied with the training experience.	195 (48.3)	181 (44.8)	23 (5.7)	2 (0.5)	3 (0.7)	136 (30.1)	263 (58.2)	49 (10.8)	4 (0.9)	0 (0)	1.78 (1.11-2.88)	.02

[a]Responses to questions about the feedback on Web-based training platform were on a 5-point Likert scale, ranging from 1 (strongly agree) to 5 (strongly disagree).

[b]OR: odds ratio.

[c]Univariate logistic regression analysis was used to compare the differences between two groups (dependent variable as two categories with combining scores 1, 2 into one category and scores 3, 4, 5 to another category).

Discussion

Principal Findings

This study suggested that in rural China, a blended approach to BPHS training was more effective in improving knowledge than a pure e-learning approach. Trainees in blended-learning group expressed a higher satisfaction level about their learning experiences than pure e-learning trainees. Our study demonstrates the feasibility of applying Internet-related technology to PHCWs' training on BPHS and explores the various training modes to improve the knowledge of PHCWs in rural China.

Currently, the inequalities in health care provision between urban and rural areas and the inequalities in the distribution of health workers remain serious problems in China [25-27]. Rural areas have both lower densities of health workers and less-educated workforce [6]. Achieving the equitable BPHS for all residents requires that every Chinese family in rural and urban areas has access to an appropriately trained and supported health worker. Our previous qualitative study showed that most PHCWs had insufficient knowledge on BPHS but had a positive attitude toward Web-based training approaches [9]. At present, e-learning has become an increasingly popular means to promote learning among health workers using online communications [15]. Blended learning, the combination of e-learning, and face-to-face instructor training, has also been presented as a promising approach for health education [10]. The differences between two novel methods include the different communication scenarios and perceived costs, with face-to-face scenarios having higher learners' costs [28]. In this study, we discussed the comparison results of blended and pure e-learning methods, focusing on two aspects: knowledge achievement and satisfaction level.

Our study suggests that the blended-learning approach is more effective than pure e-learning in terms of knowledge achievement. This is supported by a recent meta-analysis of 56 studies finding that blended learning appears to be more effective than or at least as effective as e-learning [20]. Our findings are consistent with the previous research which showed that the combination of computer-assisted instruction and traditional classroom lecture yielded a significantly greater improvement in knowledge achievement of nursing students than when either strategy is used alone in the context of congenital heart disease [16]. Similarly, Llambí et al [19] reported that Uruguayan physicians who completed a blended-learning course on tobacco cessation achieved better test scores than those who attended pure online course. Furthermore, our quantitative results in this study also showed that blended learners expressed more positive ratings about goal achievement than online learners. A possible explanation may be that blended-learning approaches allow PHCWs to have face-to-face interactions and discussions within groups. A study conducted among pharmacy students emphasized the significance of face-to-face interactions in the blended-learning approach, which were more highly rated than online interactions

XSL•FO
RenderX

[29]. Lack of face-to-face interaction was reported to be a challenge addressed in e-learning programs [30,31]. Previous studies suggested that lack of face-to-face interaction in the pure e-learning may contribute to professional isolation, a decrease in learning experience quality, and unsatisfactory learning outcomes [14,32,33]. According to constructivist learning theory, learning is a social activity, which is intimately associated with the connection with other human beings, teachers, peers, and so on [34]. The theory proposed that learners who have recognized the social aspect of learning and enhanced their interaction with others are more likely to have successful learning experiences [34]. This is further supported by our results showing that blended learners had increased interaction with others via the training than e-learners. Similarly, other studies indicated that blended-learning learners are less likely to experience feelings of isolation or reduced learning interest when compared with e-learners [20,35-37]. Consistent with the previous studies, blended learning achieved a greater learning interest in our study. The richness of blended experiences, including two forms of learning methods and allowing learners to have the face-to-face association and interaction with peers, might also promote learners' learning interest.

In our study, we found that blended-learning trainees had a higher satisfaction level about their learning experiences than pure e-learning trainees. As for the case-based problem solving courses, social and collaborative learning experiences are important to help individuals in thinking, learning, and finding a solution for problems [34]. So and Brush [38] indicated that learner perceptions of collaborative learning were related to learning satisfaction, and learners with higher perceived levels of collaborative learning tended to be more satisfied with blended courses. Although participants in the pure e-learning group could communicate with others in the discussion forums on the BPHS Web-based platform, the asynchronous communication might not make trainees feel part of a learning community. Another study revealed that the learners in the online learning group claimed less learning support and more workload than learners in the blended-learning group with the explanation that learners in the Web-based learning group might lack a sense of presence or belonging [39]. Blended learning with various instructional methods, such as the mix of the face-to-face form of classroom training and Web-based technology, was the major factor in enhancing learner satisfaction [39,40].

Consideration of learning outcome alongside the devoted costs and resources was important for educators to effectively review the educational interventions [41]. Commonly, there are five basic cost-driving categories related to both blended-learning and pure e-learning approaches: labor costs, content development and acquisition, technology and infrastructure, operations costs, and learner-support services [42]. It was reported that developing a 100% online, media-rich, self-paced Web-based content was expensive and required multiple resources and skills [43]. Meanwhile, previous studies suggested that blended learning may potentially balance out and optimize the training program development and deployment cost and time by combining different delivery modes [43,44]. However, another study pointed out that not all blended learning would be cost-effective, and that the design of learning models around staff time was the determinant [45]. In addition, the resource support in the blended learning are involved in making different forms of resources (offline and online) available for learners as well as organizing them [43]. Thus, the cost-effective analysis of the blended versus pure e-learning approaches is necessary for educators to develop a more cost-effective mode, and we suggest that it needs to be conducted in the future.

Although blended learning shows positive learning outcomes and satisfaction level in the study, barriers to the implementation among PHCWs still exist. Possible barriers to blended learning include technical difficulties, such as interrupted or limited Internet connection, poor computer literacy, and hindrance in accessing learning resource material, as reported previously [11,46]. The lack of time for PHCWs to take part due to service load was another barrier because most rural PHCWs are responsible for delivering both BPHS and medical service to residents [5]. The barriers mentioned above could explain most dropouts in the blended-learning group in our study. Making suitable arrangement between work and training is critical for PHCWs to complete the training courses. On the other side, selecting the right blend between face-to-face and online learning is also important for the successful implementation of blended courses [47], which should take into account the job characteristics of health care workers.

Strengths and Limitations

To our knowledge, this is the first study to compare the effectiveness of a blended-learning approach with a pure e-learning approach to BPHS training among rural PHCWs. The main strengths of this study include the relatively large sample size, and both subjective and objective evaluation methods applied for comparison. Furthermore, our study provided more evidence on the effects of blended learning in comparison with pure e-learning.

The study has five limitations. First, the dropouts in both groups were seen in our study. The dropout rates were similar in the blended group and pure e-learning group in course module 1 (18.5% vs 15.5%) and course module 2 (16.7% vs 14.9%). As for course module 3, we should caution that the dropout rate was differentially higher in the blended group (21.8%) than that in the pure e-learning group (13.9%). However, the background characteristics of participants who drop out in two groups in the course module 3 were similar. As well, there were no significant differences in the comparison of background characteristics between dropouts and non-dropouts. In addition, among dropouts who had completed the baseline assessment of course module 3, there was no baseline difference between those who completed (mean score 73.98 [SD 17.26]; n=930) and dropouts (mean score 72.29 [SD 23.75]; n=155) with difference, 1.69 (95% CI −1.43 to 4.81; P=.40). As there are some dropouts who had not completed the baseline assessment and their willingness had not been investigated, we still should caution the potential selective bias. Second, all subjects included in this study were from Central China, which limits its generalizability to other areas. Third, we compared the knowledge achievement between two training modes, but the effects of the courses on behavioral change or long-term

XSL•FO
RenderX

educational outcomes were not evaluated and compared in this study. Fourth, the same knowledge questionnaires for pre- and posttest were used in the study, which might have introduced a subject sensitization bias. Fifth, the economic evaluation of two different training forms was not carried out in our study. To develop a more cost-effective training mode, the economic evaluation should be performed in a future study.

Conclusions

In conclusion, blended approaches to BPHS training resulted in a better knowledge achievement and a higher satisfaction level than pure e-learning approaches among PHCWs in rural China. Using more effective training modes to improve PHCWs' knowledge on BPHS can help enhance the PHCWs' competency and accordingly improve the quality of health care in rural China in order to achieve health equity. To provide more rigorous evidence on the effects of blended learning in comparison with pure e-learning, more research is needed in the future.

Acknowledgments

The authors thank all of the participants for their contributions to this study. This study was supported by the project "Strengthening Primary Health Care Workers' Competence by Using an Internet-Based Interactive Platform in Rural China" funded by the Ministry of Science and Technology of China. The funder had no involvement in the design, analysis, or reporting of the study.

Authors' Contributions

WRY conceptualized and designed the study. XXZ, ZXZ, FS, and QL performed the study. XXZ and ZXZ analyzed the data. XXZ prepared the first draft of the paper. All authors contributed to the revision of the manuscript.

Conflicts of Interest

None declared.

References

1. Meng Q, Yuan J, Jing L, Zhang J. Mobility of primary health care workers in China. Hum Resour Health 2009;7:24 [FREE Full text] [doi: 10.1186/1478-4491-7-24] [Medline: 19292911]
2. NHFPC. 2016. Illustration: Understanding national basic public health services URL: http://www.nhfpc.gov.cn/jws/s3578/201504/49bf80b308fb47dbb138f86b1a49a2fc.shtml [accessed 2016-08-01] [WebCite Cache ID 6jQraoP6I]
3. Hongoro C, McPake B. How to bridge the gap in human resources for health. Lancet 2004;364(9443):1451-1456. [doi: 10.1016/S0140-6736(04)17229-2] [Medline: 15488222]
4. Zhao Y, Cui S, Yang J, Wang W, Guo A, Liu Y, et al. Basic public health services delivered in an urban community: a qualitative study. Public Health 2011 Jan;125(1):37-45. [doi: 10.1016/j.puhe.2010.09.003] [Medline: 21145087]
5. Yip WC, Hsiao WC, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. Lancet 2012 Mar 3;379(9818):833-842. [doi: 10.1016/S0140-6736(11)61880-1] [Medline: 22386036]
6. Strasser R, Kam SM, Regalado SM. Rural health care access and policy in developing countries. Annu Rev Public Health 2016;37:395-412. [doi: 10.1146/annurev-publhealth-032315-021507] [Medline: 26735432]
7. Li X, Liu J, Huang J, Qian Y, Che L. An analysis of the current educational status and future training needs of China's rural doctors in 2011. Postgrad Med J 2013 Apr;89(1050):202-208. [doi: 10.1136/postgradmedj-2012-131094] [Medline: 23389284]
8. Khatony A, Nayery ND, Ahmadi F, Haghani H, Vehvilainen-Julkunen K. The effectiveness of web-based and face-to-face continuing education methods on nurses' knowledge about AIDS: a comparative study. BMC Med Educ 2009;9:41 [FREE Full text] [doi: 10.1186/1472-6920-9-41] [Medline: 19591678]
9. Zhang Z, Zhan X, Li Y, Hu R, Yan W. Web-based training for primary healthcare workers in rural China: a qualitative exploration of stakeholders' perceptions. PLoS One 2015;10(5):e0125975 [FREE Full text] [doi: 10.1371/journal.pone.0125975] [Medline: 25961727]
10. Bollinger R, Chang L, Jafari R, O'Callaghan T, Ngatia P, Settle D, et al. Leveraging information technology to bridge the health workforce gap. Bull World Health Organ 2013 Nov 1;91(11):890-892 [FREE Full text] [doi: 10.2471/BLT.13.118737] [Medline: 24347719]
11. Kulier R, Gülmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, et al. Effectiveness of a clinically integrated e-learning course in evidence-based medicine for reproductive health training: a randomized trial. JAMA 2012 Dec 5;308(21):2218-2225. [doi: 10.1001/jama.2012.33640] [Medline: 23212499]
12. Wallen GR, Cusack G, Parada S, Miller-Davis C, Cartledge T, Yates J. Evaluating a hybrid web-based basic genetics course for health professionals. Nurse Educ Today 2011 Aug;31(6):638-642 [FREE Full text] [doi: 10.1016/j.nedt.2010.11.001] [Medline: 21106279]

XSL•FO
RenderX

13. Hauge LS, Frischknecht AC, Gauger PG, Hirshfield LE, Harkins D, Butz DA, et al. Web-based curriculum improves residents' knowledge of health care business. J Am Coll Surg 2010 Dec;211(6):777-783. [doi: 10.1016/j.jamcollsurg.2010.07.011] [Medline: 20851642]
14. Gaikwad N, Tankhiwale S. Interactive E-learning module in pharmacology: a pilot project at a rural medical college in India. Perspect Med Educ 2014 Jan;3(1):15-30 [FREE Full text] [doi: 10.1007/s40037-013-0081-0] [Medline: 24072666]
15. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
16. Rouse DP. The effectiveness of computer-assisted instruction in teaching nursing students about congenital heart disease. Comput Nurs 2000;18(6):282-287. [Medline: 11105402]
17. Howerton WB, Enrique PR, Ludlow JB, Tyndall DA. Interactive computer-assisted instruction vs. lecture format in dental education. J Dent Hyg 2004;78(4):10. [Medline: 16197750]
18. Shomaker TS, Ricks DJ, Hale DC. A prospective, randomized controlled study of computer-assisted learning in parasitology. Acad Med 2002 May;77(5):446-449. [Medline: 12010707]
19. Llambí L, Esteves E, Martinez E, Forster T, García S, Miranda N, et al. Teaching tobacco cessation skills to Uruguayan physicians using information and communication technologies. J Contin Educ Health Prof 2011;31(1):43-48. [doi: 10.1002/chp.20100] [Medline: 21425359]
20. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. J Med Internet Res 2016;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]
21. NHFPC. Notification on National Basic Public Health Service (2011 edition) URL: http://www.nhfpc.gov.cn/zhuzhan/wsbmgz/201304/cb5978bb42814451a26e5c97dd855254.shtml [WebCite Cache ID 6jQspPXUJ]
22. TJMU. The public health services online training platform based on Moodle URL: http://elearning.tjmu.edu.cn/gwfw/ [accessed 2016-10-02] [WebCite Cache ID 6kwmFAlPs]
23. Moodle. Using forum URL: https://docs.moodle.org/29/en/Using_Forum [accessed 2016-10-01] [WebCite Cache ID 6kvxljtVc]
24. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: L. Erlbaum Associates; 1988.
25. Wang T, Zeng R. Addressing inequalities in China's health service. Lancet 2015 Oct 10;386(10002):1441. [doi: 10.1016/S0140-6736(15)00402-X] [Medline: 26466037]
26. Zhou K, Zhang X, Ding Y, Wang D, Lu Z, Yu M. Inequality trends of health workforce in different stages of medical system reform (1985-2011) in China. Hum Resour Health 2015;13:94 [FREE Full text] [doi: 10.1186/s12960-015-0089-0] [Medline: 26645960]
27. Anand S, Fan VY, Zhang J, Zhang L, Ke Y, Dong Z, et al. China's human resources for health: quantity, quality, and distribution. Lancet 2008 Nov 15;372(9651):1774-1781. [doi: 10.1016/S0140-6736(08)61363-X] [Medline: 18930528]
28. Kerres M, Witt CD. A didactical framework for the design of blended learning arrangements. Journal of Educational Media 2003 Oct;28(2-3):101-113. [doi: 10.1080/1358165032000165653]
29. Edginton A, Holbrook J. A blended learning approach to teaching basic pharmacokinetics and the significance of face-to-face interaction. Am J Pharm Educ 2010 Jun 15;74(5):88 [FREE Full text] [Medline: 20798797]
30. Frehywot S, Vovides Y, Talib Z, Mikhail N, Ross H, Wohltjen H, et al. E-learning in medical education in resource constrained low- and middle-income countries. Hum Resour Health 2013;11:4 [FREE Full text] [doi: 10.1186/1478-4491-11-4] [Medline: 23379467]
31. Ballew P, Castro S, Claus J, Kittur N, Brennan L, Brownson RC. Developing web-based training for public health practitioners: what can we learn from a review of five disciplines? Health Educ Res 2013 Apr;28(2):276-287 [FREE Full text] [doi: 10.1093/her/cys098] [Medline: 22987862]
32. Halstead JA, Coudret NA. Implementing Web-based instruction in a school of nursing: implications for faculty and students. J Prof Nurs 2000;16(5):273-281. [doi: 10.1053/jpnu.2000.9459] [Medline: 11033937]
33. Kheng S. The challenges of upgrading from ISPO Category II level to Bachelor Degree level by distance education. Prosthet Orthot Int 2008 Sep;32(3):299-312. [doi: 10.1080/03093640802109764] [Medline: 18720252]
34. Hein G. Constructivist learning theory. 1991 Presented at: International Committee of Museum Educators Conference. Proceedings of Developing Museum Exhibitions for Lifelong Learning Conference, ICOM/CECA; October 15-22, 1991; Jerusalem, Israel URL: http://beta.edtechpolicy.org/AAASGW/Session2/const_inquiry_paper.pdf
35. Wu J, Tennyson RD, Hsia T. A study of student satisfaction in a blended e-learning system environment. Comput Educ 2010 Aug;55(1):155-164. [doi: 10.1016/j.compedu.2009.12.012]
36. Wu J, Tennyson RD, Hsia T, Liao Y. Analysis of E-learning innovation and core capability using a hypercube model. Comput Human Behav 2008 Sep;24(5):1851-1866. [doi: 10.1016/j.chb.2008.02.008]
37. Maki RH, Maki WS, Patterson M, Whittaker PD. Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. Behav Res Methods Instrum Comput 2000 May;32(2):230-239. [Medline: 10875167]
38. So H, Brush TA. Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. Comput Educ 2008 Aug;51(1):318-336. [doi: 10.1016/j.compedu.2007.05.009]

XSL•FO
RenderX

39. Lim D, Morris M, Kupritz V. Online vs. blended learning: differences in instructional outcomes and learner satisfaction. JALN 2007;11(2):27-42 [FREE Full text]
40. Thorne K. Blended Learning: How to Integrate Online & Traditional Learning. London: Kogan Page; 2003.
41. Walsh K, Reeves S, Maloney S. Exploring issues of cost and value in professional and interprofessional education. J Interprof Care 2014 Nov;28(6):493-494. [doi: 10.3109/13561820.2014.941212] [Medline: 25051088]
42. Battaglino T, Haldeman M, Laurans E. Eric. 2012. Creating sound policy for digital learning: a working paper series from the Thomas B. Fordham Institute URL: http://files.eric.ed.gov/fulltext/ED528948.pdf [accessed 2016-10-01] [WebCite Cache ID 6kvjPZzvN]
43. Singh H. Building effective blended learning programs. Educational Technology 2003;43(6):51-54 [FREE Full text]
44. Halan D. Blended learning. J Educ Technol 2005;2(1):20-24 [FREE Full text]
45. Maloney S, Nicklen P, Rivers G, Foo J, Ooi YY, Reeves S, et al. A cost-effectiveness analysis of blended versus face-to-face delivery of evidence-based medicine to medical students. J Med Internet Res 2015;17(7):e182 [FREE Full text] [doi: 10.2196/jmir.4346] [Medline: 26197801]
46. Flys T, González R, Sued O, Suarez CJ, Kestler E, Sosa N, et al. A novel educational strategy targeting health care workers in underserved communities in Central America to integrate HIV into primary medical care. PLoS One 2012;7(10):e46426 [FREE Full text] [doi: 10.1371/journal.pone.0046426] [Medline: 23115628]
47. Maarop AH, Embi MA. Implementation of blended learning in higher learning institutions: a review of literature. IES 2016 Feb;9(3):41. [doi: 10.5539/ies.v9n3p41]

Abbreviations

BPHS: basic public health services
CDC: Center for Disease Control and Prevention
e-learning: electronic learning
GDP: gross domestic product
MCQ: multiple-choice questions
OR: odds ratio
PHCWs: primary health care workers
TB: tuberculosis

Edited by G Eysenbach; submitted 07.08.16; peer-reviewed by J Kaewkungwal, C Nguyen; comments to author 25.08.16; revised version received 02.10.16; accepted 02.03.17; published 01.05.17.

Please cite as:
Zhan X, Zhang Z, Sun F, Liu Q, Peng W, Zhang H, Yan W
Effects of Improving Primary Health Care Workers' Knowledge About Public Health Services in Rural China: A Comparative Study of Blended Learning and Pure E-Learning
J Med Internet Res 2017;19(5):e116
URL: http://www.jmir.org/2017/5/e116/
doi:10.2196/jmir.6453
PMID:28461286

©Xingxin Zhan, Zhixia Zhang, Fang Sun, Qian Liu, Weijun Peng, Heng Zhang, Weirong Yan. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Using Child Personas in the Development of a Digital Peer Support Service for Childhood Cancer Survivors

Pontus Wärnestål[1], PhD; Petra Svedberg[2], RN, PhD; Susanne Lindberg[1], MSc; Jens M Nygren[2], PhD

[1]School of Information Technology, Halmstad University, Halmstad, Sweden

[2]School of Health and Welfare, Halmstad University, Halmstad, Sweden

Corresponding Author:
Jens M Nygren, PhD
School of Health and Welfare
Halmstad University
Box 823
Halmstad, 30118
Sweden
Phone: 46 35167863
Fax: 46 35167100
Email: jens.nygren@hh.se

Abstract

Background: Peer support services have the potential to support children who survive cancer by handling the physical, mental, and social challenges associated with survival and return to everyday life. Involving the children themselves in the design process allows for adapting services to authentic user behaviors and goals. As there are several challenges that put critical requirements on a user-centered design process, we developed a design method based on personas adapted to the particular needs of children that promotes health and handles a sensitive design context.

Objective: The purpose of this study was to evaluate the effects of using child personas in the development of a digital peer support service for childhood cancer survivors.

Methods: The user group's needs and behaviors were characterized based on cohort data and literature, focus group interviews with childhood cancer survivors (n=15, 8-12 years), stakeholder interviews with health care professionals and parents (n=13), user interviews, and observations. Data were interpreted and explained together with childhood cancer survivors (n=5) in three explorative design workshops and a validation workshop with children (n=7).

Results: We present findings and insights on how to codesign child personas in the context of developing digital peer support services with childhood cancer survivors. The work resulted in three primary personas that model the behaviors, attitudes, and goals of three user archetypes tailored for developing health-promoting services in this particular use context. Additionally, we also report on the effects of using these personas in the design of a digital peer support service called Give Me a Break.

Conclusions: By applying our progressive steps of data collection and analysis, we arrive at authentic child-personas that were successfully used to design and develop health-promoting services for children in vulnerable life stages. The child-personas serve as effective collaboration and communication aids for both internal and external purposes.

(J Med Internet Res 2017;19(5):e161) doi:10.2196/jmir.7175

KEYWORDS

peer; childhood; cancer; survivor; participation; user experience; service design

Introduction

Peer Support for Childhood Cancer Survivors

Advances in diagnosis, risk stratification, and treatment protocols have resulted in that most children who are diagnosed with cancer survive the disease and have the potential to live a long life with a quality of life comparable with their peers. However, the transition from a period of intensive treatment to everyday life is often associated with physical and psychological problems for which they need professional support [1-3]. Managing these physical and psychological problems as well as the social challenges that are associated with the experiences and consequences of the disease can be facilitated by social support mediated by peers on whom the child can rely and who

XSL•FO
RenderX

share a similar background [4]. The availability of such peer support is limited and resources are often offered sporadically and by the initiative of patient organizations or health care services [5]. This shortage in resources that fulfills the needs and requirements of the target group demonstrates the need to develop peer support services that are adapted to the preferences and requirements of child users and that are not restricted by clinical, geographical, age-related, and logistic limitations of face-to-face interaction associated with this user group [6].

Designing With and for Children in a Sensitive Context

Developing digital peer support (DPS) services directed toward children surviving from cancer to facilitate health-promoting social connectedness to other children with similar experiences has to involve the children themselves in the design process to allow for inclusion of their perspectives on challenges of survival and on how to integrate such services into their everyday life [7,8]. A user-centered design (UCD) process enables designers to meet the children's worldview, their cognitive and emotional developmental stage, age, and gender [9], as well as their requirements on usability and experiential quality [6]. Children can be involved through a variety of UCD techniques and methods such as cooperative and participatory design [10,11], contextual design [12], and persona-based design [13-15]. The involvement of children can be achieved through participation in different stages of the design process and for different purposes such as, defining the needs within the user group in a particular area, formulating the aim of the design, planning and setting up methodologies, assembling and analyzing data, sketching and prototyping, and planning and implementing dissemination of findings [16].

Involvement of children who have survived cancer in the context of health-promoting and social digital services is associated with several challenges, which, if unaddressed, might have a negative impact on their chances of getting an active role in a user-centered design process. First, although it is acknowledged that children have the right to be heard in matters that affect their lives, health, and care, they are often viewed as "vulnerable" subjects due to their dependence on adults. It is therefore essential to protect children from harm associated with involvement and therefore carefully handle consent to participation, confidentiality, research context, and activities [17]. Second, children who have experiences from a severe disease are seen as even more vulnerable in relation to participation and it is common that various gatekeepers restrict their participation, such as ethical boards, health care professionals, parents [18], and ultimately, the children themselves and their willingness to participate in processes developed by researchers or designers [17]. Third, the specific characteristics of this user group (eg, age-span, medical history, geographic spread, and clinical restrictions) make it difficult to recruit, meet, and engage the children on a regular basis for long-term participatory design work. UCD techniques and methods that take these considerations into account could be powerful tools for user involvement in the design context outlined previously. To address these problems, we developed a design process based on "personas" adapted to the particular requirements of addressing the needs of children, promoting health, and handling a sensitive context.

A Persona-Based Design Process for Health Care Contexts

Persona-based design is a UCD approach that provides a vivid representation of the target users, makes them concrete and life-like, and not merely described by demographic statistics. Personas are defined as "fictitious, specific, concrete representations of target users" [19]; however even though they are fictitious, personas are created directly from research data collected using both qualitative and quantitative research methods. The use of personas as a design method has been applied in different ways. One of the most well-known in the industry is goal-directed design (GDD) [13,14]. GDD has been used within health care research projects as well [20].

An essential benefit of using personas is that they have the potential to build empathy for the target users [21]. In the context of designing interactive systems, empathy is understood as an understanding of—and identification with—the user to ensure that they will be able to take advantage of the service being designed and will be able to use it with pleasure rather than frustration. Another benefit often voiced regarding persona-based design is the persona's effectiveness in conveying design requirements to various stakeholders [22,23]. However, as we predicted that several of the stakeholders would act as gatekeepers and—based on a protective stance—would limit access to children with experience from cancer treatment in the development of a DPS service, we applied an approach toward users' participation based on differences between a "salutogenic" orientation and a "pathogenic" orientation in relation to health. The original idea of salutogenesis states that it is more important to focus on peoples' resources and capacities to create health rather than focusing on risks, ill health, and disease (pathogenesis) [24]. The concept of salutogenesis has been developed toward, first, problem solving and second, capacity to use resources available [25]. In line with this we used the salutogenic orientation to focus the development of a DPS service on factors that maintain and promote health and well-being of childhood cancer survivors and the pathogenic orientation to focus on factors that are related to disease and treatment in relation to DPS services. The salutogenic orientation was thus used to incorporate aspects that are inherent to the users and their motivations, behaviors, and goals related to everyday life at home, with friends and family, and in school, whereas the pathogenic orientation was used to incorporate stakeholders' views on the challenges and limitations related to disease and treatment that a DPS service has to deal with.

Study Aim

Along these lines, our design approach takes a user-centered, cocreative perspective, where the use of child-personas plays an important role. The design context for children who have survived cancer is of a sensitive nature, where ethical considerations need to be addressed regarding what topics can be handled during interviews and workshops to ensure a healthy environment for all participants [26]. This puts additional requirements on the design methodology, as the restrictions inherent in this sensitive design context limit the ways user data can be collected compared with traditional approaches. The purpose of this study was to develop high-quality, authentic

XSL•FO
RenderX

personas for DPS service design for children who have survived cancer. This paper describes and discusses a method for coconstruction of personas with children in sensitive design contexts and the effects of using these personas in the development of a DPS service for childhood cancer survivors.

Methods

Study Design

Due to ethical reasons, a salutogenic perspective was prominent throughout the activities performed with children. Stakeholder interviews with health care professionals and parents of children with a history of cancer provided a pathogenic perspective. Personas were later created to merge these perspectives. The study was approved by the regional ethical board (dnr 2011/307).

Participants

For the initial semistructured focus group interviews, children between the ages of 8-12 that were currently under maintenance treatment or in remission from cancer were recruited (Table 1). Children that were still in an acute state of cancer treatment and children who had only received radiation therapy or surgery were excluded from the study. The children were recruited through two Swedish hospitals. Nurses selected 10 girls and 19 boys that fit the inclusion criteria, of whom 5 girls and 10 boys consented to participate. Both the children and their parents gave informed consent.

Recruitment for the stakeholder interviews was based on an organizational perspective with the identification of important professional roles (eg, medical specialist, clinic leader, consultancy nurses; Table 1). From the initial interviews, additional important roles were identified. In total, 9 health care professionals were interviewed: 1 medical specialist or doctor, 1 clinic leader (doctor), 1 oncology nurse, 2 consultancy nurses, 2 play or occupational therapists, 1 sibling supporter (nurse), and 1 representative from a local childhood cancer patient association. Additionally, 4 parents of children between the ages of 9-11 years with a history of cancer were interviewed.

From the 15 children who had participated in the initial focus group interviews, 5 were chosen for further participation in a series of 3 explorative design workshops (Table 1). These children (3 boys aged 11-12 years and 2 girls aged 11 and 13 years) were selected as they represented particularly well-functioning groups and had shown a high level of creativity during the initial focus group interviews. For the final workshop, 3 girls and 4 boys without experience from cancer treatment aged 10-12 years were recruited from a local school (Table 1).

Focus Group Interviews

The focus group interviews aimed at researching how children with a history of cancer treatment experienced friendship [27] (Table 1). The focus of the interviews was how the children interacted socially with their friends and what their everyday lives looked like. The focus groups (n=5) each comprised 3 children of similar age and same gender and were performed in a location chosen together with the children's families. In this context, small homogeneous groups are recommended to ensure that children's needs are met [18,28]. Due to geographical reasons, one of the focus group interviews was performed with groups of children with mixed genders.

The focus group interviews, each approximately 90 min, were recorded using video and sound. The interviews were semistructured and separated into three stages. First, a section meant to create familiarity within the group and familiarize the children with each other [29]. Second, a section where the children used creative and informative techniques to articulate opinions and experiences on relevant topics. Activities included, for example, brainstorming, drawing and telling, listing, and answering activity-oriented questions—techniques used to help children in expressing themselves [30,31]. Finally, a third closing section that summarized the focus group interview.

Stakeholder Interviews

Considering input from subject-matter experts and important business roles is a vital and common precursor to successful design. These interviews are often performed before user involvement. Also, for this study we were aware of the limitation of not being able to broach the pathogenic perspective with the children, and thus further importance was placed on these interviews. The semistructured stakeholder interviews (n=13) were driven by the children's views gained from the focus group interviews and gave insight into the relevant pathogenic aspects related to the disease, treatment, and later transition to everyday life, and spoke of for example long-term treatment strategies, challenges, and the possible effects of introducing DPS into the health care process [5] (Table 1).

Design Workshops

After the stakeholder interviews, we involved 5 children who had previously participated in the focus groups in 3 explorative and generative design workshops. Each workshop was run twice, once with the 3 boys and once with the 2 girls resulting in 6 workshops in total. Around 3-4 researchers participated in each workshop, where one child and one adult always cooperated, and the remaining researcher functioned as facilitator. Each workshop lasted 3 hours including a meal break (Table 1).

Comics were used as a theme throughout all workshops. Having a familiar theme recur throughout all workshops can serve to make the design process easier for the participants to understand [32]. Each workshop had a different focus: (1) building familiarity and creating proxy personas, (2) creating redemption scenarios, and (3) feedback and prototyping. After both rounds of workshops had been performed, the 7 children who had participated took part in a summary session where they gave feedback on the outcomes and early prototypes. This also served to inform the children of what their contribution had contributed to, as is recommended from an ethical standpoint [26].

A few months after the conclusion of the explorative design workshops, a fourth validation workshop was conducted with 7 children aged 10-12 years. These children had not been part of the previous design workshops or focus groups and did not have a history of cancer. The work in this workshop was performed in groups of 2-3 children with 1-2 adults for 3 hours. The aim of this workshop was to gain feedback related to feasibility on presented design concepts.

XSL•FO
RenderX

Workshop 1: Proxy Personas

The aim of the first workshop was mainly to create familiarity within the group, as not all participating researchers had met the children. The aim was to create proxy personas that the participants could identify with and that could be used as the base of the redemption scenarios of the second workshop.

Each child-adult pair created one cut-out-doll character and presented their character to the rest of the group at the end of the session. During the character creation process, the facilitator motivated the participants by asking questions about the characters that the pairs were creating. These questions ranged from concrete and tangible (eg, "what is your character's name?") to more value-oriented and abstract questions (eg, "what does your character like or dislike?" and "what is important to your character?"). The latter questions were informed by the focus group interviews, as well as the information gained from the stakeholder interviews. From this work emerged basic demographic information as well as values and motivational aspects of the characters. This information about and descriptions of the characters were after the workshop summarized and compiled into visually appealing and informative descriptions that we called proxy personas.

Workshop 2: Redemption Scenarios

The aim of the second workshop was to create redemption scenarios in the form of stories, using the characters created during the first workshop. Initially, the children were asked if they wanted to change anything about their characters. When they were satisfied with the characters, they each again teamed up with an adult and set to the task of completing comics. This process is similar to that detailed in [33]. The pairs were given the beginnings and endings of comics that detailed scenarios that were initially problematic (eg, the character feeling that their friends didn't understand them) and eventually were solved through social interaction (eg, the character feeling that their friends understood them better). The pairs were asked to fill in the middle, that is, how the problem was solved.

Each storyline was set a priori based on theories of friendship and peer support [4] and the conceptual model of friendship established from the focus group interviews [27]. The purpose of the task was to gain the children's perspective on how these positive experiences were created. The comics allowed the design teams to explore themes such as friendship, social support, and play. Yet, by using comics and characters as proxies, the children were not speaking about themselves and their own experiences, thus reducing the risk of their participation being upsetting.

The story-creation process allowed us to learn about the children's reasoning and preferences for social interaction, and how they currently received support from both family and peers. The process corresponds to context scenario creation [14], where the characters act as personas.

Workshop 3: Design Session

The focus of the third workshop was ideation. The group with 3 boys was asked to give feedback, discuss, and change design concepts presented as continuations of the comics they created in the second workshop. The group with 2 girls was asked to sketch and prototype design solutions that fit into the comics' redemptions scenarios. The aim was not to identify more design ideas but to keep exploring the children's interests and motivations. Therefore, the ideas were allowed to vary in both theme and quality. Whereas some concepts showed promise, others were not realistic. Nevertheless, all ideas were captured and made part of the subsequent analysis.

Workshop 4: Feedback on Feasibility

The fourth workshop was, as mentioned, conducted with a group of children who did not have a history of cancer and had not previously been part of the research project. The aim of the workshop was to gather feedback on suggested design concepts. We chose to include healthy children at this stage to not overuse the limited target group [26], and the sought feedback was not unique to this target group.

The children worked in teams of 2-3 and moved between 3 stations. Each station was run by 1-2 adults and was given approximately 30 min. The concepts presented at each station were in the form of low-fi prototypes, and the children gave both verbal and drawn feedback. The feedback varied from simple adaptations to completely new design concepts, both feasible and unfeasible. The workshop was concluded with an open discussion in the entire group.

Modeling

The modeling phase comprised qualitative analysis of the collected data [34] and is summarized in Table 1. Two kinds of models were the primary outcome from this phase: personas, and context and key-path scenarios [14]. These models guided the identification and exploration of a number of design concepts and also played an important role in the further design work; for example, by allowing us to maintain a user focus in the stages where users were not actively involved. The empirical data with which we entered into the modeling phase consisted of:

- A conceptual model of friendship from the perspective of children with experience of cancer treatment (from the focus group interviews)
- The pathogenic perspective of the children's experiences (from stakeholder interviews with health care professionals and parents)
- Proxy personas (characters) cocreated with children (from design workshop 1)
- Redemption scenarios depicting stories about friendship, social support, and sensitive contexts, cocreated with children (from design workshop 2)
- Prototypes of design solutions that fit the redemption scenarios, cocreated with children (from design workshop 3)
- Dialogue and interaction between children, researchers, and designers, developed and deepening over 3 design workshops (from design workshop 1-3)
- Feedback on and adaptations of design concepts by children (from design workshop 4)

XSL•FO
RenderX

Table 1. Summary of the empirical data from the discovery phase (step 1-3) and the persona construction progression in the modeling phase (step 4).

Steps	Focus groups	Stakeholder interviews	Design workshops	Modeling
Participants	5 groups with in total 5 girls and 10 boys, 8-12 years old, and treated for cancer	13 interviews with in total 9 health care professionals and 4 parents of children treated for cancer	2 groups with in total 2 girls and 3 boys, 11-13 years old, and treated for cancer, participated in 3 sequential workshops 1 group with in total 3 girls and 4 boys, 13 years old, without experience from cancer, participated in a final workshop	
Types of data	Interview transcripts Video footage	Interview transcripts	Characters (proxy personas) Redemption scenarios (comics) Video footage	Affinity diagrams
Findings	Conceptual descriptions of friendship	Roles and effects of a DPS[a] from a pathogenic perspective "Wicked" design challenges Onboarding and positioning of service in relation to existing health care processes	Scenario insights on behaviors, attitudes, motivations regarding friendship, relations, peer support, and redemption strategies	Categories and behavioral dimensions for boys and girls in the user group
Progression of persona construction	Discovering central salutogenic concepts (ie, friendship, peer support) that need stakeholder input	Defining the context (ie, vulnerability, friendship with peers, and surrounding environments such as hospital, home, and school Friendship, cancer or disease aspects	Building character and story: Narrative constructs based on the concept and context captured in scenarios	Complete primary personas (Anton, Julia, and Anna)

[a]DPS: digital peer support.

The data from the focus group interviews were analyzed using a qualitative content analysis approach [35]. The analysis process produced a conceptual model of friendship, with the specific perspective of children with a history of cancer [27]. This model consisted of three generic categories: (1) "common interests and experiences," (2) "mutual empathic actions," and (3) "mutual trust and understanding" that describe the progressive process of becoming friends. The model was central in the subsequent work, including the framing of the themes for the stakeholder interviews and the redemption scenarios during the design workshops.

The qualitative analysis of the stakeholder interviews identified a number of so-called "wicked problems" of design in this context [5]. There is, for example, the issue of screen time, where we want to encourage outdoor physical activity to improve the children's wellbeing, yet are at risk of increasing screen time by the introduction of a new digital service. The analysis of the stakeholder interviews further identified a suitable onboarding process for the DPS service and positioned the service as strategic and salutogenic [5].

The design workshops were analyzed qualitatively in order to model the rich gathered data [36]. The proxy persona descriptions and comics were combed for details on behavior, motivation, goal fulfillment, technology use, social media interaction, social interaction, and communication channels (analog and digital). Three researchers wrote down all terms relating to this aspect that they could identify in the material, and to this the pathogenic input from the stakeholder interviews was added, along with additional concepts derived from quotes from the dialogue during the design workshops. All terms were written on post-it notes and assembled on a wall, where they were grouped by similarity in an iterative and collaborative fashion. Each change was made with an articulated intention (eg, "I'm moving term X to category Y because of Z").

From this, clusters emerged and formed categories into an affinity diagram [36]. A total of 40 categories were formed and these were in turn organized into eight themes. For example, the theme "creativity" contained the categories building, drawing, and taking photos. Some categories in turn contain subcategories with multiple connections. The category "building" is for example separated into digital construction (with atomic data points such as Minecraft) and analog construction (eg, playing with Lego or building a tree house). Minecraft also belongs to the category computer games in the theme "consume," whereas building a tree house also belongs to the theme "physical outdoor activities."

The rich, multi-level affinity diagram that was created at this stage formed the basis for the creation of authentic, high-quality personas [14,36,37]. It was also noted early that certain categories were intimately tied to gender; several apparent differences between the data from the two groups of children emerged. For example, from both groups, the themes of technology, creativity, and social interaction were identified, but the girls were more oriented toward photography and sharing photographs on Instagram, whereas the boys built Minecraft worlds while communicating using Skype.

XSL•FO
RenderX

Results

Coconstructed Child Personas

Following the persona-creation process advocated by Wärnestål et al [34] and Cooper [14], three primary personas were created: 1 boy (Anton) and 2 girls (Julia and Anna). The main distinction between them was their main goal for interacting with the DPS service. Table 2 shows a summary of the project's three primary personas, covering the complete service lifecycle from initial user onboarding (Anton), through continuous use (Julia), to an "alumni" perspective after exiting the peer support service (Anna). Figure 1 shows the overview page of Anton's persona description.

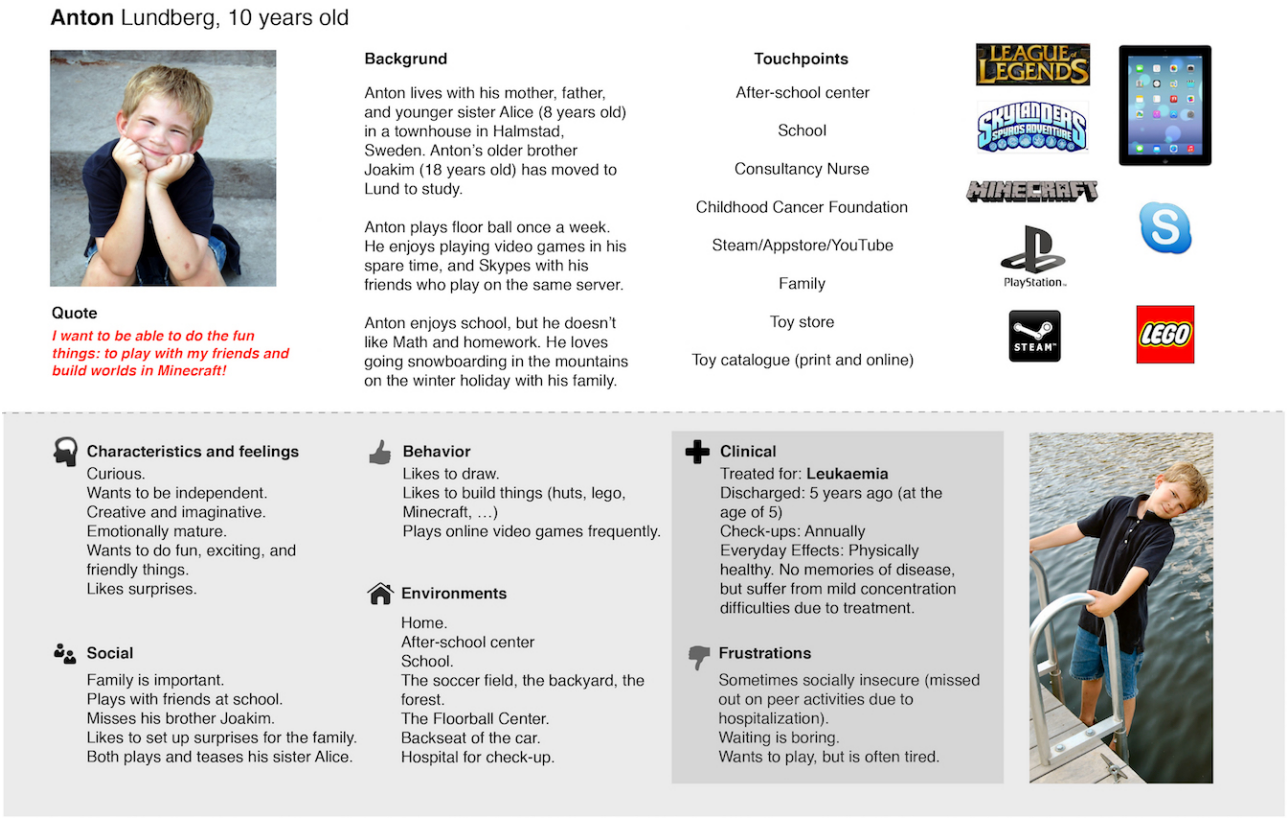
The progression of the personas' richness is based on the four steps summarized in Table 1. The persona descriptions of Anton, Julia, and Anna contain interpreted data from focus group interviews, stakeholder insights derived from the interview analysis, as well as a distillation of the proxy persona attributes and social redemption scenarios discovered in the 3 design workshops. Of particular interest for child-persona construction in this design context are results relating to: (1) pathogenic versus salutogenic perspectives, (2) platforms for Web-based communication, and (3) implications for avatars and conversational user-system interaction.

Table 2. Persona overview of the three primary personas.

Persona A	Persona B	Persona C
"Anton"	"Julia"	"Anna"
Male, 10 years old	Female, 12 years old	Female, 16 years old
Role:	Role:	Role:
Player (entering the DPS[a] service)	Mentor (experienced player)	Alumni (leaving the DPS service)

[a]DPS: digital peer support.

Figure 1. Persona description (translated from Swedish) of Anton, one of the project's three primary personas. The persona description merges the pathogenic perspectives (summarized in "Clinical" and "Frustration" categories) with the salutogenic perspectives.



Salutogenic and Pathogenic Perspectives

The personas manifest the synthesis of the pathogenic and salutogenic perspectives. The pathogenic orientation is most evident under the "clinical" and "frustration" headings of the persona descriptions. The other four headings are mainly derived from the salutogenic orientation. However, content in these headings are verified and complemented by pathogenic perspectives. For example, the characteristic "emotionally mature" was verified by both parents and health care professionals. The characteristic "wants to play" (described by children, and part of Anton's persona) was complemented with

XSL•FO
RenderX

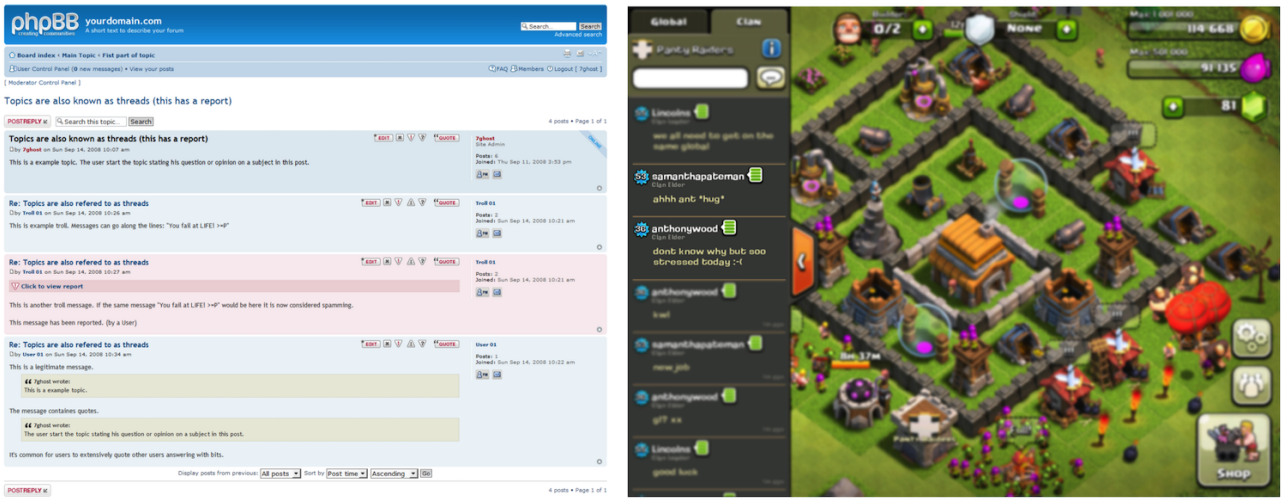
"but is often tired" in the "frustrations" heading from parent interviews. The result of such syntheses emphasizes the necessity of both the salutogenic and pathogenic perspectives when developing personas for a DPS service in this domain.

Child-Centric Views on Web-Based Communication

Due to its qualitative and rich nature, a number of child-centric insights that ran counter to what we learned in the interviews with adults (parents and health care providers), surfaced in the workshops. These aspects were incorporated in the personas to maintain an authentic end-user voice in the service design process that followed the persona generation. Figure 2 shows one such difference between child and adult perspectives on Web-based communication.

In the interviews with adults, the topic of "Web-based communication" between children came up several times. However, the references and examples used ranged from text-based forums (such as phpBB in Figure 2), mature social media platforms such as Facebook, to even more traditional media such as telephone calls. This stands in contrast to the media choices introduced and referred to by the children. Facebook was not used by a single child in our study, and the preferred communication platforms were either in-game conversations (such as the Clash of Clans chat interface in Figure 2) or newer social media platforms such as Kik and Snapchat. The two latter platforms were mentioned at all by neither health care providers nor parents.

Figure 2. Comparison between text-based forum (generic legacy phpBB on desktop web browser, referenced in the stakeholder interviews), and chat in the tablet-based game Clash of Clans (referenced in the children workshop sessions). Example of supportive chat between the players is highlighted in the game interface (e.g. second and third utterances in the chat area to the left in the Clash of Clans interface).

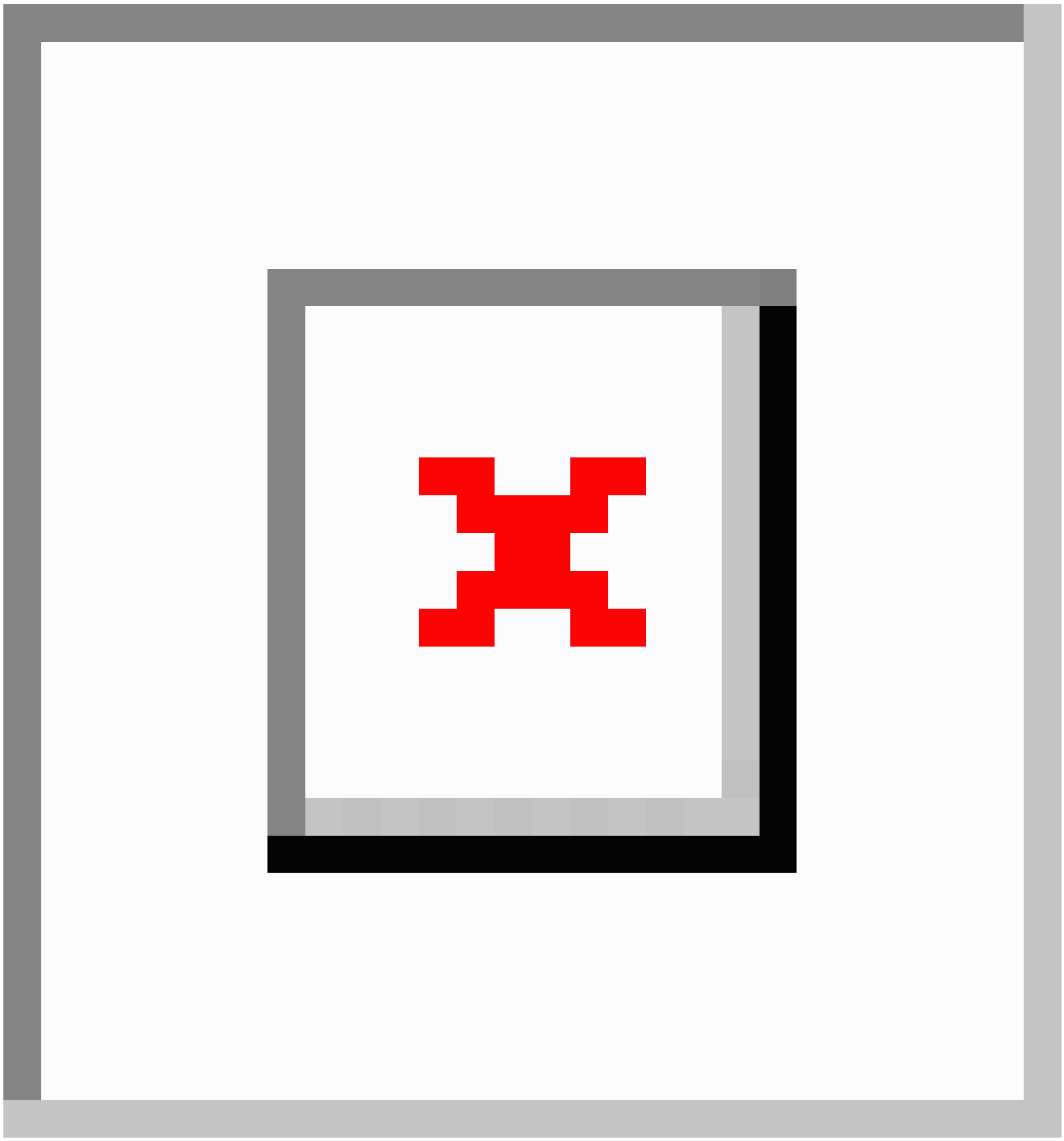


Design Implications for "Give Me a Break"

Give Me a Break is a tablet-based, Web-based service developed for childhood cancer survivors based on the three personas described in the previous sections of this paper. The service is a functional prototype and has been tested with real users [38]. A detailed description of the service is beyond the scope of this paper. However, we highlight some features of the service that exemplify the effects of using the coconstructed personas. Figure 3 shows the gradual revelation of the onboarding sequence. Bobo, a friendly and conversational robot, greets the user and explains the sign-up process in discrete steps. This solution is based on the Anton persona's need to carefully understand things in manageable steps, and all three personas' desire to engage in conversation rather than standard form-filling sign-ups. Bobo helps the user craft a Web-based avatar that the user navigates the virtual environment with. Bobo follows the user into the game world and explains the core concepts that the user can explore. As the user gets acquainted with the mechanics of the game experience, Bobo's presence diminishes to let the user focus on conversational interactions with others instead. Figure 3 shows one of the scenes in the environment. In order to trigger conversations between users, which are the underlying mechanic to achieve peer support effects through friendship [26], user avatars display content from their respective interests listed in their profiles that Bobo helped set up in the onboarding process. By automatically displaying a user's interest thought bubble (eg, "I like to draw" or "I love to play Minecraft"), the avatar provides an actionable conversation starter for surrounding users. This resonates with both Anton and Julia, who both are driven by social interaction, but are cautious around new people and sometimes unsure of how to approach others.

XSL•FO
RenderX

Figure 3. Screenshots (a, left) and (b, right) from a Swedish high-fidelity prototype of the digital peer support (DPS) tablet app "Give Me a Break." Screenshot (a) shows the user's first encounter with the tutor robot "Bobo" in the onboarding process. Screenshot (b) shows the virtual environment where player avatars display their interests in thought bubbles in order to spark interest and conversation. In this particular screen, the purple dinosaur displays the user's interest in drawing.



Discussion

Principal Findings

The methodological implications for a child-centric design process in this study is that a UCD process needs to be customized to fit the design space for (1) child users; (2) the domain of health-promoting and digital, social services; and (3) sensitive contexts. These three qualities are not explicitly addressed by generic persona-driven methods developed for adult, professional users in a productivity- or efficiency-oriented domain. The traditional principles rather regard goals such as minimizing work [14] and to increase productivity and reduce costs. For children's interaction with technology in terms of digital toys, games, and other more experience-oriented products, such as social media platforms, principles beyond productivity and effectiveness have to be acknowledged [39,40].

Cocreative aspects have been voiced as an important part of child-computer interaction, such as Druin's work on cooperative inquiry [7,41], and the initiatives stemming from the Scandinavian participatory design approach [9,42]. Such advances address the needs of developmental abilities of children when involving them in the design process. However,

XSL•FO
RenderX

it is difficult to directly implement standard interview and participatory design techniques for young children because of difficulties with abstraction and verbalizing conceptual problem solving [39]. Furthermore, in the context of designing for children in sensitive contexts relating to disease, there are pathogenic aspects that designers and researchers need to incorporate in any design efforts for the target group, that need to be collected in other ways due to ethical reasons as outlined previously. In our approach, the synthesis of salutogenic and pathogenic input to the child personas help designers to maintain an authentic user-centered approach in such contexts.

The rest of this section discusses the effects of using the personas in the design of a DPS service for childhood cancer survivors. First, we discuss how personas were used as communication aid within the design team, as well as for outside communication of the project. Second, experiences on how personas worked as an idea-generating tool are described. Third, we describe how personas were used as tools for making design decisions throughout the design process.

Personas as Communication Aid

Due to their familiar narrative-based form, persona descriptions provide a common language for discussing effects of the design. By referring to the primary personas, a consistent picture of the target user group could be formed. Personas were used as communication aid between a wide range of disciplines and roles (interaction design, research team, care-givers, parents, game developers, and so on), and helped build a common vision in the within an interdisciplinary team and external actors.

Within-Team Communication

Within a multi-disciplinary group, comprising researchers and practitioners from nursing, medical science, and human-computer interaction—as well as children users—character and story are the "lowest common denominator." Personas and scenarios provide the lingua franca for service innovation and development, where designers, researchers, children, parents, and caregivers have a common ground in the conversation about the service. Since the children had taken a very active part in the development of the proxy personas and scenario ideas, it also empowered them in the design discussions.

The persona descriptions also help enforce a "project language," which is necessary to avoid misunderstandings and misinterpretations. We found that using personas and accompanying scenarios formed a productive foundation for workshops and idea generation sessions. Furthermore, the use of personas helped us communicate around the design's effectiveness.

Involving External Actors

External collaborators (such as the game production company that built the first high-fidelity prototype of the DPS service, or business investors) clearly understand the essence and intent of the service when presented with personas and scenarios.

Personas and scenarios were used effectively in business modeling workshops for communicating and developing business models surrounding the service. By employing a narrative in the form of scenarios where the primary personas interact with the service, it was clear how business model decisions could affect the user experience of the service. This allowed us to measure the design's effectiveness and see the implications in terms of organizational structure and other strategic planning activities related to the DPS service development.

The embodiment of the users' goals and needs makes it easier for the team to focus on the common understanding of the end users. This is especially important in multidisciplinary teams, as well as when interacting with external actors and developers. The personas therefore contributed to building consensus and commitment to the design.

Personas as Catalysts for Dealing With Legal and Ethical Positioning

Our three personas functioned as rich archetypes, which effectively steered the design team members away from stereotypical renderings of the user group. The persona descriptions helped both stakeholders and designers to reduce the users to caricatures, and instead allowed the team to talk about the rich and authentic personalities of Anton, Julia, and Anna. By relating to and designing for authentic people (fictional characters—yet based on rich empirical data), both researchers and designers voiced that they increasingly understood and cared about what would happen to the personas given a specific design suggestion. In short, the team developed empathy for the personas, in a way that the elastic concept of "the user" would not. This was evident in the discussions regarding ethical and legal positioning. By thinking and reasoning about the consequences of a design decision in terms of an authentic user, new aspects were uncovered that could be addressed at an early stage. One clear example of this is the way that the members of the team routinely would ask: "How would Anton feel if the service was implemented this (or that) way?"

Personas and Scenarios as Idea Generators

Having personas based on salutogenic input from the children, it was natural to evolve the solution into playful interaction based on positive health outcomes. This complemented a more traditional approach of letting stakeholder input (which in this project would have been mostly pathogenic) dictate design decisions. Instead, the pathogenic aspects served as a backdrop to check ideas generated from the primary salutogenic persona needs.

Positioning, wicked problem identification, and onboarding (stakeholder input) is not enough to provide a solution that resonate with this target group. The salutogenic perspective, the children's point of view regarding what kind of service they want to have, and the interactive and aesthetic qualities (ie, what the children are used to and what they appreciate in terms of digital service interaction with games and other apps) were not captured in a relevant form in the stakeholder interviews. It proved to be quite the opposite: both parents' and medical professionals' views on digital interaction and digital services were sometimes both naive and outdated. Phrases such as "what kids do on the Internet these days is beyond me" and "I don't understand all these new apps that the kids are using" are

XSL•FO
RenderX

examples of this sentiment. In the interviews, our stakeholders referred to text-based discussion forums and Wikis as possible platform ideas for the service, whereas the social interaction platforms that were suggested by the children codesigners never even mentioned these. Instead, they mentioned more contemporary social media platforms, as well as in-game chats. Figure 2 illustrates the difference in reference frames between the stakeholder and end-user (children) input.

Scenarios that exemplify how Anton, Julia, and Anna interact with relevant social media services (as of early 2016 these include Snapchat, Instagram, Skype, Facetime, and various in-game messaging solutions), could then serve as a means of communicating this effectively to the stakeholders.

Persona-Driven Design Decisions

The developed personas guided design decisions of both major significance for the whole project and at a level dealing with minor design decisions regarding function and aesthetics of the digital service. In this section we exemplify how personas influenced the service design at these different levels. Example 1 is a major service design decision, helping setting the entire vision of the project. Example 2 is a medium-level service design decision that highlights how the stakeholder interview analysis gives insight on when and how the onboarding takes place, modified by persona behaviors and characteristics, rendering the onboarding scenario tailored to fit Anton's profile creation. Example 3 is lower-level interaction design decision, helping enhancing usability and user experience in the initial interactions in the platform.

Example 1: Shaping the Entire Service Experience

An important synthesized insight from children dialogues and stakeholder input is the need for training social skills for this particular user group. Since they have missed large portions of time in school due to being hospitalized, they miss opportunities of training social skills during school breaks and extracurricular activities. In the words of the children themselves: "on the breaks you learn how to have fun with friends." This, along with interviews with young adults that want to reconnect with lost acquaintances on other platforms, the idea of a (long-term) virtual real-time playground emerged. By examining the different attitudes, interests, and behaviors of the three primary personas, it was clear that newcomers and "old-timers" would have different needs and skillsets on a DPS platform. Hence, the idea of player, mentors, and alumni emerged and were integrated into the design of Give Me a Break. These roles were crystallized and refined as a direct result of scenario building around the personas and shaped the entire peer support service to be centered around player onboarding, mentor interaction between both players and mentors, as well as supporting contact facilitation between alumni in other channels and platforms after they have stopped using the original peer support service.

Example 2: Addressing Persona Behaviors in the Onboarding Context

During service onboarding, the initial concept was a fairly standard user profile setup. Setting up user profile interests was initially designed as a "wizard-style" step-by-step interaction before entering the service and engaging with other users. When checking this procedure with the Anton persona, the team realized this was a boring, noncreative way of doing it, and stood in contrast the Anton's behavior pattern of creating and exploring. So instead, the design team integrated the profile building as part of the playground interaction. Setting up user profile interests can now be done through a number of explorative and engaging ways, for example walking up to other people and watch their interest appear in bubbles above their avatar, or walking past an "Instagram signpost" that reminds the user to add his or her Instagram user name, or getting tips from a nonplayer helper character in the shape of a tutor robot called "Bobo" that accompanies Anton's avatar as he is familiarizing himself with the playground environment (see Figure 3).

Example 3: Understandability and Ease-of-Use

For any digital service, a user needs to be able to understand the interaction design. Anton's need to understand what to do on the playground is thus a basic usability requirement, and the challenge for a service and interaction designer is to accommodate the specific needs and provide the right kind of assistance and pedagogic vectors for Anton so that he becomes successful in using the service.

In order to design a customized experience for this particular user group, the personas were critical. Anton's onboarding scenario, which includes offline usage mode in case Anton doesn't have Internet connection, and the interaction design with the robot Bobo is a direct effect of (1) Anton's onboarding context, (2) Anton's mild insecurity due to his condition (stakeholder input), and (3) Anton's explorative and creative mindset (salutogenic input).

Strength and Limitations

The point of user personas is to be a useful abstraction and visualization of salient aspects of what users want, their relationship with the service being envisioned and built, and how they behave in relation to the service and other human beings. Interviews and workshop material provided us with patterns and insights regarding attitudes, goals, behaviors, skills, and needs of our users. Our three personas represent behavior patterns specific to salutogenic play (from a long-term, strategic perspective). As the service platform is iteratively refined and its development continuously guided by user tests, new insights can naturally be incorporated in the framework provided by our personas and scenarios.

Critique could be voiced regarding the lack of statistical significance in our findings. This critique can be met, as the kind of understanding provided by this method may not surface through a strictly quantitative approach. We argue that the method of using personas works as a complement for gaining critical insights for service design where a human interactive and social component is key. Qualitative studies put trustworthiness in focus and deals with the traditional lenses of objectivity, dependability, credibility, relevance, and transferability [39,43].

XSL•FO
RenderX

Conclusions

The motivation behind the research reported on herein stems from methodological constraints put forward by the specifics of vulnerable children in health-related, sensitive contexts. As design thinking and UCD practice find their way into processes that aim to cater for patient experience in health care [44], there is a growing need to customize methods and techniques originally devised for more traditional applications and domains.

This paper provides insights and resolutions to difficulties that can arise when empirical data from end users are restricted due to clinical and ethical reasons. By combining the salutogenic (from the children) and pathogenic (from adult stakeholders) perspectives, we could learn about aspects related to both health and disease that are important for the design of authentic, child-centered personas that could be employed for investigating complex themes of friendship and wellbeing, peer support and relationships, and the role of social technology in daily life. The design-oriented method described in this paper is ultimately about understanding human behavior in relation to a specific situation and context. It is therefore applicable even in processes where a digital artifact or service is not necessarily the outcome.

Acknowledgments

We gratefully acknowledge the grants from the Swedish Research Council, the Swedish Research Council Formas, the Swedish Childhood Cancer Society, the Knowledge Foundation, and the Swedish Governmental Agency for Innovation Systems (Vinnova). We are also grateful for the app development help from game design agency Hello There.

Conflicts of Interest

None declared.

References

1. Eiser C. Beyond survival: quality of life and follow-up after childhood cancer. J Pediatr Psychol 2007 Oct;32(9):1140-1150. [doi: 10.1093/jpepsy/jsm052] [Medline: 17644535]
2. American Academy of Pediatrics Section on Hematology/Oncology Children's Oncology. Long-term follow-up care for pediatric cancer survivors. Pediatrics 2009;123(3):906-915.
3. Svedberg P, Einberg E, Wärnestål P, Stigmar J, Castor A, Enskär K, et al. Support from healthcare services during transition to adulthood - experiences of young adult survivors of pediatric cancer. Eur J Oncol Nurs 2016 Apr;21:105-112. [doi: 10.1016/j.ejon.2016.02.008] [Medline: 26952685]
4. Solomon P. Peer support/peer provided services underlying processes, benefits, and critical ingredients. Psychiatr Rehabil J 2004;27(4):392-401. [Medline: 15222150]
5. Wärnestål P, Nygren JM. Nygren, Building an experience framework for a digital peer support service for children surviving from cancer. In: Proceedings of the 12th International Conference on Interaction Design and Children. 2013 Presented at: International Conference on Interaction Design and Children; 2013; New York p. 269-272.
6. Stinson JN, White M, Breakey V, Chong AL, Mak I, Low KK, et al. Perspectives on quality and content of information on the internet for adolescents with cancer. Pediatr Blood Cancer 2011 Jul 15;57(1):97-104. [doi: 10.1002/pbc.23068] [Medline: 21328524]
7. Druin A. The role of children in the design of new technology. Behav Inf Technol 2002;21(1):1-25.
8. Söderbäck M, Coyne I, Harder M. The importance of including both a child perspective and the child's perspective within health care settings to provide truly child-centred care. J Child Health Care 2011 Jun;15(2):99-106. [doi: 10.1177/1367493510397624] [Medline: 21685225]
9. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. J Biomed Inform 2008 Aug;41(4):624-635 [FREE Full text] [doi: 10.1016/j.jbi.2007.10.004] [Medline: 18082468]
10. Bødker S. Creating conditions for participation: conflicts and resources in systems development. Hum Comput Interact 1996;11(3):215-236.
11. Schuler D, Namioka A. Participatory design: principles and practices. Hillsdale, NJ: L. Erlbaum Associates; 1993:334.
12. Beyer H, Holtzblatt K. Contextual design: defining customer-centered systems. San Francisco, CA: Morgan Kaufmann Publishers; 1998:496.
13. Goodwin K, Cooper A. Designing for the digital age: how to create human-centered products and services. Indianapolis, IN: Wiley Pub; 2009:768.
14. Cooper A. About face : the essentials of interaction design. New York: Wiley; 2014:720.
15. Pruitt J, Adlin T. The persona lifecycle: keeping people in mind throughout product design. Amsterdam: Elsevier; 2006:744.
16. Moules T, O'Brien N. Participation in perspective: reflections from research projects. Nurse Res 2012;19(2):17-22. [doi: 10.7748/nr2012.01.19.2.17.c8904] [Medline: 22338804]
17. Fargas-Malet M, McSherry D, Larkin E, Robinson C. Research with children: methodological issues and innovative techniques. Journal of Early Childhood Research 2010 Jun;8(2):175-192. [doi: 10.1177/1476718X09345412]

XSL•FO
RenderX

18. Coyne I. Accessing children as research participantsxamining the role of gatekeepers. Child Care Health Dev 2010;36(4):452-454. [doi: 10.1111/j.1365-2214.2009.01012.x] [Medline: 20642564]
19. Pruitt J, Grudin J. Personas: practice and theory. In: Proceedings of the 2003 conference on Designing for user experiences. 2003 Presented at: DUX - Conference on Designing for User Experiences; 2013; San Francisco, CA p. 1-15. [doi: 10.1145/997078.997089]
20. Fore D, Goldenhar LM, Margolis PA, Seid M. Using goal-directed design to create a novel system for improving chronic illness care. JMIR Res Protoc 2013 Oct 29;2(2):e43 [FREE Full text] [doi: 10.2196/resprot.2749] [Medline: 24168835]
21. Norman D. The design of everyday things. London: MIT Press; 1998.
22. Cooper A. The inmates are running the asylum. Indianapolis, IN: Sams; 2004.
23. Pruitt J, Adlin T. The persona lifecycle: keeping people in mind throughout product design. Amsterdam: Elsevier; 2006.
24. Antonovsky A. Health, Stress and Coping. London: Jossey-Bass; 1979:225.
25. Lindström B, Eriksson M. Salutogenesis. J Epidemiol Community Health 2005 Jun;59(6):440-442 [FREE Full text] [doi: 10.1136/jech.2005.034777] [Medline: 15911636]
26. Lindberg S, Thomsen M, Åkesson M. Ethics in health promoting PD: designing digital peer support with children cured from cancer. 2014 Presented at: 13th Participatory Design Conference; 2014; Windhoek, Namibia p. 1-100. [doi: 10.1145/2661435.2661449]
27. Einberg E, Svedberg P, Enskär K, Nygren JM. Friendship relations from the perspective of children with experience of cancer treatment. J Pediatr Oncol Nurs 2015 May;32(3):153-164. [doi: 10.1177/1043454214554009]
28. Heary CM, Hennessy E. The use of focus group interviews in pediatric health care research. J Pediatr Psychol 2002;27(1):47-57. [Medline: 11726679]
29. Horstman M, Aldiss S, Richardson A, Gibson F. Methodological issues when using the draw and write technique with children aged 6 to 12 years. Qual Health Res 2008 Jul;18(7):1001-1011. [doi: 10.1177/1049732308318230] [Medline: 18552326]
30. Driessnack M. Draw-and-tell conversations with children about fear. Qual Health Res 2006 Dec;16(10):1414-1435. [doi: 10.1177/1049732306294127] [Medline: 17079802]
31. Colucci E. "Focus groups can be fun": the use of activity-oriented questions in focus group discussions. Qual Health Res 2007 Dec;17(10):1422-1433. [doi: 10.1177/1049732307308129] [Medline: 18000081]
32. Lindberg L. Participatory design workshops with children with cancer: lessons learned. In: Proceedings of the 12th International Conference on Interaction Design and Children. 2013 Presented at: Interaction Design and Children; 2013; New York p. 332-335.
33. Moraveji N, Li J, Ding J, O'Kelley P, Woolf S. Comicboarding: using comics as proxies for participatory design with children. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: ACM; 2007; San Jose, CA p. 1371-1374.
34. Wärnestål P, Svedberg P, Nygren JM. Co-constructing child personas for health-promoting services with vulnerable children. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems. 2014 Presented at: CHI'14; 2014; Toronto, Ontario, Canada p. 3767-3776.
35. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
36. Kolko J. Exposing the Magic of Design: A Practitioner's Guide to the Methods and Theory of Synthesis. USA: Oxford University Press; 2011:208.
37. Goodwin K, Cooper AW. Designing for the digital age: how to create human-centered products and services. Indianapolis, IN: Wiley Pub; 2009:739.
38. Lindberg S, Svedberg P, Bergquist M, Nygren JM. Evaluating digital peer support for children cured from cancer. Int J Hum-Comput Int 2017 Jan 11:1-13. [doi: 10.1080/10447318.2017.1278892]
39. Antle A. Child-based personas: need, ability and experience. Cogn Tech Work 2007 Apr 11;10(2):155-166. [doi: 10.1007/s10111-007-0071-2]
40. Moser C, Fuchsberger V, Tscheligi M. Using probes to create child personas for games. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology. 2011 Presented at: ACE; 2011; Lisbon, Portugal p. 1-8. [doi: 10.1145/2071423.2071472]
41. Guha M, Druin A, Fails JA. Cooperative inquiry revisited: reflections of the past and guidelines for the future of intergenerational co-design. Int J Hum-Comput Int 2013 Jan;1(1):14-23. [doi: 10.1016/j.ijcci.2012.08.003]
42. Iversen O, Dindler C. A Utopian agenda in child–computer interaction. Int J Hum-Comput Int 2013 Jan;1(1):24-29. [doi: 10.1016/j.ijcci.2012.08.002]
43. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. In: Proceedings of the SIGCHI conference on Human factors in computing systems. 2007 Presented at: CHI'07; 2007; San Jose, CA. [doi: 10.1145/1240624.1240704]
44. Bate P, Robert GB, Maher L. Bringing User Experience to Healthcare Improvement: The Concepts, Methods and Practices of Experience-based Design. Abingdon: Radcliffe Publishing; 2007:224.

XSL•FO
RenderX

Abbreviations

DPS: digital peer support
GDD: goal-directed design
UCD: user-centered design

Edited by G Eysenbach; submitted 15.12.16; peer-reviewed by M Seid, J Taylor; comments to author 03.01.17; revised version received 22.03.17; accepted 08.04.17; published 18.05.17.

Please cite as:
Wärnestål P, Svedberg P, Lindberg S, Nygren JM
Effects of Using Child Personas in the Development of a Digital Peer Support Service for Childhood Cancer Survivors
J Med Internet Res 2017;19(5):e161
URL: http://www.jmir.org/2017/5/e161/
doi:10.2196/jmir.7175
PMID:28526663

©Pontus Wärnestål, Petra Svedberg, Susanne Lindberg, Jens M Nygren. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 18.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of Trust in an Online Breast Cancer Forum: A Qualitative Study

Melanie Lovatt[1], BA (Hons), MA, PhD; Peter A Bath[2], BSc (Technical), MSc, PhD; Julie Ellis[3], BA (Hons), MA, PhD

[1]Faculty of Social Sciences, University of Stirling, Stirling, United Kingdom

[2]Health Informatics Research Group, Information School, University of Sheffield, Sheffield, United Kingdom

[3]Department of Sociological Studies, University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Peter A Bath, BSc (Technical), MSc, PhD
Health Informatics Research Group
Information School
University of Sheffield
Regent Court
211 Portobello
Sheffield,
United Kingdom
Phone: 44 114 222 2636
Fax: 44 114 278 0300
Email: p.a.bath@sheffield.ac.uk

Abstract

Background: Online health forums provide peer support for a range of medical conditions including life-threatening and terminal illnesses. Trust is an important component of peer-to-peer support, although relatively little is known about how trust forms within online health forums.

Objective: The aim of this paper is to examine how trust develops and influences sharing among users of an online breast cancer forum.

Methods: An interpretive qualitative approach was adopted. Data were collected from forum posts from 135 threads on 9 boards on the UK charity, Breast Cancer Care (BCC). Semistructured interviews were conducted with 14 BCC forum users. Both datasets were analyzed thematically using Braun and Clarke's approach and combined to triangulate analysis.

Results: Trust operates in 3 dimensions, structural, relational, and temporal, and these intersect with each other and do not operate in isolation. The structural dimension relates to how the affordances and formal rules of the site affected trust. The relational dimension refers to how trust was necessarily experienced in interactions with other forum users: it emerged within relationships and was a social phenomenon. The temporal dimension relates to how trust changed over time and was influenced by the length of time users spent on the forum.

Conclusions: Trust is a process that changes over time and which is influenced by structural features of the forum, as well as informal but collectively understood relational interactions among forum users. The study provides a better understanding of how the intersecting structural, relational, and temporal aspects that support the development of trust facilitate sharing in online environments. These findings will help organizations developing online health forums.

(J Med Internet Res 2017;19(5):e175) doi:10.2196/jmir.7471

KEYWORDS

trust; online information sharing; breast cancer; online health communities; qualitative research

Introduction

Online health forums provide peer support relating to a variety of medical conditions, including long-term illnesses, acute health problems, and life-threatening and terminal illnesses. Previous studies show that users value forums as a way of accessing emotional and informational support from people who are going through or have already undergone similar experiences [1-3].

XSL•FO
RenderX

Advice and support are solicited and provided by users sharing personal and sometimes very intimate experiences, and trust is therefore regarded as a necessary component in facilitating the sharing of peer-to-peer support [4,5]. However, less is known about the processes by which trust forms in such communities [6]. In this paper, we address this gap in knowledge by exploring how trust develops among users of an online forum for people with breast cancer.

With the advent and increased usage of online platforms including social media, forums, and discussion boards, the disclosure of personal information in online social spaces has become commonplace. Consequently, researchers have become interested in the phenomenon of trust in online environments. Some research has investigated trust in the system or structure, whereby the technological affordances of the platform (for instance, Facebook or a specific discussion board) are deemed more or less trustworthy, thus persuading or dissuading users from sharing their information [7,8]. Other studies have explored the similarities and differences between assessing trustworthiness in online and face-to-face social interactions [9] and investigated how people judge the trustworthiness of information online [10].

Trust is particularly significant within the context of online health forums, given the sensitive nature of the content that is shared and the extreme circumstances that individuals are facing. Studies have explored different aspects of trust in informing people's decisions to share their personal experiences in such spaces. These include how technical affordances and the presence of moderators affect user inclinations to trust sites [11], the influence that individual personality traits and risk beliefs have on their inclination to disclose health information online [12], and the interpersonal nature of trust in online health forums, focusing on how user perceptions of relationships with other users affect whether and how they trust each other [5,13]. These studies have tended to regard trust either as a variable that can influence sharing or a behavioral outcome between and among forum users. Such approaches have been criticized for conceptualizing trust as a fixed and stable phenomenon, rather than taking into account its processual and relational nature, whereby trust emerges and develops over time and through changing interpersonal relationships [14].

A small number of studies have adopted a processual approach to conceptualizing trust, which is potentially more helpful in understanding how trust develops among forum users. Such research has identified that trust develops in separate stages or processes according to the trajectories of forum users and their changing relationships with other forum users. Radin [15], for example, identified 3 stages of trust on a breast cancer forum, in which users moved from lurking, to self-disclosure, to initiating virtual and face-to-face visits with other users. However, Radin used social capital theory to argue that people self-disclose by weighing up the potential risks and benefits of sharing about themselves which, we argue, implies a rational decision-making process that does not accurately account for user actions. Radin's research also did not explain how users assess each other's trustworthiness. Fan et al [16] focused on relational dimensions of trust to argue that users decide whether or not to trust each other based on the credibility of the information they post and their characteristics as posters (eg, frequency of posts and similarity to other users) and that some experienced forum users become trusted because of their established reputation and relationship with other users.

These approaches are useful in showing the interplay between relational and temporal dimensions in trust formation, in which trust develops between and among users and within online communities over time. Focusing on these 2 dimensions alone, however, risks downplaying the influence of the structure (for example, the platform) within which these trusting interactions take place. Sociological studies have conceptualized trust as a multidimensional phenomenon that operates at intersecting structural, relational, and temporal levels, whereby individual decisions concerning trust are influenced by social relationships over time, within particular sociopolitical contexts [14,17]. Rather than trust being considered an outcome or function of human actions, such approaches foreground the processes, contexts, and relationships within which trust becomes relevant and how individuals negotiate decisions about trust. We argue that a theoretical approach that conceptualizes trust as processual and relational, rather than as a fixed variable, is necessary to understand how trust emerges and develops. The overall aim of this paper is to examine how trust develops and influences sharing among users of an online health discussion forum.

Methods

Study Design

This study formed part of a larger project called a "Shared Space and a Space for Sharing" [18], which investigated how and why people in a range of extreme circumstances share information online and the importance of trust and empathy in the process of sharing. In this paper, we focus on how people with life-threatening and terminal illnesses use online support forums to share information, emotions, and experiences.

To gain an in-depth understanding of how trust manifests on the forum and develops between users, we chose a case-study approach and conducted thematic analyses of forum posts and interviews with 14 forum users. While most research has used either forum posts or interviews, few studies have combined these datasets [15,16]. We argue that using both provides a better understanding of how trust operates on the forum and that each dataset informs the other. For instance, while trust was rarely explicitly mentioned on the forum and therefore difficult to see by analyzing the posts alone, interviewee reflections on how they assessed and demonstrated trustworthiness shed light on how trust is done on the forum.

Study Setting

After seeking permission from the UK charity Breast Cancer Care (BCC) [19], we chose an online forum hosted by the organization for our case study. We selected this forum because it is one of the longest established Web-based forums, it is open access, allowing anyone to view the site and read the message boards, and the terms and conditions made it clear that the information may be used for research purposes. While anyone with Internet access can view the material, users must register and log in to post messages. At the time of the study, there were

XSL•FO
RenderX

an estimated 200,000 registered users, although not all of these are active at any one time (BCC, personal communication). Most users are women living with breast cancer. Other users include men with breast cancer, breast cancer survivors, and relatives and friends of those with the disease. The site is moderated by staff at BCC, whose responsibilities include ensuring that users do not disclose personal information that would reveal the identities of themselves or others, removing spam content, and intervening when necessary to manage conflict among users. The forum is organized into sections (eg, going through treatment), boards (eg chemotherapy, surgery), and threads (particular topics or queries initiated by users) within the boards. This description of the forum is how it was organized at the time of the data collection (2015). Since then, there have been a number of changes to the site, including the sections that are present in the forum [19].

Discussion Forum

Purposive sampling was used to identify a range of discussion boards that would provide suitable diversity of the topics discussed. In 2015, we collected 233 archived threads across a range of years (2006-2014) from 10 different boards on the BCC forum (see Table 1 for details). The variation in dates reflects the fact that different boards were created over time. For each board, we started with the first post made to that space and collected subsequent posts until we had reached our sample quota. This was a pragmatic decision designed to sample consistently across different boards. The forum data were analyzed thematically according to the approach detailed by Braun and Clarke [20]. ML, PB, and JE familiarized themselves with the data in the sampled threads and noted initial ideas independently (phase 1). These were then discussed to compare the initial ideas emerging from the data and to enhance interresearcher reliability. ML then selected 15 threads from 9 of the boards (135 threads in total) for further in-depth analysis and coding (phase 2) (see Table 1). These threads were selected according to which were most pertinent to the project's interests in the concepts of sharing, trust, and empathy. ML then coded the messages using NVivo 10 (QSR International), and a third of the messages were coded by PB and compared for analytical rigor. Following discussion and agreement between ML and PB, ML then grouped the codes into overarching themes (phase 3). These were then reviewed (phase 4) and refined by rereading the data extracts to ensure a close fit between the data and our conceptual interpretation of them (phase 5).

Interviews

Interview participants were recruited purposefully via a message posted on the BCC forum, explaining the purpose of the project and inviting potential participants to contact the project team. Inclusion criteria were that participants be at least 18 years of age, a user of the BCC forum, and either have a diagnosis of breast cancer themselves or be a relative or friend of someone with the condition. Thus, our final sample was self-selecting and was comprised of users who proactively responded to the recruitment message. While it is possible that one or more of the interviewees may have written forum posts that we analyzed as part of our sample, we deliberately did not seek this information. As it was not necessary to analyze their forum activity to achieve the project aims, we decided not to ask interviewees to reveal their forum user ID as this would have compromised their online anonymity.

Table 1. Details of forum posts included in the analyses.

			Phase 1 analysis	Phase 2 analysis
Name of section	Name of board	Dates of posts	Number of threads	Number of threads
Going through treatment	Chemo monthly threads board	2013	1	
Welcome to the forum	New members board	2012-2013	25	15
Have I got breast cancer?	Appointments and waiting for test results board	2007-2008	25	15
I am recently diagnosed	Diagnosed with breast cancer board	2007	25	15
Going through treatment	Surgery board	2007	25	15
Living with and beyond breast cancer	Sex and relationships board	2012-2014	22	15
I have secondary breast cancer	End of life board	2009-2010	25	15
Supporting someone with breast cancer	Family, partners, and friends board	2006-2007	25	15
Talk to people like me	Younger women and families board	2007	25	15
Talk to people like me	Men's board	2006	35	15

XSL•FO
RenderX

Table 2. Details of interview participants (all interviewees were women).

Name of participant (pseudonym)	Age range	Date of diagnosis	Type of interview
Anne	50-59	December 2014	Face-to-face
Beth	40-49	March 2012	Face-to-face
Christine	50-59	March 2014	Skype
Danielle	40-49	January 2015	Face-to-face
Eleanor	50-59	October 2012	Face-to-face
Frances	50-59	January 2014	Telephone
Gayle	50-59	June 2013	Face-to-face
Hazel	60-69	2012	Face-to-face
Isobel	40-49	January 2014	Telephone
Janice	60-69	March 2014	Telephone
Kathryn	50-59	December 2014[a]	Telephone
Libby	40-49	August 2014	Telephone
Nancy	40-49	December 2014	Telephone
Olivia	50-59	February 2010	Face-to-face

[a]Original diagnosis in 1992.

Interviews were conducted by JE with 14 forum users and took place between April and June 2015: 7 were conducted in person, 6 were by phone, and 1 was by Skype. Face-to-face interviews were conducted in a venue chosen by the interviewee. Table 2 presents the characteristics of the sample of interviewees, all of whom were women who had been diagnosed with breast cancer. All but one of the women were white British, with one white European. The mean age of participants was 52 years (range 40-67 [SD 8.0] years). The interviews were semistructured as we used a flexible schedule which guided the interview toward issues relevant to our study but allowed participants to introduce new topics. If a participant raised an issue of particular interest, this was incorporated into the schedule. Questions related to participant use of the BCC forum, their experiences of sharing online, their relationships with other forum users, and their experiences of and attitudes toward trust and empathy on the forum.

The shortest interview lasted 49 minutes and the longest was 2 hours and 30 minutes; the mean duration was 1 hour and 7 minutes. All interviews were audiorecorded and transcribed verbatim. Transcripts were analyzed thematically by ML following the 5-phase process described above [20], and interpretations and themes were discussed and agreed with PB and JE.

Analyzing Both Datasets

Once both datasets had been analyzed in turn, we analyzed them together in order to triangulate our analysis and arrive at a better understanding of how trust operated on the forum. Combining the datasets in this way helped to shed light on themes emerging from the analysis of the forum threads that we had not initially associated with trust. For instance, one of the key themes which emerged from analysis of the forum posts was humor. While not immediately seeing the relevance of this to trust, analysis of the interview transcripts revealed that one way in which participants presented themselves and perceived others as trustworthy users was by including amusing stories in their posts. Consequently, combining datasets revealed aspects of the forum data that had hitherto been invisible to us. It was at this stage of analyzing both datasets that we developed our conceptual framework of the structural, relational, and temporal dimensions of trust, as we found this a useful approach to understanding how the themes from both datasets related to trust and also to each other.

Ethical Considerations

The data collected from the online forum were publicly available, and under the terms of use of the BCC forum, users were informed that material posted was publicly accessible and so collecting informed consent from individual users was not necessary. This complied with the recommendations of the Research Ethics Committee. In order to preserve user anonymity, we deleted usernames and other identifying features (eg, names of organizations or other individuals). In accordance with the Research Ethics Committee's requirements, we have reworded quotations in instances where it might be possible to trace a user by inputting the quotation into an Internet search engine. In such cases, care has been taken to protect the anonymity of the user while retaining the meaning and nuance of the original message.

Interviewees received information sheets about the study prior to providing informed consent and were informed that they could leave the project at any time. Although the purpose of the interviews was to ask participants to discuss their experience of the forums rather than their experience of the illness per se, we wanted to give interviewees the opportunity to talk about the illness as a way of contextualizing their forum use. We were therefore mindful that some participants might become distressed and developed a protocol with BCC of what to do in

XSL•FO
RenderX

such instances; however, this was not required during any of the interviews.

Ethical approval was granted by the University of Sheffield Research Ethics Committee (analysis of forum posts; application 001955) and UK Ministry of Defence Research Ethics Committee (interviews; application 614/MODREC/14).

Results

Overview

Following our analyses, we conceptualized trust as operating in 3 different dimensions: structural, relational, and temporal. By structural, we mean how the affordances (the technological design of the forum which allows or prevents users from taking certain actions) and formal rules of the site affected trust. Relational refers to how trust was necessarily experienced in interactions with others; that is, it emerged within relationships and was a social rather than an individual phenomenon. The temporal dimension relates to how trust changed over time and was influenced by the length of time users had spent on the forum. These 3 dimensions intersect with each other and do not operate in isolation. For example, the trust which users had in the forum's structure affected the extent to which they trusted and related to other users. Further, the more time the users spent on the forum, the more they got to know each other, meaning that the relational and temporal dimensions are inseparable. While for analytical purposes we separate the dimensions into 3 different themes, we draw out instances where they influence each other. As we triangulated analysis of interviews and forum posts, we intersperse both sets of data in the following sections. Anonymized quotations from interviews are indicated by a pseudonym and age group following the quote in square brackets; quotations from the forum are presented without usernames and followed by [forum post].

Structural

Although the forum is accessible to anyone with Internet access, users must register before they can post messages. New users have their first 3 messages premoderated before their posts automatically go live on the forum. This enables moderators to identify and block spam accounts or other nongenuine users. Spam content is also screened out automatically by the forum's filter system. Additionally, moderators remind new users that the forum is public and strongly advise them not to use their real names as usernames or to provide any other identifying information. Thereafter, moderators typically keep their interventions to a minimum, but these initial security measures were perceived by participants as contributing to the safety of the forum and inclined them to trust other users, as they believed them to be genuine.

> *I do trust the site, yes...when you're completing, you know, the name that you're going to use on there, it's very much that you only put as much information as you want...so I feel we're quite secure in that way, um, and the things that people are putting, um, you trust what they say.* [Frances, 50-59 years]

While few posts explicitly mentioned trust in the forum, users often assured new posters that they found the site supportive and helpful.

> *...do share any worries or problems in the forum—all of us understand and there is always someone who you can rely on to help you...* [forum post]

New members of the forum are also given an initial message of support from a forum moderator.

> *...I am sure you will get plenty of useful advice and information from other people on the forum. You are also welcome to contact our Freephone helpline on 0808 800 6000—you can talk to someone confidentially about how you are feeling.* [forum post]

This welcoming message from a moderator, in addition to welcome messages from other users of the forum, gave new members confidence in the forum and helped them to develop trust in the online community.

> *Within minutes I got loads of responses that were really reassuring and I thought, oh great, that's really helpful...* [Anne, 50-59 years]

This structural aspect of the forum was important at a time when the new member had made their first post and emphasizes the importance of temporality, in that getting a swift reply from forum users in response to their first post was reassuring. It also highlights the relational aspects of helping the new member develop trust and relationships within the community.

However, despite the forum's formal procedures and guidance to users, not all interviewees realized that the forum was public, and one participant described being upset when an offline acquaintance discovered some personal information about her from reading the forum.

> *Well, up until I found out about, you know, the incident, I felt sort of 99.9 percent safe but now obviously I don't because I realize that people can come in and identify you...so it's not as safe as I thought it was.* [Christine, 50-59 years]

While overall the forum was regarded as a safe space in which participants felt they could trust each other, the incident described above shows that it was not able to guarantee participant anonymity completely and this could influence the trust which participants had in the site and how they shared information about themselves. In this instance, it appeared that the user had placed too much trust in the site, from her misunderstanding of how open the forum was, and had shared more information than she would otherwise have done.

The information shared on the forum was also generally trusted because users believed that the BCC moderators would act to ensure that no false or misleading advice was posted. In the context of a cancer diagnosis, trust in the informational content of the forum was perceived as being particularly important, as users may have been more desperate to try different treatments.

> *The trust on the information that you can pick up is different, um, because...around cancer there is a lot of um—well, it's a scary thing. It's life and death and people um, when their backs are to the wall will try*

XSL•FO
RenderX

> *anything and everything. You know, it's sort of drinking your own piss type territory. So, any, um, advice, I think has to be, um, looked at quite carefully and I think the...moderators actually monitor that so I think most of what is on there—all of what is on there is factual, informative or um, um, it's been moderated; maybe been taken down...[the moderators] might say "this isn't helpful" and that is a very good thing because it does, um, clean up to some extent the site and it makes it more trustworthy.* [Hazel, 60-69 years]

As the quote implies, because the forum was for people with cancer it meant that being able to trust the content posted was particularly important. The consequences of acting on misinformation may have resulted in serious negative health effects or using products that were not supported by an appropriate evidence base.

Forum users also provided guidance on the extent to which information on the Web could be trusted.

> *...Be very careful with what you read on the Web. Much of this is out of date, anything that's over a year old is out of date...* [forum post]

This kind of advice, coupled with the presence of moderators and the feedback from other forum users, helped users to develop trust in the forum.

The availability of the forum 24 hours a day meant that someone could post a message whenever they needed to share their feelings with others (eg, in the early hours of the morning when they felt anxious and could not sleep). It was also possible that there might be others online at that time who were able to comfort and provide some reassurance, helping users to trust that this could be a reliable source of support throughout the day and night.

> *My first post was...like two in the morning when I couldn't sleep and I was like...okay, can't sleep, rollercoaster whatever, anybody else feeling this? Is this normal? I'm crying half the night, can't sleep...and you kind of—somebody's up there all the time regardless (laughing) of what time of the night or day it is and...somebody came back saying yeah, it's normal, hear you, been there, it will get better. It's kind of that reassurance...* [Danielle, 40-49 years]

Relational

The site facilitated peer-support via the sharing of personal experiences, emotions, and information; it was therefore vital that forum users trusted each other. While trust was not often explicitly mentioned on the forum, interview participants indicated the ways in which they assessed the trustworthiness of other users and also demonstrated their own trustworthiness. In an online rather than face-to-face medium, in which they were unable to see gestures or facial expressions, trust occurred in the following ways: using an appropriate tone, being reciprocal, not claiming expertise that users did not have, and seeking or demonstrating similarity with other users.

Appropriate Tone

Using an appropriate tone in posts was regarded by participants as a key way in which they could demonstrate or assess trustworthiness. Inappropriate posts included those perceived as ranting or being overly negative. Such users were regarded as being too engrossed in their own concerns rather than responding to and supporting other users.

> *The other thing is it's not okay to be a negative person on the website...Everyone's allowed a one-off meltdown so long as you get over it the next day or the day after. You are not allowed to be permanently miserable because if you are, you get pretty much ignored...So, the whole trust thing, "Can I trust you with personal information?" "Yeah." "Can I trust you to want to interact with me?" "Yeah, only if I do it in the right way" (laughs).* [Anne, 50-59 years]

Although it was clear from the interviews that it was acceptable for a person to express negative feelings on occasions, it was not acceptable to be negative continually. To be trusted and for other users to respond to their posts, individuals were expected to be aware of and respond to the needs and feelings of others. This was also revealed in many instances of posts in which users apologized for ranting.

> *Sorry for ranting but I want to know what one out of 3 sentinel lymph nodes having cancer means?* [forum post]
>
> *Sorry to witter on but I really want to describe the context.* [forum post]

These demonstrate user awareness of not appearing overly preoccupied with their own situations and to be mindful of the effect this might have on others. As demonstrated in the quote from Anne, being perpetually negative was not tolerated and hindered trust, although occasional rants were acceptable and could foster trust.

Humor was also frequently used in posts as a way of conveying the right tone, even in very serious circumstances. Interviewees explained the value of humor in lightening what could otherwise be a very dark space and demonstrating trustworthiness.

> *...for me there was a sort of balance, but even if you were having a really shit week, you always said something hilarious or something...There's always something stupid and ridiculous that happens.* [Olivia, 50-59 years]
>
> *...I tend to turn to humor a little bit if I can because if you can't laugh, what are you going to do? (laughs).* [Libby, 40-49 years]

The following posts were typical of the humor presented on the forum.

> *Having been diagnosed on Valentine's Day (the consultant didn't even write it in a card for me!) for breast cancer, the plastic surgeon recommended a skin saving mastectomy/reconstruction.* [forum post]
>
> *Seriously it was much better than I was expecting. Obviously, no one wants to be in hospital but I felt relieved that the cancer was out. My boob and I were*

XSL•FO
RenderX

> *no longer friends after diagnosis so I wasn't that sorry to see it go in the end!* [forum post]

As well as humorous language, users frequently included emojis in their posts to help them convey appropriate emotions.

> *You've always got to have a smiley...smileys or emoticons are brilliant and now you can get the moving ones and all sorts...if you're annoyed you can have a big, red, angry face or whatever but...I mean, I was always a smiley user. If you read my posts there's probably a smile or something stuck on it...they weren't very good ones but...I think you can get some of it across by—by your emoticons or smileys...* [Beth, 40-49 years]

Language was crucial in demonstrating to each other that users knew the informal rules of the forum: that it was acceptable to share negative emotions such as anger or distress as long as this was done in an acceptable way, meaning that they would appear trustworthy. Conversely, forum users who were perceived as being overly negative in their posts or who did not moderate their comments either with humor or an explanation were mistrusted by other forum users because they transgressed the accepted norms.

Reciprocity

Trust was also influenced by the extent to which users displayed reciprocity. As a peer support forum, it was imperative that users provided as well as received support. Reciprocity could be demonstrated through language, and, relating to the previous theme, users had to convey a reciprocal tone in order to be trusted.

> *I think trust as well, for me, um, when people would go on and they'd just been diagnosed and they were ranting and all the rest of it and then they went on and on and on being me, me, me members of the forum and in the end I just wouldn't respond to them anymore because they never responded to anybody else. They never contributed.* [Olivia, 50-59 years]

Reciprocity could also be demonstrated through behavior such as explicitly responding to others or being a frequent user of the forum. Such reciprocal behavior conveyed that the person was trustworthy which, in turn, facilitated the sharing of support and the development of a sense of belonging to the forum.

> *The people I suppose that I trust are the ones that are there regularly first off...I suppose the people that actually try and answer your questions...it is just about "Do I feel this person understands what I'm trying to say and is trying to help me?" and if they are, then I do—try and do the same for them, and other people it's a bit more of a soapbox. It's a bit more, you know, "I feel terrible and you all should run around and help me," but actually, you've never bothered to post an answer to anybody else's thing or whatever and you just think, well, you know. Again, I might answer the first few but eventually I'm probably not going to bother because it's all about sharing and trying to help each other.* [Anne, 50-59 years]

> *I thought that frequency was really good to develop trust, that if you're on there fairly frequently people kind of trust that you're serious, that you—you know, that you care about them, that you—you're part of the forum; you kind of belong if you like.* [Olivia, 50-59 years]

Reciprocity was also a temporal as well as a relational characteristic as users applied their knowledge of each other's past actions to help them predict future actions, thus influencing their immediate decisions on the forum regarding trust. Reciprocity and feelings of regard for other users were evident in the frequent posts which explicitly thanked others for their support and offered well wishes.

> *Thank you once again to everyone who responded, it is so comforting to know that people out there who have never met me care enough to spend their time carefully thinking through and writing replies. If anyone needs advice from me about something they are going through, or about to go through then I will be very pleased to help.* [forum post]

> *Thanks everyone for replying. I feel a lot better now. This is a wonderful forum. I'd been crying so much but you have all made me feel better. Good luck with all your treatments.* [forum post]

Reciprocity therefore indicated that users were good forum users; they were concerned for each other and not just for themselves, which in turn helped to foster trust on the forum.

Not Claiming Expertise You Do Not Have

Just as forum users perceived that moderators helped to facilitate trust by monitoring and removing misinformation, users also assessed each other's trustworthiness according to their claims to expertise. Interviewees reported that they were inclined to distrust users who appeared to provide medical information without the caveat that they could only speak from personal experience.

> *I worry about people who give answers on things they're not qualified to comment on. So, I think anyone can advise anybody else of what they think but (unclear) don't present it as, you know, fact when it...isn't...particularly when you're talking about life-threatening illnesses...you know, clearly everyone's head is going to be in a mess, so...that's the absolute worst time to be presenting yourself as an expert in anything. Qualify it, you know, if you're going to say it and I am astonished at the number of people—"Oh, I'd never do this" or "You must always do that." I just think "How do you know?" (laughs).* [Anne, 50-59 years]

Another interviewee reported that if she read a post containing medical advice or information that she believed to be incorrect, she intervened to try and present a fuller account out of concern that other users might trust the content

> *...she was telling these people on the site not to eat sugar because sugar feeds the cancer, and I know...what she was sort of thinking...but that would have put the fear of God into a lot of people...I don't*

XSL•FO
RenderX

> *always trust what I read but that's usually—it's a medical thing and I think, well I know better than that but um, it's hard when there's other people that wouldn't know and then obviously read things like that...my concern is that, you know, there are people on there that—if they're trusting everything they're told, it can be a little bit of a scary place, um, because, you know...they then suddenly think oh God, I'm not going to eat any more sugar or I've got to go and do this...quite often if I see a comment like that I will comment and direct them to a site that will explain to them what that person's been trying to say with regards to, you know, the cancer needing energy and what have you...* [Frances, 50-59 years]

Frances' interventions also suggest a feeling of responsibility toward other forum users which could engender trust, both in making people aware that not all information should be taken at face value but also directing them to websites containing more reliable information.

Forum posts often contained medical information in which users emphasized that they could only speak from their own experience.

> *I can tell you about my decisions about hormone therapy but need to emphasize that this was my personal decision.* [forum post]
>
> *The four-week gap is very long (in my opinion).* [forum post]

In this way, forum users framed their advice as being derived from their own experiential knowledge rather than their being an expert.

Similarity of Other Users

There was also evidence to suggest that trust was influenced by the extent to which users shared similar characteristics or situations, with some interviewees saying that they were more inclined to trust other users who were of a similar age or who had a comparable family situation

> *Interviewee: Interestingly...when I've looked at the people that I reply to and speak to, generally on the forum, they're often very similar...So, when I look at—if I look and analyze or think about, um, the people that I liaise with, um, maybe on a weekly basis, that they seem to be similar people, you know.*
>
> *Interviewer: And that helps you to trust? Is that what you're—? Yeah.*
>
> *Interviewee: Yeah, yeah. I think so, yeah.* [Eleanor, 50-59 years]

There were frequent references in user posts to characteristics such as their age and how many children they had, and the insight provided by the quote above suggests that inclusion of these details may have informed user decisions on who to trust (ie, because they were in a similar situation, not only with respect to being affected by breast cancer but also in terms of their personal circumstances). Examples of posts in which users emphasized their similarity to each other follow.

> *What time is your appointment tomorrow? I'll be thinking of you—our situations seem similar—I have a 7-year-old and a 2-year-old.* [forum post]
>
> *I was reading your post and felt the need to reply. I lost my beloved mum just over a year ago after her breast cancer spread to her liver and stomach. Like you I am an only child and was very close to my mother.* [forum post]

The inclusion of such details may help to place users and make them appear more relatable and familiar to each other. This was something that seemed important in an online setting, where users are more limited in their ability to assess the extent to which other users are their kind of people than in face-to-face support groups.

Temporal

The BCC forum included a range of users, from those who were new to the site following a recent diagnosis through to users who had been using the site for several years. The ways in which users experienced and perceived trust did not necessarily remain constant during the time they spent on the forum but changed over the course of their illness and as they gained experience of using the forum and getting to know the other users.

Lurking or Deciding to Post

Interviewees had different experiences of and perspectives on initially joining the forum and deciding if and when to start posting. Some users joined the forum and started posting relatively quickly after diagnosis. For Janice (60-69 years), the extreme circumstance of receiving a cancer diagnosis and undergoing chemotherapy provoked the formation of trust among users "because chemo is such an intense experience that that trust is actually forged, that sort of initial bit." Another interviewee described herself as "an innately trusting person, so my default is I trust somebody until they prove otherwise...Um, I think the developing of trust—I think you just test it out don't you? You put something out there and you see what people say back…" [Olivia, 50-59 years]. In such cases, the trust required to start posting on the forum developed quite quickly.

Other users lurked for a period of time in order to get an idea of what the forum was like and to judge "Are these my kind of people?" [Anne, 50-59 years].

In their initial posts, forum users often mentioned that they had watched the forum for a while before deciding to post.

> *This is my first post as I have learned a lot from reading your posts but have always been too nervous to join in.* [forum post]
>
> *I'm excited to say this is my very first post! Hope you don't mind me joining in but I've been lurking and reading everyone else's messages for about 11 months now and thought it was about time I joined in.* [forum post]

Making the initial decision to trust the forum and its users enough for new users to post was therefore influenced by the nature of the reason for accessing the forum (ie, breast cancer) and users' natural inclination or disinclination to trust people.

XSL•FO
RenderX

Interestingly, users who posted did not always consider or were not concerned that their posts might be read by lurkers or anyone else who accessed the forum. The trust they developed was in other people who posted on the forum, but they did not appear to take account of other people who were not visible.

Developing Trust Over Time

The trust which users had in the forum and toward other users changed over time, and participants spoke of trust growing as they got to know other users better.

> *I do trust the women that I'm in the group with. Um, I suppose that builds as time goes on. Um, you don't know them to begin with and then, by now, you know, six months down the line you know them and we know each other on Facebook as well now. So, we know each other's real names (laughs) for instance. Um, so I do trust them...I think just, um, because I do only more or less post in the one thread with the same people, it's—you know, it becomes familiar with this set of people and—yeah, you're chatting about something with somebody that you know essentially.* [Libby, 40-49 years]

This suggests that users did not necessarily develop trust in the forum as a whole or with everyone but used particular sections or threads or where the same people tend to post and so become familiar with each other. This is an example of how the forum's structure intersected with the relational and temporal dimensions of trust formation, particularly in providing separate spaces, within which users got to know people and develop relationships over time.

Forum posts often included details of users' everyday lives, such as their holidays or hobbies, and these—as well as the information about their treatment and health condition—may have helped users to get to know each other better. This may have helped the development of trust.

> *I have to like someone before I'll trust them and I have to know them quite well really, yeah...going back to how you perceive people by how they write, which you have to—and you bear in mind that...you start speaking to somebody from, say—for almost six—daily for six months, which you probably wouldn't speak to your best friend daily for six months. So, you learn an awful lot about people each day and...you just have a conversation really and you...get a good um, idea of what people are like, unless they're very good at inventing a story about themselves, which I'm sure some people could, but on this particular kind of thread you wouldn't—you wouldn't make things up really.* [Beth, 40-49 years]

This also reiterates the importance of users demonstrating their trustworthiness through their writing and how this facilitated trusting relationships. Forum posts would sometimes allude to users knowing each other based on previous posts.

> *You are definitely right David. I've read some of your messages before and know that you are a warm caring person who cherishes his wife and family and is not ashamed to tell everyone.* [forum post]

In this way, trusting relationships developed over time.

Discussion

Principal Findings

The aim of this paper was to examine how trust develops and influences sharing among users of an online health discussion forum. In this paper, trust as a processual practice emerges as a complex concept involving a number of elements. Interviewee accounts in particular suggested that trust plays an important role in how people share information, experiences, and emotions in online health forums. In our study, we explored how trust manifested on the BCC forum in 3 main intersecting dimensions: structural, relational, and temporal. The structural affordances of the forum (eg, the presence of moderators and security features) inclined users to trust the site as a safe space where they could share personal details. Within the forum's structure, users conveyed and assessed trustworthiness relationally, that is, through their interactions with other users. Relational trust involved using appropriate language and tone in forum posts, behaving reciprocally, not falsely claiming expertise, and seeking similarities with other users. Trust was also necessarily a process that developed over time, from the initial decision by users to join the forum to ongoing changes in how users related to each other as they got to know one other.

In addition to confirming findings in previous studies, our paper makes 2 new key contributions to the literature on trust in online health forums. First, by triangulating analyses of forum posts and interviews with users, our research reveals the characteristics involved in assessing and conveying trust such as not ranting and deploying humor. Interviewees spoke explicitly of how these were instrumental in assessing and negotiating trust, and this was also evident in forum posts. While previous studies have identified that tone, humor and the temporal trajectory of forum users facilitate sharing on online breast cancer forums, they do not show explicitly, as we do, how these characteristics are related to trust [21-23]. Additionally, while studies of face-to-face interactions between health professionals and patients with cancer have identified that humor relates to trust, different conclusions have been made as to whether trust must be present before humor is used or if it emerges as a result of humor [24,25]. Our research supports arguments that humor contributes to the creation of trust and does not just result from it. Our findings also support research that suggests that people are more likely to perceive online information as credible when it is posted by users who are judged to have similar characteristics to themselves [26].

Second, while previous studies have considered how, separately, structural [11], relational [16], and temporal [15] dimensions influence trust on online health forums, our study demonstrates that these dimensions intersect and cannot be adequately understood in isolation from each other. Here we draw on sociological theories of intersectionality, which emphasize how analysis of empirical data is enhanced by considering how different dimensions interact with each other to shape how a particular phenomenon, experience, or identity is manifested [27]. We argue that through intersecting with each other, the structural, relational, and temporal dimensions become more

XSL•FO
RenderX

than their distinctive components, revealing that trust is processual and fluid rather than a fixed, unchanging variable. The structural affordances of the online spaces within the forum intersected with relational and temporal dimensions of trust by enabling people to get to know and trust each other by responding empathetically over a period of time. In addition to the structure of the forum allowing relationships to develop over time, this also allowed users to communicate with each other 24 hours a day. Consequently, the forum's structure facilitated relational interactions between users at times when users' family members and friends may have been unavailable to provide support. The organization and moderation of the BCC forum was relatively light-touch and, after new users registered and had their initial posts checked, the forum's structure may have been largely invisible to users apart from when moderators occasionally intervened. We suggest that this reveals the permissive rather than restrictive nature of the structure and is not indicative of the forum's insignificance as an influence on sharing practices. We find sociological conceptualizations of trust useful in highlighting these intersecting dimensions. For instance, as Brownlie and Howson [17] suggested in their study of trust in the context of MMR vaccines, "leaps of faith cannot be understood outside interactions and relationships nor isolated from the systems or institutions within which these unfold." Similarly, Khodyakov [14] has argued that conceptualizing trust as a process requires understanding temporal characteristics of trust and how they influence trusting relationships within systems or structures. Our 3-dimensional theoretical framework helps us to make sense of the processual nature of trust on such forums. In doing so it contributes to the gap identified in a recent meta-synthesis of qualitative studies of online communities for people with long-term health conditions, which concluded that while trust and reciprocity do exist in such sites, "far less is known about the process that facilitates them" [6].

Practical Implications of our Study

Our study has practical implications for how organizations that create and maintain online peer-support forums can help to make these trusted spaces where users feel able to share information and experiences. A growing literature suggests that organizations that manage online communities can cultivate trust through particular design features or management practices [16,28]. Our study adds to this knowledge through highlighting the importance of having in place appropriate structural aspects of forums—for example, a registration process and formal moderation. Moderators can encourage trusting relationships between users by removing offensive, inaccurate, or inappropriate posts. The structural aspects of forums help users to trust the online environment within which they elicit information from others about their condition, treatment, side effects, etc. This also helps people feel confident about sharing their own information, experiences, and emotions, and organizations that develop online discussion forums for patients and carers should ensure that these structural aspects are in place to help foster trust in the forum.

On the one hand, our findings show ways in which structural features built into forums can influence how users trust each other in a top-down approach. On the other hand, our research demonstrates how trust is also a bottom-up phenomenon that emerges out of ongoing relationships and informal rules. This suggests that structural elements may not influence trusting relationships in a simple cause-and-effect relationship but may be hard to predict and dependent on relational and temporal aspects. Our more nuanced findings indicate that future collaborative work involving organizations and researchers is needed to explore possibilities for forum design and management that are based on a more processual understanding of the agentic and emergent nature of user relationships. Ultimately, the interdependence of the 3 dimensions in the formation of trust suggests that organizations developing online discussion forums need to be aware of the importance of these as intersecting components for developing trust among users of forums.

Our study also has implications for clinical practice. People increasingly turn to online health forums on receiving a diagnosis, and while recent research suggests much of the information shared on such sites is of good quality [29], our findings can be useful for clinicians advising patients on what characteristics (eg, tone, claims to expertise by users) to look for before deciding to participate in online forums. Through using health forums, patients can become better informed about their condition and play a more active role in the decision-making process.

Strengths and Limitations of Our Study

The strengths of this study lie in its in-depth exploration of one online peer-support group and in combining analyses of forum posts and interviews with forum users. Using both datasets together enabled us to strengthen our analysis and the validity of our claims. We demonstrate instances where analysis of the interview transcripts helped us better understand themes present in the forum posts and vice versa. This approach is useful in analyzing the inner workings of online forums, where it would be difficult to understand user intentions and decisions from their forum posts alone. We argue that using these complementary approaches can be particularly beneficial in researching intangible phenomena such as trust, which are not necessarily explicitly discussed or made visible on forums [30].

Our study is limited to a single case study that concerns one health condition, breast cancer. The forum is predominantly used by women and all of the interviewees were women. All interview participants self-selected, and this may have introduced elements of bias. For instance, forum users with strong opinions toward the forum may have been more likely to respond to the recruitment message than others. It is likely that trust manifests differently on other forums used by people with different illnesses and of different demographic groups [31]. In particular, we are aware that experiences of having breast cancer are shaped by distinct sociocultural discourses, and therefore our findings must be interpreted within this broader context. Other conditions may have certain characteristics that shape collective identities, influencing the content and nature of online interactions. We hope to explore this in future research. However, we suggest that the theoretical framework used to interpret our findings here has wider

XSL•FO
RenderX

applicability to other online health forums where trust is of significance.

Implications for Further Research

Further research could examine how trust operates in relation to what is shared in the online environment—for example, whether it is factual knowledge and how important this is to personal health and well-being and the personal and experiential nature of information and emotions that are shared. It is important to understand the extent to which the person sharing these different types of information is placing trust in the forum users as well as the trust that people place in the information. What is clear from our research is that trust and how it appears in online environments is highly contextual and content specific, and so further research is needed to explore how the process of trust may change on different forums with different demographic groups. Future research could also examine in greater detail the role of moderators in shaping trust in online spaces. Finally, new technologies and platforms are likely to change how trust manifests, so researchers should be responsive to technological change and how this might affect trust and sharing online.

Conclusion

This study contributes new knowledge to the underresearched area of how trust forms and develops on online health forums. Our findings show that the development of trust is a process which is influenced by structural features of the forum, and informal but collectively understood relational interactions between forum users. It is also apparent that trust changes over time. We suggest that this 3-dimensional framework of trust could be applied to other studies of trust in online health settings. Our findings are of value to organizations hosting online health forums that seek to develop a better understanding of what promotes trust and facilitates sharing in online environments.

Acknowledgments

We thank BCC for their support in allowing us to use their online forum within our study and for enabling us to recruit forum users to take part in the interviews. We are grateful to the people who took part in the interviews. We are grateful to colleagues Professor Nigel Ford and Dr. Andrew Cox, Information School, University of Sheffield, for helpful feedback on an earlier draft of this paper and to the 3 reviewers who provided helpful feedback that enabled us to improve the paper.

The "A Shared Space and A Space for Sharing" project is one of several projects funded through the EMoTICON network (grant number ES/M00354X/1), which is funded through the following UK cross-council programs: Partnership for Conflict, Crime, and Security Research (led by the Economic and Social Research Council); Connected Communities (led by the Arts and Humanities Research Council); and Digital Economy (led by the Engineering and Physical Sciences Research Council) in partnership with the UK Defence Science and Technology Laboratory and the Centre for the Protection of National Infrastructure.

Conflicts of Interest

None declared.

References

1. Coulson, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007 Oct;68(2):173-178. [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
2. Gill PS, Whisnant B. A qualitative assessment of an online support community for ovarian cancer patients. Patient Relat Outcome Meas 2012;3:51-58 [FREE Full text] [doi: 10.2147/PROM.S36034] [Medline: 23185122]
3. Sillence E. Giving and receiving peer advice in an online breast cancer support group. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):480-485. [doi: 10.1089/cyber.2013.1512] [Medline: 23659724]
4. Feng J, Lazar J, Preece J. Empathy and online interpersonal trust: a fragile relationship. Behav Inform Technol 2004;23(2):97-106.
5. Zhau J. Trust, empathy, social identity, and contribution of knowledge within patient online communities. Behav Inform Technol 2013;32(10):1041-1048.
6. Allen C, Vassilev I, Kennedy A, Rogers A. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
7. Haciyakupoglu G, Zhang W. Social media and trust during the Gezi protests in Turkey. J Comput-Mediated Commun 2015;20(4):450-466.
8. Lankton NK, NKMcKnight D. What does it mean to trust facebook? Examining technology and interpersonal trust beliefs. SIGMIS Database 2011;42(2):32-54.
9. Mesch GS. Is online trust and trust in social institutions associated with online disclosure of identifiable information online? Comput Hum Behav 2012;28(4):1471-1477.
10. McPherson AC, Gofine ML, Stinson J. Seeing is believing? A mixed-methods study exploring the quality and perceived trustworthiness of online information about chronic conditions aimed at children and young people. Health Commun 2014;29(5):473-482. [doi: 10.1080/10410236.2013.768325] [Medline: 24099647]

XSL•FO
RenderX

11. Leimeister JM, Ebner W, Krcmar H. Design, implementation, and evaluation of trust-supporting components in virtual communities for patients. J Manage Inform Syst 2005;21(4):101-135.
12. Bansal G, Zahedi F, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010;49(2):138-150.
13. Chiu CM, Hsu M, Wang E. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. Decis Support Syst 2006;42(3):1872-1888.
14. Khodyakov D. Trust as a process: a three-dimensional approach. Sociology 2007;41(1):115-132.
15. Radin P. “To me, it's my life”: medical communication, trust, and activism in cyberspace. Soc Sci Med 2006 Feb;62(3):591-601. [doi: 10.1016/j.socscimed.2005.06.022] [Medline: 16039031]
16. Fan H, Lederman R, Smith SP, Chang S. How trust is formed in online health communities: a process perspective. Commun Assoc Informat Syst 2014;34(28):531-561.
17. Brownlie J, Howson A. “Leaps of faith” and MMR: an empirical study of trust. Sociology 2005;39(2):221-239.
18. A Shared Space and a Space for Sharing. URL: http://www.space4sharingstudy.org/ [accessed 2017-05-03] [WebCite Cache ID 6qBkelfhq]
19. Breast Cancer Care Forum. URL: https://forum.breastcancercare.org.uk/ [accessed 2017-01-24] [WebCite Cache ID 6nl4kefTk]
20. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101.
21. Sandaunet AG. A space for suffering? Communicating breast cancer in an online self-help context. Qual Health Res 2008;18(12):1631-1641.
22. Rubenstein EL. Rituals of introduction and revolving roles: socialization in an online breast cancer community. Libr Inform Sci Res 2015;37:353-362.
23. Demjén Z. Laughing at cancer: humour, empowerment, solidarity and coping online. J Pragmatics 2016;101:18-30.
24. Dean RA, Gregory DM. Humor and laughter in palliative care: an ethnographic investigation. Palliat Support Care 2004 Jun;2(2):139-148. [Medline: 16594243]
25. Tanay MA, Roberts J, Ream E. Humour in adult cancer care: a concept analysis. J Adv Nurs 2013 Sep;69(9):2131-2140. [doi: 10.1111/jan.12059] [Medline: 23215893]
26. Flanagin AJ, Hocever K, Nancy S. Connecting with the user-generated Web: how group identification impacts online information sharing and evaluation. Inform Commun Soc 2014;17(6):683-694.
27. Choo HY, Ferree MM. Practicing intersectionality in sociological research: a critical analysis of inclusions, interactions, and institutions in the Study of Inequalities. Sociol Theor 2010;28(2):129-149.
28. Hsu MH, Ju TL, Yen CH, Chang CM. Knowledge sharing behavior in virtual communities: the relationship between trust, self-efficacy, and outcome expectations. Int J Hum-Comput Stud 2007;65(2):153-169.
29. Cole J, Watkins C, Kleine D. Health advice from Internet discussion forums: how bad is dangerous? J Med Internet Res 2016 Jan 06;18(1):e4 [FREE Full text] [doi: 10.2196/jmir.5051] [Medline: 26740148]
30. Chen AT. What's in a virtual hug? A transdisciplinary review of methods in online health discussion forum research. Libr Inform Sci Res 2014;36:120-130.
31. Brady E, Segar J, Sanders C. “You get to know the people and whether they're talking sense or not”: negotiating trust on health-related forums. Soc Sci Med 2016 Aug;162:151-157 [FREE Full text] [doi: 10.1016/j.socscimed.2016.06.029] [Medline: 27359321]

Abbreviations

BCC: Breast Cancer Care

Edited by G Eysenbach; submitted 08.02.17; peer-reviewed by P Dalrymple, N Zdenkowski, T Gronlund; comments to author 28.02.17; revised version received 11.04.17; accepted 28.04.17; published 23.05.17.

Please cite as:
Lovatt M, Bath PA, Ellis J
Development of Trust in an Online Breast Cancer Forum: A Qualitative Study
J Med Internet Res 2017;19(5):e175
URL: http://www.jmir.org/2017/5/e175/
doi:10.2196/jmir.7471
PMID:28536093

©Melanie Lovatt, Peter A Bath, Julie Ellis. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Enhancing Comparative Effectiveness Research With Automated Pediatric Pneumonia Detection in a Multi-Institutional Clinical Repository: A PHIS+ Pilot Study

Stephane Meystre[1], MD, PhD; Ramkiran Gouripeddi[2], MBBS, MS; Joel Tieder[3], MPH, MD; Jeffrey Simmons[4], MSc, MD; Rajendu Srivastava[5,6], MPH, MD; Samir Shah[4], MSCE, MD

[1]Medical University of South Carolina, Charleston, SC, United States
[2]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States
[3]Seattle Children's Hospital and University of Washington, Seattle, WA, United States
[4]Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[5]Department of Pediatrics, University of Utah, Salt Lake City, UT, United States
[6]Primary Children's Hospital, Salt Lake City, UT, United States

Corresponding Author:
Stephane Meystre, MD, PhD
Medical University of South Carolina
135 Cannon St, 4th Floor
Charleston, SC,
United States
Phone: 1 843 792 0015
Fax: 1 843 792 5587
Email: meystre@musc.edu

Abstract

Background: Community-acquired pneumonia is a leading cause of pediatric morbidity. Administrative data are often used to conduct comparative effectiveness research (CER) with sufficient sample sizes to enhance detection of important outcomes. However, such studies are prone to misclassification errors because of the variable accuracy of discharge diagnosis codes.

Objective: The aim of this study was to develop an automated, scalable, and accurate method to determine the presence or absence of pneumonia in children using chest imaging reports.

Methods: The multi-institutional PHIS+ clinical repository was developed to support pediatric CER by expanding an administrative database of children's hospitals with detailed clinical data. To develop a scalable approach to find patients with bacterial pneumonia more accurately, we developed a Natural Language Processing (NLP) application to extract relevant information from chest diagnostic imaging reports. Domain experts established a reference standard by manually annotating 282 reports to train and then test the NLP application. Findings of pleural effusion, pulmonary infiltrate, and pneumonia were automatically extracted from the reports and then used to automatically classify whether a report was consistent with bacterial pneumonia.

Results: Compared with the annotated diagnostic imaging reports reference standard, the most accurate implementation of machine learning algorithms in our NLP application allowed extracting relevant findings with a sensitivity of .939 and a positive predictive value of .925. It allowed classifying reports with a sensitivity of .71, a positive predictive value of .86, and a specificity of .962. When compared with each of the domain experts manually annotating these reports, the NLP application allowed for significantly higher sensitivity (.71 vs .527) and similar positive predictive value and specificity .

Conclusions: NLP-based pneumonia information extraction of pediatric diagnostic imaging reports performed better than domain experts in this pilot study. NLP is an efficient method to extract information from a large collection of imaging reports to facilitate CER.

(J Med Internet Res 2017;19(5):e162) doi:10.2196/jmir.6887

KEYWORDS

natural language processing; pneumonia, bacterial; medical informatics; comparative effectiveness research

XSL•FO
RenderX

Introduction

Community-acquired pneumonia (CAP) is a leading cause of hospitalization among children in the United States [1,2]. Despite this prevalence, the effectiveness of common management strategies [3] is unknown. Multicenter studies using administrative data are inexpensive to conduct and could help compare treatment effectiveness and overcome the challenge of measuring adverse outcomes [4,5]. However, these studies are limited by the potential for subject misclassification. International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) discharge diagnosis codes are commonly used to identify patients [4,5]. Improper use of these codes may lead to false positive or false negative cases [6]. In studies of pediatric CAP, this might lead to systematic biasing by inadvertently including patients without pneumonia or excluding patients with pneumonia in the study cohort [7]. Furthermore, use of these discharge diagnosis codes only precludes more accurate risk adjustment than might be available through admission chest radiograph results, for example [8].

The PHIS+ repository augments the Pediatric Health Information System (PHIS), an administrative database from the Children's Hospital Association, with clinical data [9]. PHIS+, consists of laboratory [9] and microbiological testing results [10], as well as imaging reports from 6 pediatric hospitals across multiple care settings (inpatient, outpatient, emergency department, and ambulatory surgery) over a 5-year study period. The clinical data in the PHIS+ repository are standardized and harmonized using biomedical terminologies and common data models. But, unlike laboratory results, which are available in discrete formats for comparative effectiveness research analyses, imaging reports are available only in narrative clinical text and lack standardization in structure and format. To allow for efficient and rapid access to these data, we developed a Natural Language Processing (NLP) application to determine the diagnosis of bacterial pneumonia from pediatric diagnostic imaging reports by extracting pneumonia characteristics (ie, presence, symmetry, and size of pleural effusion and pulmonary infiltrate) [11].

NLP has been used to extract different types of clinical information from various sources of narrative text in adult patients [12]. Studies have applied Bayesian networks and NLP to detect bacterial pneumonia in adults [13], and several used an NLP application called MedLEE [14] to extract community-acquired pneumonia severity scores in adults [15] and pneumonia information from chest radiology reports in a neonatal intensive care unit [16], or to identify patients with tuberculosis [17]. Recent efforts applied NLP to extract pneumonia information from radiology reports in an adult intensive care unit [18], detect probable pneumonia cases and help manual chart review [19], and also included electronic health record structured data to detect pneumonia cases [20]. These studies reported accuracy metrics with large variations, sensitivity ranging from .45 to .95, and positive predictive value (PPV) from .075 to .86 (best PPV was .86 with a sensitivity of .75 [18], and best sensitivity was .95 with a PPV of .78 [13]). They typically focused on only one type of clinical note, at only one health care organization or hospital, and included the complete development of large complex NLP systems. Only one of these prior studies included children evaluated for pneumonia [19], but it required a manual review of a subset of the radiology reports already analyzed by the NLP system. A good recent review of NLP applications to radiology reports can be found in [21]. The goal of this study was to develop an automated, scalable, and accurate method to determine the presence or absence of pneumonia in children, using a large variety of chest imaging reports from the newly developed PHIS+ repository in order to facilitate the conduct of adequately powered comparative effectiveness research aimed for treatment options of hospitalized children.

Methods

Study Sites

Six free-standing children's hospitals were included: Boston Children's Hospital (Boston, MA, USA); Children's Hospital of Philadelphia (Philadelphia, PA, USA); Children's Hospital of Pittsburgh (Pittsburgh, PA, USA); Cincinnati Children's Hospital Medical Center (Cincinnati, OH, USA); Primary Children's Hospital, Intermountain Healthcare (Salt Lake City, UT, USA); and Seattle Children's Hospital (Seattle, WA, USA).

Reference Standard Preparation

The imaging procedures from the six contributing hospitals in the PHIS+ repository were already mapped to Current Procedural Terminology (CPT) codes [22]. We first selected relevant chest diagnostic imaging (chest radiograph, computerized tomography, and ultrasound) procedure CPT codes (see Multimedia Appendix 1), and then extracted a stratified random collection of imaging study reports mapped to these CPT codes. One report was extracted for each randomly selected patient. A preliminary power analysis indicated that a selection of 270 imaging reports would allow a 95% CI of ±4% width with an expected sensitivity of 90%, assuming mention of pneumonia in 25% of the reports (pneumonia is the information we extracted mentioned the least frequently). A total of 282 reports were eventually selected, deidentified using De-ID software (DE-ID Data Corp) [23] and provided as plain text files for NLP-based information extraction.

Reference Standard Annotation

The 282 deidentified diagnostic imaging reports were annotated by domain experts to evaluate the pneumonia information extraction application. Annotations included all mentions of pulmonary infiltrate, their local context (eg, negation, as in "no infiltrate"), and their symmetry (ie, unilateral or bilateral); pleural effusions, their local context, and their size (ie, small or moderate or large); mentions of pneumonia and their local context (eg, "consistent with pneumonia" or "no evidence of pneumonia"); and whether the report supported the diagnosis of bacterial pneumonia (Figure 1).

The domain experts, three attending pediatric hospital medicine physicians, were trained while also iteratively refining the annotation instructions on the basis of their experience. They first annotated a set of 15 reports, with low interannotator agreement. Examples of disagreements between domain experts are listed in Figure 2.

XSL•FO
RenderX

After having discussed disagreements and updated the annotation instructions, they annotated a second set of 10 other reports and reached fair agreement (pairwise proportions of agreement: .65-.78 for infiltrates, .12-.7 for effusions, and .43-.74 for mentions of pneumonia). Finally, after a final round of disagreement discussions and instructions refinement, they annotated 10 new reports and reached excellent agreement (.96-.98 for infiltrates, .94-1 for effusions, and .92-1 for mentions of pneumonia). The training phase then ended, and annotation of the complete 282 reports collection followed (including reannotation of the initial 15+10+10 reports). At this stage, the rare disagreements were discussed among all domain experts to reach consensus for the reference standard. The annotated information included the following (Figure 1; Final annotation guideline in Multimedia Appendix 2):

- Mentions of "pneumonia" (or synonyms—eg, "pneumonitis"), without adjectives (except if required to define the concept; eg, "lung infection" needs "lung" to be precise enough).
- Mentions of "pleural effusion" (or synonyms—eg, "empyema"; or terms that imply the existence of a pleural effusion if "pleural effusion" or a synonym is not mentioned—eg, "loculation," "free fluid"), without adjectives.
- Mentions of "pulmonary infiltrate" (or synonyms like "opacity," "consolidation"), without adjectives or remote synonyms like "small airways disease," "interstitial markings," "peribronchial thickening," or "atelectasis."
- Context surrounding each pneumonia, effusion, or infiltrate annotation (referred to as "local context") was annotated as *present* (ie, affirmed, not negated, current), *absent* (ie, negated, excluded), *speculative* (ie, hypothetical, a possibility, to rule it out), or *historical* (ie, in the past, not current anymore).
- Pleural effusion size was annotated as *small*, *moderate-large*, or *not mentioned*.
- Symmetry of infiltrates was annotated as *unilateral*, *bilateral*, or *not mentioned*.
- Overall, each report was annotated as to whether it did or did not generally support the diagnosis of bacterial pneumonia (true or false).

Figure 1. Diagnostic imaging report annotations example.

CLINICAL HISTORY: **AGE-year-old male with pneumonia and pleural effusion.
Chest tube in place. Question size of pleural effusion.
COMPARISON: Ultrasound of the chest dated **DATE[Mar 20 2010].
FINDINGS: Compared to the previous examination, the size of the left pleural effusion has decreased, measuring up to approximately 2cm in thickness at the costophrenic sulcus. Consolidated and atelectatic lung is noted adjacent to the small effusion. No definite septations or complexity in the pleural fluid.
IMPRESSION: Decreased size of left pleural effusion, with adjacent consolidation/atelectasis.

Figure 2. Examples of domain expert annotation disagreements.

Topic	Examples	Cause(s)	Solution(s)
Inconsistent term spans	…the possible pneumonia. …the possible pneumonia. …the possible pneumonia.	Lack of precision when selecting text to annotate	• Automate selection of alphanumeric characters.
Effusion terms	…and pleural effusion. …and pleural effusion. …no fluid in pleural spaces. …moderate degree of pericardial effusion[Negated],	Annotation guideline understanding differences	• Annotation guideline clarification. • Discussions among experts.
Pulmonary infiltrate terms	…some small airways disease. …some small airways disease. …with interstitial markings… …with interstitial markings…	Lack of agreement about terms to annotate as "pulmonary infiltrate"	• Annotation guideline clarification. • Discussions among experts.
Mentions of "fluid"	…small amount of fluid. …small amount of fluid. …fluid collection.	Lack of agreement about "fluid" mentions annotation	• Annotation guideline clarification. • Discussions among experts.
Pneumonia mention modifiers	…rule-out pneumonia. …has a history of pneumonia.	Disagreement about Pneumonia annotation (possible or historical?)	• Annotation guideline modifications.

XSL•FO
RenderX

Clinical Information Extraction Application Development

We developed an application based on NLP to automate the extraction of information. This application was based on the Apache UIMA (Unstructured Information Management Architecture) framework [24] using components either developed specifically for this application or adapted from another NLP application: Textractor [25]. Components included text preprocessing (sections detection, lists annotation, sentence segmentation, tokenization, part-of-speech tagging, and chunking), dictionary look-up, local context analysis, annotation attributes and patient information (hospital and patient code) extraction, machine learning features extraction, and the final classification (Figure 3).

During text preprocessing, sections were detected using a collection of regular expressions representing possible headers for patient history sections. Lists were also detected using regular expressions, and their entries segmented as individual sentences. Segmentation of the text in sentences was adapted from Textractor, which is based on a machine learning algorithm (maximum entropy, MaxEnt [26]). Sentences are then "tokenized," split in words or other meaningful groups of alphabetical or numeric characters. Each token is then assigned a part-of-speech tag with another module adapted from Textractor that is based on maximum entropy (itself adapted from OpenNLP [26]). Finally, noun phrase "chunks" are detected with a third module adapted from Textractor, which is also based on maximum entropy (also originally adapted from OpenNLP [26]).

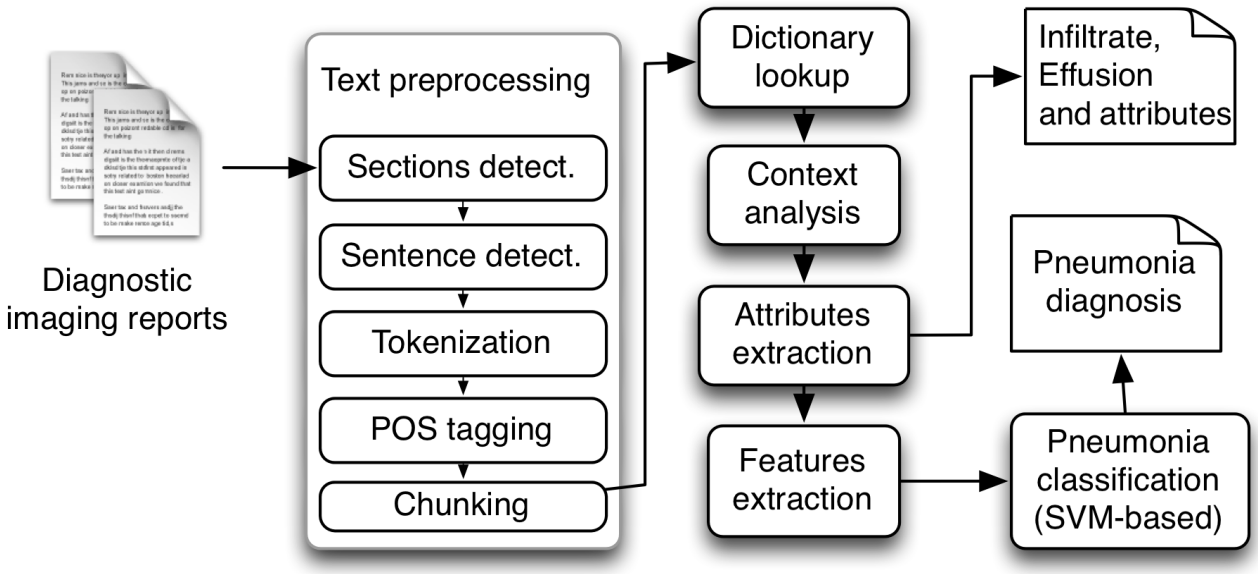
The dictionary lookup module searches a list of terms for matches with the noun phrase "chunks" detected in the text. The list of terms (ie, dictionary) was originally based on a subset of the Unified Medical Language System (UMLS) Metathesaurus [27] filtered by semantic type to include only disease or syndrome, finding, or pathologic function. This dictionary was later replaced with a list of terms built manually by clinicians (based on their domain knowledge), an approach that allowed for improved accuracy.

The local context analysis was based on the ConText algorithm [28], as implemented in Textractor. This algorithm looks for keywords that indicate local context such as negation (eg, denied, no, absent), and then assigns this context to concepts found in a window of words following or preceding the keyword. For example, in the sentence "Findings consistent with viral or reactive airways disease without focal pneumonia," the keyword "without" indicates negation and precedes the annotated concept "pneumonia," which will therefore be considered negated, or absent.

The extraction of annotation attributes (effusion size and infiltrate symmetry) and patient information (hospital and patient code) was based on a set of regular expressions developed specifically and implemented similarly to ConText, assigning these attributes to the appropriate annotated concepts.

Finally, the classification of reports as supporting the diagnosis of bacterial pneumonia (or not) was based on a Support Vector Machine (SVM) classifier with lexical and semantic features. These features included a "bag-of-words" (ie, list of words occurring more than once in our reports collection, without stopwords like "and," "from," "each") and the annotated concepts with their attributes (eg, "pleural effusion" annotation with "small" quantity attribute). The classifier was an implementation of LIBSVM [29], with the radial basis function (RBF) kernel.

Figure 3. Components of the pneumonia clinical information extraction application.



Application Performance Improvements

When initially evaluating the pneumonia classification accuracy, sensitivity was not satisfactory. Therefore, we compared several different machine learning algorithms, refined parameters for the SVM, and filtered the machine learning features (bag-of-words), as well as the dictionaries used by our application.

XSL•FO
RenderX

Machine learning algorithms compared included decision trees, rule learners, naïve Bayes, Bayesian networks, and SVMs, all implemented in the Weka software (version 3.7; University of Waikato, New Zealand) [30]. Features used were the same with each algorithm and included the annotated concepts, and their attributes and local context. Refining the SVM parameters (ie, the penalty parameter *C*, and the radial basis function parameter gamma; final values allowing for best accuracy: *C*=11.5, gamma=.1) consisted in realizing a grid search for selecting the best values of these parameters (using the Grid Parameter Search tool available with LIBSVM).

The "bag-of-words" is an important set of features for machine learning, and the original version included 2103 different words. Even after excluding stopwords, most remaining words have no meaning associated with the diagnosis or radiological signs of pneumonia. To focus our classification on more meaningful words for our task, we manually reviewed all words in the initial bag-of-words (named BOW0) and created three versions with increasing levels of domain specificity. The first refined bag-of-words (BOW1) included 99 words, the second (more specific) bag-of-words (BOW2) included 37 words, and the third (most specific) bag-of-words (BOW3) included only 23 words. The three refined bag-of-words are listed in Multimedia Appendix 3. All were annotated as unigrams.

Finally, refining our dictionary of terms focused on mentions of pulmonary infiltrate, removing terms that caused many false positive matches, but few correct matches.

Performance Evaluation Approach

We used a cross validation approach with 5 "folds" for training and validation. This approach starts with random partition of our collection of 282 notes into 5 subsets of approximately the same size. Then, one subset is retained for testing and the remaining four subsets are used for training. This process is repeated 5 times (ie, "folds"), with each subset used only once for testing. In each "fold," we compared the information extraction application output with the manual reference standard annotations, and classified each annotation as true positive (application output matches the reference standard), false positive (application output not found in the reference standard), or false negatives (reference standard annotation missed by the application). We also counted true negatives for the overall classification when the reference standard and the application both classified the report as not supporting the diagnosis of bacterial pneumonia. Finally, we used counts of true positives, true negatives, false positives, and false negatives, and computed various accuracy metrics at the end of the whole process (not after each fold and then averaged across folds). Accuracy metrics included sensitivity (ie, recall), positive predictive value (ie, precision), the F_1-measure (a harmonic mean of sensitivity and positive predictive value [31]), and the accuracy (proportion of agreement) of the local context category and the attributes category (effusion size and infiltrate symmetry).

For the concept-level evaluation, application automatic annotations and reference standard manual annotations were compared and considered a match when the annotated text overlapped exactly (except preceding or following white space or punctuation) and the annotated information categories (eg, "Effusion") were the same. For the document-level evaluation, reports were classified as supporting the diagnosis of bacterial pneumonia or not. They were considered a match when their binary classification corresponded to the reference standard classification. For document-level evaluation of domain experts, their initial classification (ie, before adjudication of differences between annotators and reference standard development) were compared with final reference standard classifications.

Results

Reference Standard Development

The 282 radiology imaging reports annotated, originated from each of the 6 health care organizations in approximately the same numbers (48 from the Boston Children's Hospital, 48 from the Children's Hospital of Philadelphia, 47 from the Children's Hospital of Pittsburgh, 48 from the Cincinnati Children's Hospital Medical Center, 47 from the Primary Children's Hospital, and 44 from the Seattle Children's Hospital). Annotations included 72 mentions of pneumonia or synonyms (0.255 per report on average), 312 mentions of pulmonary infiltrate or synonyms (1.106), and 369 mentions of pleural effusion or synonyms (1.309). Among the 282 reports, 24.5% (69/282) supported the diagnosis of bacterial pneumonia. Agreement among annotators for the 247 (282 minus 35 reports used for annotators training) not previously seen imaging reports reached 82 of 121 pneumonia mentions (67.8%), 502 of 610 infiltrate mentions (82.3%), and 526 of 670 effusion mentions (78.5%).

Performance at the Concept Level

Concepts evaluated here included the automatic annotations by our application of mentions of pneumonia, pleural effusion, pulmonary infiltrate, and corresponding local context and attributes. The average sensitivity and positive predictive value were approximately 93-94%, with higher accuracy for mentions of pneumonia, and lower accuracy for mentions of pleural effusion (Table 1). The local context was correct in about 92% (65/71) to 94.1% (272/289) of the cases, and the attribute category in about 72.3% (209/289) to 92.5% (321/347) of the cases.

XSL•FO
RenderX

Table 1. Concept level accuracy evaluation results.

Metrics	Terms mentioned in radiology imaging reports			
	Pneumonia	Infiltrate	Effusion	All included terms
True positives	71	289	347	707
False positives	0	20	37	57
False negatives	1	23	22	46
Sensitivity	.986	.926	.940	.939
Positive predictive value	1.000	.935	.904	.925
F_1-measure[a]	.993	.931	.922	.932
Context accuracy	.916	.941	.931	.929
Attribute accuracy	N/A[b]	.723	.925	.824

[a]F_1-measure is a harmonic mean of sensitivity and positive predictive value [31].

[b]N/A: not applicable.

Performance at the Document Level

This classification was evaluated with various configurations of our application. Sensitivity was quite low (.42) with our initial configuration (Table 2), motivating us to experiment with the aforementioned performance improvement approaches.

When using the SVM classifier with all features (ie, concepts with local context and attributes, and bag-of-words), the more specific bag-of-words (BOW2 and BOW3) allowed for higher positive predictive value and specificity, but sensitivity was the highest at .652 with the least filtered bag-of-words (BOW1). The configuration allowing for the highest sensitivity and F_1-measure was based on the least filtered bag-of-words and a refined dictionary (Best system in Table 2).

We also compared different machine learning algorithms with a limited set of features (ie, no bag-of-words as not all algorithms tested could use it). Most of them allowed for higher sensitivity than the SVM algorithm (as implemented in Weka sequential minimal optimization [SMO] [32]), but their positive predictive value was always lower (see Multimedia Appendix 4).

Table 2. Document-level classification results.

Metrics	BOW0[a]	BOW1[b]	BOW2[c]	BOW3[d]	Best system[e] (95% CI)	Domain experts average
True positives	29	45	31	30	49	36
True negatives	207	200	210	209	205	206
False positives	6	13	3	4	8	7
False negatives	40	24	38	39	20	33
Sensitivity	.420	.652	.449	.435	.710 (.683-.737)	.527
Positive predictive value	.829	.776	.912	.882	.860 (.833-.886)	.848
F_1 measure	.556	.709	.602	.583	.778	.650
Specificity	.972	.939	.986	.981	.962 (.951-.974)	.966
Accuracy	.837	.869	.855	.847	.901 (.883-.918)	.862

[a] BOW0: Initial bag-of-words.

[b]BOW1: First refined bag-of-words.

[c]BOW2: Second (more specific) refined bag-of-words.

[d]BOW3: Third (most specific) refined bag-of-words.

[e]BOW1 with refined dictionary.

XSL•FO
RenderX

The decision tree algorithm (pruned C4.5 decision tree [33]) automatically created the decision tree and allowed for a classification F_1-measure of .552 (Figure 4).

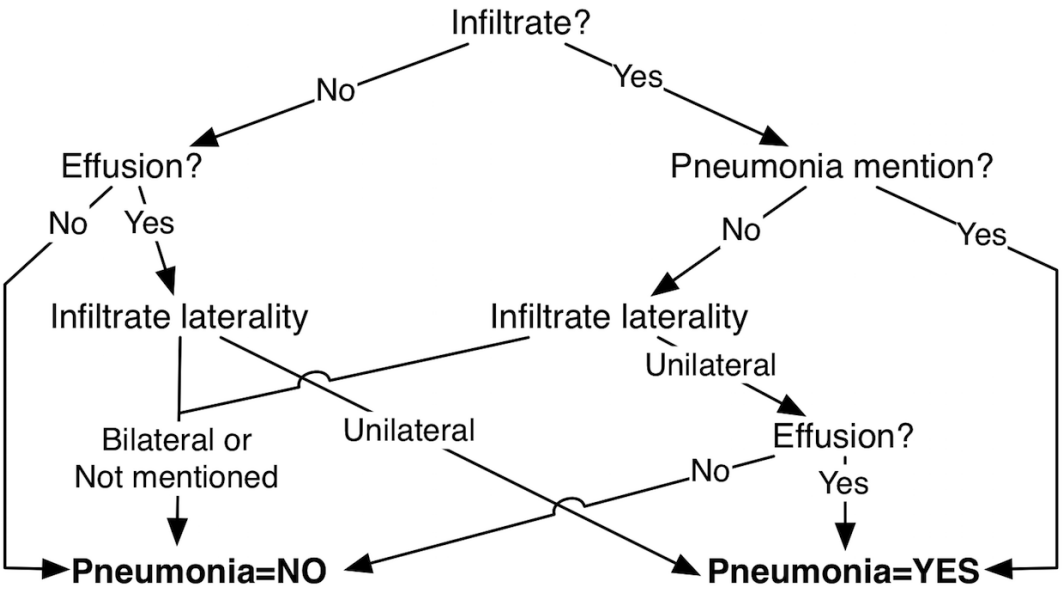
The rule learner (Repeated Incremental Pruning to Produce Error Reduction [RIPPER] [34]) automatically learned three rules that allowed for a classification F_1-measure of .613:

- IF (Effusion=Present) AND (Symmetry=Unilateral) THEN Supports pneumonia=Yes
- IF (Infiltrate=Present) AND (Pneumonia mention=Present) THEN Supports pneumonia=Yes
- OTHERWISE Supports pneumonia=No

The Naïve Bayes algorithm implemented in Weka is based on John and Langley algorithm [35] and the Bayesian network implementation is based on several different algorithms such as Cooper K2 algorithm [36]. The Bayesian network allowed for the highest sensitivity (.739).

In Weka, the SVM implements John Platt's sequential minimal optimization (SMO) algorithm [32]. In our experiment, where the bag-of-words was not part of the features used here, it reached the highest positive predictive value (.811), but also had low sensitivity.

Figure 4. Pruned decision tree for pneumonia classification.



Error Analysis

The most common errors our application made were false negatives, erroneously classifying reports as not supporting the diagnosis of bacterial pneumonia when they actually did support it. Among the 20 false negatives, most were cases of pneumonia that were not as clear, with only 48% of the expert annotators originally agreeing that they were positive cases. This average agreement was 86% for cases that were correctly classified. Most false negatives had no pleural effusion and some had infiltrates mentioned as “airspace disease,” which domain experts specifically decided to exclude as a clear indicator of bacterial pneumonia. Others had pleural effusions mentioned as “fluid” (without the mention of “pleural”), which were difficult to differentiate from other fluid locations in the thorax.

False positive errors (ie, erroneously classifying reports as supporting the diagnosis of bacterial pneumonia when they actually did not support it) were rarer, often caused by local context analysis errors (eg, “pleural effusion has completely resolved” not recognized as an absence of pleural effusion).

Discussion

Principal Findings and Comparison With Prior Work

The most accurate version of our NLP-based pneumonia information extraction application performed better than human domain experts, with significantly higher sensitivity (Fisher exact test, with P=.04.

We found variation in the language used in chest imaging reports both within and across the six children’s hospitals. This was due to inherent differences in imaging modalities, radiologists reporting, and hospital practice. Despite this variability in language, the most accurate version of our NLP-based diagnostic imaging reports classification application eventually reached a sensitivity of .71, positive predictive value of .86, and a specificity of .96. It was based on an SVM classifier with a refined set of features that included a filtered bag-of-words of 99 words, and the annotated concepts with their attributes. When tested in its first version, it only reached a sensitivity of .42.

Experiments to improve classification accuracy included refining the features and parameters used by the SVM classifier, and testing other algorithms. These algorithms included decision trees, rule learners, naïve Bayes, Bayesian networks, and SVMs. They allowed for sensitivity between .42 and .74, positive

XSL•FO
RenderX

predictive value between .66 and .81, and specificity between .88 and .97. Even if the Bayesian network reached a slightly higher sensitivity than the most accurate version of our classifier (.739 vs .71), its positive predictive value was significantly lower (.78 vs .86), and the overall accuracy and F_1-measure were therefore lower. These metrics are consistent with or significantly better than earlier studies such as the extraction of pneumonia information from chest radiology reports in a neonatal intensive care unit by Mendonça and colleagues [16], who reported .71 sensitivity but only .075 positive predictive value, or the extraction of pneumonia findings from chest radiology reports by Fiszman and colleagues [37], who reported .90 positive predictive value but only .34 sensitivity.

The performance reached by the most accurate version of our NLP-based reports classification application may seem low when considering the classification task it performed (ie, classifying diagnostic imaging reports as supporting the diagnosis of bacterial pneumonia or not), but this task was actually more difficult than it may appear. When comparing the three domain experts (ie, attending physicians) annotating these reports with the final reference standard, their average sensitivity was lower than the automatic classifier (Table 2). The positive predictive value and specificity were comparable. This comparison demonstrates the difficulty of the classification task, and the excellent performance of our application when compared with human experts.

Limitations

Our evaluation had several limitations. First, although we had a small sample of annotated diagnostic imaging reports, this sample size allowed for CIs between .023 and .054 only (95% CI; Table 2). This pilot study only included imaging reports from 282 patients, but allowed for sufficiently precise assessment of the accuracy of our system to then apply it to a much larger population of more than 10,000 patients. Comparing our approach with domain experts would benefit from increased precision and could be based on an additional evaluation based on a new larger testing set. Next, the 5-fold cross-validation approach we used only yields meaningful results if the testing set and training set are drawn from the same population, which was our case (both were randomly drawn from our collection of diagnostic imaging reports). Cross-validation could also be misused if selecting features using the complete dataset, and using some data for both training and testing. We avoided both problems by selecting features manually (without examining the dataset, only the experts' domain knowledge), and by ensuring that each report was used only exactly once for testing in our cross-validation approach. The BOW refinement process was purely manual and based on clinical domain knowledge, an approach that would not generalize easily to other applications. Finally, this pilot study was realized on a subset of clinical notes from a unique small population in 6 health care organizations, possibly making additional adaptations required to generalize to a larger population (eg, retraining the machine learning algorithms, refining the dictionaries used).

Conclusions

We developed and used an NLP-based information extraction application to generate discrete and accurate data to identify pediatric patients with CAP. Our main objective was good positive predictive value and improved sensitivity when compared with human domain experts. The pneumonia information extraction application used methods and resources that were trained and evaluated with our reports collection, using a 5-fold cross-validation approach. It allowed for classifying pediatric diagnostic imaging reports with a higher accuracy than that by human domain experts (ie, higher sensitivity and similar positive predictive value and specificity) in this pilot study. After this study, it was used to extract information and classify a much larger collection of diagnostic imaging reports (more than 10,000) in the PHIS+ database, for subsequent community-acquired pneumonia research comparing the effectiveness of different treatment options.

Acknowledgments

This study was approved by the Institutional Review Board of the Children's Hospital of Philadelphia (CHOP), as the primary recipient of the PHIS+ grant funding. A business associates' agreement was used between each hospital and the Children's Hospital Association to authorize sharing of data with identifiers, and a data use agreement governed the sharing of deidentified hospital clinical data. This project was funded under grant number R01 HS019862 from the AHRQ. We thank Ron Keren, MD, MPH, for his advice and leadership of the PHIS+ project. We also thank the Pediatric Research in Inpatient Settings (PRIS) Research Network (www.prisnetwork.org).

Authors' Contributions

SMM conceived the NLP system and led its development. This work was done while he was part of the University of Utah Biomedical Informatics Department. RG was responsible for the data access, preparation, and analysis. JST, JMS, RS, and SSS offered their clinical domain expertise. JST, JMS, and SSS annotated the reference standard. SSS was responsible for the clinical project and evaluation. SMM drafted the initial manuscript. RG, JST, JMS, RS, and SSS provided critical revision of the manuscript. All authors gave the final approval of the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1

Current Procedural Terminology Codes used to select relevant imaging studies.

[PDF File (Adobe PDF File), 213KB - jmir_v19i5e162_app1.pdf]

Multimedia Appendix 2

Annotation guideline.

[PDF File (Adobe PDF File), 49KB - jmir_v19i5e162_app2.pdf]

Multimedia Appendix 3

Refined bag-of-words.

[PDF File (Adobe PDF File), 220KB - jmir_v19i5e162_app3.pdf]

Multimedia Appendix 4

Document level classification accuracy with different machine learning algorithms.

[PDF File (Adobe PDF File), 37KB - jmir_v19i5e162_app4.pdf]

References

1. Lee GE, Lorch SA, Sheffler-Collins S, Kronman MP, Shah SS. National hospitalization trends for pediatric pneumonia and associated complications. Pediatrics 2010 Aug;126(2):204-213 [FREE Full text] [doi: 10.1542/peds.2009-3109] [Medline: 20643717]
2. Keren R, Luan X, Localio R, Hall M, McLeod L, Dai D, Pediatric Research in Inpatient Settings (PRIS) Network. Prioritization of comparative effectiveness research topics in hospital pediatrics. Arch Pediatr Adolesc Med 2012 Dec;166(12):1155-1164. [doi: 10.1001/archpediatrics.2012.1266] [Medline: 23027409]
3. Bradley JS, Byington CL, Shah SS, Alverson B, Carter ER, Harrison C, Pediatric Infectious Diseases Societythe Infectious Diseases Society of America. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. Clin Infect Dis 2011 Oct;53(7):e25-e76. [doi: 10.1093/cid/cir531] [Medline: 21880587]
4. Ambroggio L, Taylor JA, Tabb LP, Newschaffer CJ, Evans AA, Shah SS. Comparative effectiveness of empiric β-lactam monotherapy and β-lactam-macrolide combination therapy in children hospitalized with community-acquired pneumonia. J Pediatr 2012 Dec;161(6):1097-1103. [doi: 10.1016/j.jpeds.2012.06.067] [Medline: 22901738]
5. Williams DJ, Hall M, Shah SS, Parikh K, Tyler A, Neuman MI, et al. Narrow vs broad-spectrum antimicrobial therapy for children hospitalized with pneumonia. Pediatrics 2013 Nov;132(5):e1141-e1148 [FREE Full text] [doi: 10.1542/peds.2013-1614] [Medline: 24167170]
6. Kaafarani HM, Rosen AK. Using administrative data to identify surgical adverse events: an introduction to the Patient Safety Indicators. Am J Surg 2009 Nov;198(5 Suppl):S63-S68. [doi: 10.1016/j.amjsurg.2009.08.008] [Medline: 19874937]
7. Williams DJ, Shah SS, Myers A, Hall M, Auger K, Queen MA, et al. Identifying pediatric community-acquired pneumonia hospitalizations: accuracy of administrative billing codes. JAMA Pediatr 2013 Sep;167(9):851-858 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.186] [Medline: 23896966]
8. McClain L, Hall M, Shah SS, Tieder JS, Myers AL, Auger K, et al. Admission chest radiographs predict illness severity for children hospitalized with pneumonia. J Hosp Med 2014 Sep;9(9):559-564 [FREE Full text] [doi: 10.1002/jhm.2227] [Medline: 24942619]
9. Narus SP, Srivastava R, Gouripeddi R, Livne OE, Mo P, Bickel JP, et al. Federating clinical data from six pediatric hospitals: process and initial results from the PHIS+ Consortium. AMIA Annu Symp Proc 2011;2011:994-1003 [FREE Full text] [Medline: 22195159]
10. Gouripeddi R, Warner PB, Mo P, Levin JE, Srivastava R, Shah SS, et al. Federating clinical data from six pediatric hospitals: process and initial results for microbiology from the PHIS+ consortium. AMIA Annu Symp Proc 2012;2012:281-290 [FREE Full text] [Medline: 23304298]
11. Meystre S, Gouripeddi R, Shah S, Mitchell J. Automatic pediatric pneumonia characteristics extraction from diagnostic imaging reports in a multi-institutional clinical repository. 2013 Presented at: 2013 Joint Summits on Translational Science; March 18-22, 2013; San Francisco.
12. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform 2008:128-144. [Medline: 18660887]
13. Kim W, Wilbur WJ. Corpus-based statistical screening for phrase identification. J Am Med Inform Assoc 2000;7(5):499-511 [FREE Full text] [Medline: 10984469]

XSL•FO
RenderX

14. Friedman C. A broad-coverage natural language processing system. Proc AMIA Symp 2000:270-274 [FREE Full text] [Medline: 11079887]
15. Chapman WW, Haug PJ. Comparing expert systems for identifying chest x-ray reports that support pneumonia. Proc AMIA Symp 1999:216-220 [FREE Full text] [Medline: 10566352]
16. Mendonça EA, Haas J, Shagina L, Larson E, Friedman C. Extracting information on pneumonia in infants using natural language processing of radiology reports. J Biomed Inform 2005 Aug;38(4):314-321 [FREE Full text] [doi: 10.1016/j.jbi.2005.02.003] [Medline: 16084473]
17. Johnson SB, Friedman C. Integrating data from natural language processing into a clinical information system. Proc AMIA Annu Fall Symp 1996:537-541 [FREE Full text] [Medline: 8947724]
18. Liu V, Clark MP, Mendoza M, Saket R, Gardner MN, Turk BJ, et al. Automated identification of pneumonia in chest radiograph reports in critically ill patients. BMC Med Inform Decis Mak 2013 Aug 15;13:90 [FREE Full text] [doi: 10.1186/1472-6947-13-90] [Medline: 23947340]
19. Dublin S, Baldwin E, Walker RL, Christensen LM, Haug PJ, Jackson ML, et al. Natural Language Processing to identify pneumonia from radiology reports. Pharmacoepidemiol Drug Saf 2013 Aug;22(8):834-841 [FREE Full text] [doi: 10.1002/pds.3418] [Medline: 23554109]
20. DeLisle S, Kim B, Deepak J, Siddiqui T, Gundlapalli A, Samore M, et al. Using the electronic medical record to identify community-acquired pneumonia: toward a replicable automated strategy. PLoS One 2013;8(8):e70944 [FREE Full text] [doi: 10.1371/journal.pone.0070944] [Medline: 23967138]
21. Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural language processing technologies in radiology research and clinical applications. Radiographics 2016;36(1):176-191 [FREE Full text] [doi: 10.1148/rg.2016150080] [Medline: 26761536]
22. American Medical Association. AMA-ASSN. CPT - Current Procedural Terminology URL: http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page [accessed 2013-02-15] [WebCite Cache ID 6ESKL24Ap]
23. DE-IDATA. DE-ID Software URL: http://www.de-idata.com/ [accessed 2016-10-25] [WebCite Cache ID 6lWrAN6St]
24. Apache. UIMA (Unstructured Information Management Architecture) URL: http://uima.apache.org/ [accessed 2016-10-25] [WebCite Cache ID 6lWqpTAtM]
25. Meystre SM, Thibault J, Shen S, Hurdle JF, South BR. Textractor: a hybrid system for medications and reason for their prescription extraction from clinical text documents. J Am Med Inform Assoc 2010;17(5):559-562 [FREE Full text] [doi: 10.1136/jamia.2010.004028] [Medline: 20819864]
26. Apache. Welcome to Apache OpenNLP URL: http://opennlp.apache.org [accessed 2014-01-01] [WebCite Cache ID 6MIJrWJJY]
27. Friedman C, Cimino JJ, Johnson SB. A conceptual model for clinical radiology reports. Proc Annu Symp Comput Appl Med Care 1993:829-833 [FREE Full text] [Medline: 8130594]
28. Chapman W, Chu D, Dowling J. ConText: an algorithm for identifying contextual features from clinical text. 2007 Presented at: BioNLP '07 Proceedings of the Workshop on BioNLP 2007; June 29, 2007; Prague, Czech Republic p. 81-88.
29. Chang CC, Lin CJ. NTU CSIE. LIBSVM : a library for support vector machines URL: https://www.csie.ntu.edu.tw/~cjlin/libsvm/ [accessed 2016-10-25] [WebCite Cache ID 6lWr1s08H]
30. Waikato. Weka 3: Data Mining Software in Java URL: http://www.cs.waikato.ac.nz/ml/weka/ [accessed 2017-04-14] [WebCite Cache ID 6piyFTt5G]
31. van Rijsbergen CJ. Openlib. 1979. Information retrieval URL: http://openlib.org/home/krichel/courses/lis618/readings/rijsbergen79_infor_retriev.pdf [accessed 2017-04-26] [WebCite Cache ID 6q0RSjrif]
32. Platt J. Fast training of support vector machines using sequential minimal optimization. In: Advances in kernel methods. Cambridge, MA: MIT Press; 1999:185-210.
33. Quinlan J. C4.5: programs for machine learning. San Francisco, CA: Morgan Kaufmann Publishers Inc; 1993.
34. Cohen WW. Fast Effective Rule Induction. 1995 Presented at: Proceedings of the Twelfth International Conference on Machine Learning; 1995; Tahoe City, CA p. 115-123.
35. John G, Langley P. Estimating continuous distributions in Bayesian classifiers. 1995 Presented at: UAI'95 Proceedings of the Eleventh conference on Uncertainty in artificial intelligence; August 18-20, 1995; Montreal, Canada p. 338-345.
36. Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Machine Learning 1992 Oct;9(4):309-347.
37. Fiszman M, Haug PJ. Using medical language processing to support real-time evaluation of pneumonia guidelines. Proc AMIA Symp 2000:235-239 [FREE Full text] [Medline: 11079880]

Abbreviations

BOW: bag-of-words
CAP: community-acquired pneumonia
CER: comparative effectiveness research

XSL•FO
RenderX

CPT: Current Procedural Terminology
ICD-9-CM: International Classification of Diseases, 9th revision, Clinical Modification
NLP: Natural Language Processing
PHIS+: Pediatric Health Information System, augmented
PPV: positive predictive value
RBF: radial basis function
SVM: Support Vector Machine
UIMA: Unstructured Information Management Architecture

Edited by CL Parra-Calderón; submitted 25.10.16; peer-reviewed by NP Cruz-Díaz, J op den Buijs; comments to author 18.12.16; revised version received 26.01.17; accepted 06.03.17; published 15.05.17.

Please cite as:
Meystre S, Gouripeddi R, Tieder J, Simmons J, Srivastava R, Shah S
Enhancing Comparative Effectiveness Research With Automated Pediatric Pneumonia Detection in a Multi-Institutional Clinical Repository: A PHIS+ Pilot Study
J Med Internet Res 2017;19(5):e162
URL: http://www.jmir.org/2017/5/e162/
doi:10.2196/jmir.6887
PMID:28506958

©Stephane Meystre, Ramkiran Gouripeddi, Joel Tieder, Jeffrey Simmons, Rajendu Srivastava, Samir Shah. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Pharma Websites and "Professionals-Only" Information: The Implications for Patient Trust and Autonomy

Mark Alan Graber[1], MSHCE, MD; Eliyakim Hershkop[2], BA; Rachel Ilana Graber[3], MSW, MA

[1]Departments of Emergeny and Family Medicine, University of Iowa Carver College of Medicine, Iowa CIty, IA, United States

[2]Medical School, Technion Israel Institute of Technology, Hiafa, Israel

[3]Unaffiliated, Washington DC, DC, United States

Corresponding Author:
Mark Alan Graber, MSHCE, MD
Departments of Emergeny and Family Medicine
University of Iowa Carver College of Medicine
200 Hawkins Drive, 1008 RCP Iowa City, Iowa 52242
Iowa CIty, IA,
United States
Phone: 1 319 321 6572
Fax: 1 319 353 7006
Email: mark-graber@uiowa.edu

Abstract

Background: Access to information is critical to a patient's valid exercise of autonomy. One increasingly important source of medical information is the Internet. Individuals often turn to drug company ("pharma") websites to look for drug information.

Objective: The objective of this study was to determine whether there is information on pharma websites that is embargoed: Is there information that is hidden from the patient unless she attests to being a health care provider? We discuss the implications of our findings for health care ethics.

Methods: We reviewed a convenience sample of 40 pharma websites for "professionals-only" areas and determined whether access to those areas was restricted, requiring attestation that the user is a health care professional in the United States.

Results: Of the 40 websites reviewed, 38 had information that was labeled for health care professionals-only. Of these, 24 required the user to certify their status as a health care provider before they were able to access this "hidden" information.

Conclusions: Many pharma websites include information in a "professionals-only" section. Of these, the majority require attestation that the user is a health care professional before they can access the information. This leaves patients with two bad choices: (1) not accessing the information or (2) lying about being a health care professional. Both of these outcomes are unacceptable. In the first instance, the patient's access to information is limited, potentially impairing their health and their ability to make reasonable and well-informed decisions. In the second instance, they may be induced to lie in a medical setting. "Teaching" patients to lie may have adverse consequences for the provider-patient relationship.

(J Med Internet Res 2017;19(5):e178) doi:10.2196/jmir.7164

KEYWORDS

trust; ethics; personal autonomy; readability

Introduction

About 67% of physician office visits involve a prescription medication [1]. This rises to 80% when the visit is to an emergency department [1]. Ideally, the provider should discuss the prescriptions' benefits and side effects with the patient. However, the scheduling realities of single visits restrict patient contact time and the information that can be presented. Unfortunately, providers often do not discuss medication side effects, and so on with patients [2,3].

Inevitably, and reasonably, patients seek out additional information. One source of this information is the Internet, including pharma websites. It is not an exaggeration to say that the use of the Internet to search for medical information is ubiquitous. In 2012, 72% of Internet users looked for medical information on the Web, and this percent has been increasing with time [4-6]. Whereas the quality of Web-based information

XSL•FO
RenderX

varies greatly, unfettered access to information on the Internet is critical to patients trying to educate themselves. In order to make a truly autonomous decision, patients must have whatever information they feel is necessary. Whereas providers may not have the time to review all of the information a patient may want to know, searching the Internet is not subject to a time limitation.

The purpose of this study was to investigate a convenience selection of pharma websites in order to determine whether there is information labeled as for "professionals-only" (eg, not for patients), and what one needs to do in order to access this information. We discuss our findings within the ethical framework of Western medicine, including respect for autonomy and truth telling.

Methods

We examined a convenience sample of 40 pharma websites looking for the presence of a "professionals-only" section. Drugs websites to review were selected based on the year of approval by the Food and Drug Administration (FDA; within the past 3 years), as well as a sample of commonly prescribed medications. There was no attempt to randomize the selection. All of the researchers reviewed each site to determine whether (1) a "professionals-only" section exists, and (2) whether access to this information requires certifying that one is a health care professional.

This project did not require an institutional review board (IRB) submission because no human (or animal) subjects were involved. As noted in the University of Iowa IRB Statement of Compliance, "The University of Iowa Institutional Review Boards are duly constituted with written procedures for initial and continuing review of human subjects research" [7]. This research does not include any human subjects.

Results

Of the 40 websites, 36 (90%) contained information or a link to information that was restricted to professionals only (Table 1). Twenty-four (67%, 24/36) of these 36 sites required an attestation that the user was a US health care professional before the information could be accessed (Table 1). Another 12 (33%, 12/36) of the 36 contained "professionals-only" information that was accessible without such an attestation. Figures 1 and 2 are screenshots that show of the type of verification required in order to access "professionals only" information. Researchers agreed on the categorization of all sites.

Figure 1. An example of a "professionals only" attestation statement from a drug company website.



Figure 2. An example of a "professionals only" attestation statement from a drug company website.



XSL•FO
RenderX

Table 1. Results of website evaluation.

Trade name	Drug name	Website	Embargoed information	Comments
Abilify	Aripiprazole	www.abilify.com	Yes[a]	
Amyvid	Florbetapir F18 injection	www.amyvid.com	Yes	
Anoro Ellipta	Umeclidinium and vilanterol inhalation powder	www.myanoro.com	Yes	
Briinta	Ticagrelor	www.briinta.com	Other[b]	
Brintellix	Vortioxetine	www.brintellix.com	Other	
Corlanor	Ivabradine	www.corlanor.com	Yes	
Cometriq	Cabozantinib	www.cometriq.com	Yes	
Crestor	Rosuvastatin	www.crestor.com	Yes	
Cyramza	Ramucirumab injection	www.cyramzahcp.com	Yes	
Dexilant	Dexlansoparzole	www.dexilant.com	Other	
Duavee	Conjugated estrogens or bazedoxifene	www.duavee.com	Other	
Eliquis	Apixiban	www.eliquis.com/eliquis	Yes	
Entresto	Sacubitril or valasartan	www.entresto.com	Other	
Erivedge	Vismodegib	www.erivedge.com/	Yes	
Farxiga	Dapagliflozin	www.farxiga.com	Yes	
Focalin XR	Dexmethylphenidate	www.focalinxr.com	No[c]	
Gazyva	Obinutuzumab	www.gazyva.com	Yes	
Invokana	Canagliflozon tablets	www.invokana.com	Yes	
Harvoni	Edipasvir and sofosbuvir	www.harvoni.com	Yes	
Kadcyla	Ado-trastuzumab emtansine	www.kadcyla.com	Yes	
Kalydeco	Ivacaftor	www.kalydeco.com	Yes	
Kynamro	Mipomersen sodium	www.kynamro.com	Yes	Requires sign-up
Lipitor	Atorvastatin	www.lipitor.com	No	
Myalept	Metreleptin for injection	www.myalept.com	Yes	
Nesina	Alogliptin	www.nesinahcp.com	Other	
Otezla	Apremilast	www.otelza.com	Other	
Picato	Ingenol mebutate	www.picato.com	Yes	
Pomalyst	Pomalidomide 29	www.pomalyst.com	Yes	
Pradaxa	Dabigitran	www.dabigitran.com	Yes	
Praluent	Alirocumab	www.praluent.com	Yes	
Pristiq	Desvenlofaxine	www.pristiq.com	Other	
Savaysa	Edoxaban	www.savaysa.com	Other	
Sovaldi	Sofosbuvir	www.sovaldi.com	Other	
Stribild	Elvitegravir, cobicistat, emtricitabine, tenofovir, and disoproxil	www.stribild.com	Other	
Tanzeum	Albiglutide	www.tanzeum.com	Yes	
Tecfidera	Dimethyl fumarate	www.tecidera.com	Yes	
Voraxaze	Glucarpidase	www.voraxze.com	Yes	
Vraylar	Cariprazine	www.vraylar.com	Yes	
Vyvanse	Lisdexamfetamine	www.vyvanse.com	Other	
Xarelto	Rivoroxiban	www.xarelto.com	Other	

XSL•FO
RenderX

[a]Yes = embargoed "professionals-only" information requiring attestation.

[b]Other = professionals-only information but no attestation required.

[c]No = no "professionals-only" information.

Discussion

Principal Findings

Restricting access to information as implemented by these websites has several implications. In the first instance, patients may not have access to desired information. Anything that limits a patient's knowledge has the potential to adversely impact autonomy [8,9]. It can be argued that patient-oriented pharma websites contain complete information and already present all of the information that a patient needs in a useable form. Whereas it might be intuited that this solves the problem of information access, patient-oriented pharma websites may present a skewed view of drug risks and benefits and often contain advertising designed to entice the patient [10-12]. One might also argue that the lack of access to "professionals-only" information can be mitigated by the US FDA (the agency responsible for the approval and safety of drug in the United States) having made it a requirement that pharma advertising allow patient access to the package insert, including Web-based advertising [13]. However, the language of packaging information (and medical information on the Web in general) is written at a level that is often incomprehensible to the general public [14-20]. Thus, providing unfettered access to the "professionals-only" information will not in itself solve the problem of lack of accessibility; enhanced readability and a better presentation of information would generally be needed for the information consuming public. This is not an unsurmountable barrier; difficult information can be presented in a manner that is accessible to the general public [19]. We can enhance individuals' ability to make autonomous choices by improving the quality of information presentation.

One can argue that consumers can get the same information elsewhere on the Internet (eg, wikis, sites such as drugs.com). This creates a fundamental dilemma, however. For patients to know what information is missing from their current knowledge base, they first need to access the "professionals-only" information. Thus, the argument that the same information is available elsewhere fails. This may be the case but the uncertainty in patients' minds will continue.

A second, and perhaps more troubling, dilemma arises when there is "professional-only" information on pharma websites that requires an attestation as to one's status as a health care provider. The patient can then either (1) not access the information or (2) lie about their status as a professional. Either outcome is unacceptable. Introducing lying into the therapeutic space can have adverse consequents. Requiring patients to lie in order to obtain information sets a bad precedent. If lying is tacitly accepted as part of the medical system, it has the potential to undermine one of the pillars on which medicine is built; truth telling. Worse still, the system that a patient would like to have faith in is enticing them to lie. Forcing patients to lie may challenge one's faith in the truthfulness of other aspects of care and the medical system [21,22].

Finally, "professionals-only" information can suggest that health care professionals have some dark, mysterious secrets to be hidden from the public. Patients might legitimately wonder what it is that we health care professionals have to hide. Such concerns might further undermine trust in the medical system. Such distrust, whether generated by the perception that there is information hidden from the patient or because of the requirement to lie in the medical sphere, translates into worse patient outcomes [23,24].

It is true that patients can become overwhelmed with information. One solution is to have providers act as interpreters of information; this is the model preferred by many patients [21]. This puts the onus on providers; unfortunately, providers often do not do a particularly good job of laying out the benefits and harms of therapy either [25]. However, a provider-patient partnership with the provider interpreting information for the patient and directing the patient to reliable Web-based information can be a viable model [26].

Limitations

We did not look at the content of "professionals-only" websites and compare this with the content of the patient-oriented websites. For the purposes of our study, such a comparison is a secondary consideration and would not change the study outcome; our main purpose here is to highlight the ethical issues surrounding "professionals-only" websites needing an attestation. A follow-up study could be designed to examine the issue of content. However, no matter the outcome of such an analysis, the basic dilemma remains; patients have to lie to get to "professionals-only" content or do without this information. If the content of the patient and the professional is the same, why require an attestation which may prompt the patient to lie? If the content of the professionals-only and patient-only websites are different, why are we not providing patients with the same information that we are providing to professionals (albeit presented in a manner appropriate for lay-people)?

A second weakness of our study is that we only looked at a small sample of pharma websites. It is possible, but unlikely, that many other pharma websites do not have embargoed information. Nonetheless, we have shown that there is a potential problem related to embargoed information, lying, and trust.

Conclusions

In conclusion, pharma websites often have a "professionals-only" section where access requires user attestation that she is a health care provider. Three problems follow. First, limited access potentially restricts patient information and therefore autonomy. Second, nonaccess may suggest that professionals have "something to hide." Finally, such a system may "train" or induce patients to lie in other medical interactions. Distrust of the medical system may ensue.

XSL•FO
RenderX

Acknowledgments

The authors wish to thank John Stone MD, PhD for his comments on the paper.

Authors' Contributions

MAG developed the idea, collected the data, and wrote the first draft of the paper. EH and RIG collected the data and contributed to the final manuscript.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Center for Disease Control and Prevention. Atlanta: Centers for Disease Control and Prevention; 2016 Sep 12. Therapeutic drug use URL: https://www.cdc.gov/nchs/fastats/drug-use-therapeutic.htm [accessed 2016-12-14] [WebCite Cache ID 6mki8CbR6]
2. Curtis LM, Mullen RJ, Russell A, Fata A, Bailey SC, Makoul G, et al. An efficacy trial of an electronic health record-based strategy to inform patients on safe medication use: the role of written and spoken communication. Patient Educ Couns 2016 Sep;99(9):1489-1495. [doi: 10.1016/j.pec.2016.07.004] [Medline: 27444235]
3. Ferguson S, Trupin L, Yazdany J, Yelin E, Barton J, Katz P. Who receives contraception counseling when starting new lupus medications? The potential roles of race, ethnicity, disease activity, and quality of communication. Lupus 2016 Jan;25(1):12-17 [FREE Full text] [doi: 10.1177/0961203315596079] [Medline: 26190169]
4. Pew Research Center. 2013 Dec 16. Internet, science and tech, health fact sheet URL: http://www.pewinternet.org/fact-sheets/health-fact-sheet/ [accessed 2016-12-14] [WebCite Cache ID 6mkiGE4mJ]
5. Kowalski C, Kahana E, Kuhr K, Ansmann L, Pfaff H. Changes over time in the utilization of disease-related Internet information in newly diagnosed breast cancer patients 2007 to 2013. J Med Internet Res 2014 Aug 26;16(8):e195 [FREE Full text] [doi: 10.2196/jmir.3289] [Medline: 25158744]
6. AlGhamdi KM, Moussa NA. Internet use by the public to search for health-related information. Int J Med Inform 2012 Jun;81(6):363-373. [doi: 10.1016/j.ijmedinf.2011.12.004] [Medline: 22217800]
7. The University of Iowa. Human Subjects Office URL: https://hso.research.uiowa.edu/institutional-review-boards-irbs [accessed 2017-03-11] [WebCite Cache ID 6otDzvNtT]
8. Beauchamp TL, Childress JF. Autonomy. In: Principles of Biomedical Ethics, 6th edition. New York: Oxford University Press; 2009:99-148.
9. Smets EM, Deveugele M, Kripalani S, Cameron KA. New insights on information provision, decision making and patient autonomy. Patient Educ Couns 2016 Sep;99(9):1439-1440. [doi: 10.1016/j.pec.2016.08.008] [Medline: 27578368]
10. Charbonneau DH. An analysis of benefits and risk information on pharmaceutical web sites for the treatment of menopause. Health Info Libr J 2013 Sep;30(3):212-219 [FREE Full text] [doi: 10.1111/hir.12024] [Medline: 23981022]
11. Waack K, Ernst M, Graber M. Informational content of official pharmaceutical industry web sites about treatments for erectile dysfunction. Ann Pharmacotherapy 2004;38:26-34. [doi: 10.1345/aph.1E251]
12. Graber MA, Weckmann M. Pharmaceutical company internet sites as sources of information about antidepressant medications. CNS Drugs 2002;16(6):419-423. [Medline: 12027787]
13. US Department of Health and Human Services. FDA. 1999 Aug. Consumer-Directed Broadcast Advertisements Guidance for Industry - Consumer-Directed Broadcast Advertisements URL: http://www.fda.gov/RegulatoryInformation/Guidances/ucm125039.htm [accessed 2016-12-14] [WebCite Cache ID 6mkiOI9rl]
14. Trivedi H, Trivedi A, Hannan MF. Readability and comprehensibility of over-the-counter medication labels. Ren Fail 2014 Apr;36(3):473-477. [doi: 10.3109/0886022X.2013.872571] [Medline: 24456020]
15. Basara LR, Juergens JP. Patient package insert readability and design. Am Pharm 1994 Aug;NS34(8):48-53. [Medline: 7942501]
16. Wallace LS, Keenum AJ, Roskos SE, Blake GH, Colwell ST, Weiss BD. Suitability and readability of consumer medical information accompanying prescription medication samples. Patient Educ Couns 2008 Mar;70(3):420-425. [doi: 10.1016/j.pec.2007.11.017] [Medline: 18178363]
17. Wallace LS, Roskos SE, Weiss BD. Readability characteristics of consumer medication information for asthma inhalation devices. J Asthma 2006;43(5):375-378. [doi: 10.1080/02770900600709856] [Medline: 16801142]
18. Graber MA, Roller CM, Kaeble B. Readability levels of patient education material on the World Wide Web. J Fam Pract 1999 Jan;48(1):58-61. [Medline: 9934385]
19. Graber MA, D'Alessandro DM, Johnson-West J. Reading level of privacy policies on Internet health Web sites. J Fam Pract 2002 Jul;51(7):642-645. [Medline: 12160504]

XSL•FO
RenderX

20. Shrank WH, Avorn J. Educating patients about their medications: the potential and limitations of written drug information. Health Aff (Millwood) 2007 May;26(3):731-740 [FREE Full text] [doi: 10.1377/hlthaff.26.3.731] [Medline: 17485751]
21. Goold SD. Trust and the ethics of health care institutions. Hastings Cent Rep 2001;31(6):26-33. [Medline: 12945452]
22. Eyal N. Using informed consent to save trust. J Med Ethics 2014 Jul;40(7):437-444. [Medline: 25180355]
23. Tran DQ, Barry V, Antun A, Ribeiro M, Stein S, Kempton CL. Physician trust and depression influence adherence to factor replacement: a single-centre cross-sectional study. Haemophilia 2016 Sep 30;23(1):98-104. [doi: 10.1111/hae.13078] [Medline: 27686244]
24. Armstrong K, Rose A, Peters N, Long JA, McMurphy S, Shea JA. Distrust of the health care system and self-reported health in the United States. J Gen Intern Med 2006 Apr;21(4):292-297 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00396.x] [Medline: 16686803]
25. Tan K, Petrie KJ, Faasse K, Bolland MJ, Grey A. Unhelpful information about adverse drug reactions. BMJ 2014 Aug 12;349:g5019. [Medline: 25115975]
26. 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]

Abbreviations

US FDA: United States Food and Drug Administration

Edited by G Eysenbach; submitted 14.12.16; peer-reviewed by I Adeleke, JW Li; comments to author 11.02.17; revised version received 11.03.17; accepted 06.04.17; published 24.05.17.

Please cite as:
Graber MA, Hershkop E, Graber RI
Pharma Websites and "Professionals-Only" Information: The Implications for Patient Trust and Autonomy
J Med Internet Res 2017;19(5):e178
URL: http://www.jmir.org/2017/5/e178/
doi:10.2196/jmir.7164
PMID:28539303

©Mark Alan Graber, Eliyakim Hershkop, Rachel Ilana Graber. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adherence to AHA Guidelines When Adapted for Augmented Reality Glasses for Assisted Pediatric Cardiopulmonary Resuscitation: A Randomized Controlled Trial

Johan N Siebert[1*], MD; Frederic Ehrler[2*], PhD; Alain Gervaix[1], MD; Kevin Haddad[1], RN; Laurence Lacroix[1], MD; Philippe Schrurs[3], MD; Ayhan Sahin[3], MD; Christian Lovis[2], MPH, FACMI, MD; Sergio Manzano[1], MD

[1]Geneva Children's Hospital, Department of Pediatric Emergency Medicine, University Hospitals of Geneva, Geneva, Switzerland

[2]Division of Medical Information Sciences, Department of Radiology and Medical Informatics, University Hospitals of Geneva, Geneva, Switzerland

[3]Geneva Medical Center, University Hospitals of Geneva, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
Johan N Siebert, MD
Geneva Children's Hospital
Department of Pediatric Emergency Medicine
University Hospitals of Geneva
Avenue de la Roseraie, 47
Geneva, 1205
Switzerland
Phone: 41 795534072
Fax: 41 223725423
Email: Johan.Siebert@hcuge.ch

Abstract

Background: The American Heart Association (AHA) guidelines for cardiopulmonary resuscitation (CPR) are nowadays recognized as the world's most authoritative resuscitation guidelines. Adherence to these guidelines optimizes the management of critically ill patients and increases their chances of survival after cardiac arrest. Despite their availability, suboptimal quality of CPR is still common. Currently, the median hospital survival rate after pediatric in-hospital cardiac arrest is 36%, whereas it falls below 10% for out-of-hospital cardiac arrest. Among emerging information technologies and devices able to support caregivers during resuscitation and increase adherence to AHA guidelines, augmented reality (AR) glasses have not yet been assessed. In order to assess their potential, we adapted AHA Pediatric Advanced Life Support (PALS) guidelines for AR glasses.

Objective: The study aimed to determine whether adapting AHA guidelines for AR glasses increased adherence by reducing deviation and time to initiation of critical life-saving maneuvers during pediatric CPR when compared with the use of PALS pocket reference cards.

Methods: We conducted a randomized controlled trial with two parallel groups of voluntary pediatric residents, comparing AR glasses to PALS pocket reference cards during a simulation-based pediatric cardiac arrest scenario—pulseless ventricular tachycardia (pVT). The primary outcome was the elapsed time in seconds in each allocation group, from onset of pVT to the first defibrillation attempt. Secondary outcomes were time elapsed to (1) initiation of chest compression, (2) subsequent defibrillation attempts, and (3) administration of drugs, as well as the time intervals between defibrillation attempts and drug doses, shock doses, and number of shocks. All these outcomes were assessed for deviation from AHA guidelines.

Results: Twenty residents were randomized into 2 groups. Time to first defibrillation attempt (mean: 146 s) and adherence to AHA guidelines in terms of time to other critical resuscitation endpoints and drug dose delivery were not improved using AR glasses. However, errors and deviations were significantly reduced in terms of defibrillation doses when compared with the use of the PALS pocket reference cards. In a total of 40 defibrillation attempts, residents not wearing AR glasses used wrong doses in 65% (26/40) of cases, including 21 shock overdoses >100 J, for a cumulative defibrillation dose of 18.7 Joules per kg. These errors were reduced by 53% (21/40, P<.001) and cumulative defibrillation dose by 37% (5.14/14, P=.001) with AR glasses.

Conclusions: AR glasses did not decrease time to first defibrillation attempt and other critical resuscitation endpoints when compared with PALS pocket cards. However, they improved adherence and performance among residents in terms of administering the defibrillation doses set by AHA.

XSL•FO
RenderX

(J Med Internet Res 2017;19(5):e183) doi:10.2196/jmir.7379

KEYWORDS

resuscitation; emergency medicine; pediatrics; biomedical technologies; equipment and supplies; eyeglasses

Introduction

Clinical practice guidelines aim to improve quality of care, reduce variation of practice, and provide evidence-based health care [1]. The American Heart Association (AHA) guidelines for cardiopulmonary resuscitation (CPR) are nowadays recognized as the world's most authoritative resuscitation guidelines [2,3]. They are evidence-based, synthesized by experts, and include a large number of algorithms intended to provide step-by-step processes to various life-threatening emergency situations in a systematic fashion. These algorithms are also summarized on pocket reference cards in order to be used as quick reference tools that emergency physicians may have access to during resuscitations. However, despite their availability, suboptimal quality of resuscitation is still common for both adult and pediatric patients [4]. Immediate (level 1) triage represents 175,000 patient visits every year in US pediatric emergency departments (PED) [5]. Among them, 5800 to 10,000 cases are due to in-hospital cardiac arrest (IHCA) [6,7], and 6700 to 15,000 cases to out-of-hospital cardiac arrest (OHCA) [8-10], including 6000 related to non-traumatic causes [11]. Quality CPR with adherence to AHA resuscitation guidelines optimizes the management of critically ill patients and increases their chances of survival [12,13], whereas deviation is associated with decreased likelihood of survival from cardiac arrest (CA) [14]. Currently, the median hospital survival rate from pediatric IHCA is 36% [4], whereas it is below 10% for OHCA [15,16].

As a result, the scientific community has proposed new resuscitation strategies relying on information technologies and devices aiming at improving and ensuring adherence to AHA guidelines [17-20]. Among possible emerging information technologies that could support caregivers, augmented reality (AR) glasses have recently gained a great deal of interest within the scientific community. AR glasses are wearable and connected devices that display interactive images to the visual field of users by overlaying visual information without significantly disturbing the ordinary vision. They feature some functionalities similar to those offered by smartphones and tablets by running self-contained mobile apps. Despite recent communications and studies related to the use of these glasses in various medical fields [21-26], their contribution to resuscitation in emergency medicine has not yet been investigated. Their potential to wirelessly display and interact in real time conditions with data generated by the actions performed in a specific environment could be of great interest for assisting caregivers in resuscitation rooms, while freeing users' hands and allowing them to "see the scene through the screen". Augmented reality might bring useful information to caregivers' attention without disturbing the care process, by allowing constant access to the resuscitation scene. In a previous article, we have described the adaptation of AHA Pediatric Advanced Life Support (PALS) algorithms for AR glasses [27]. This study aimed at investigating whether this adaptation for AR glasses would increase adherence to AHA guidelines by reducing deviation and time to initiation of critical life-saving maneuvers during pediatric CPR when compared with the use of PALS pocket reference cards, in a simulated model.

Methods

Study Design

We conducted a prospective, randomized controlled trial with 2 parallel groups of voluntary pediatric residents comparing time to first defibrillation attempt while using AR glasses (Google Glass, allocation group A) or AHA PALS conventional pocket reference cards (allocation group B) during a standardized simulation-based pediatric CA scenario (Multimedia Appendix 1). No changes were made to the AR glasses or the intervention during the study.

Selection of Participants

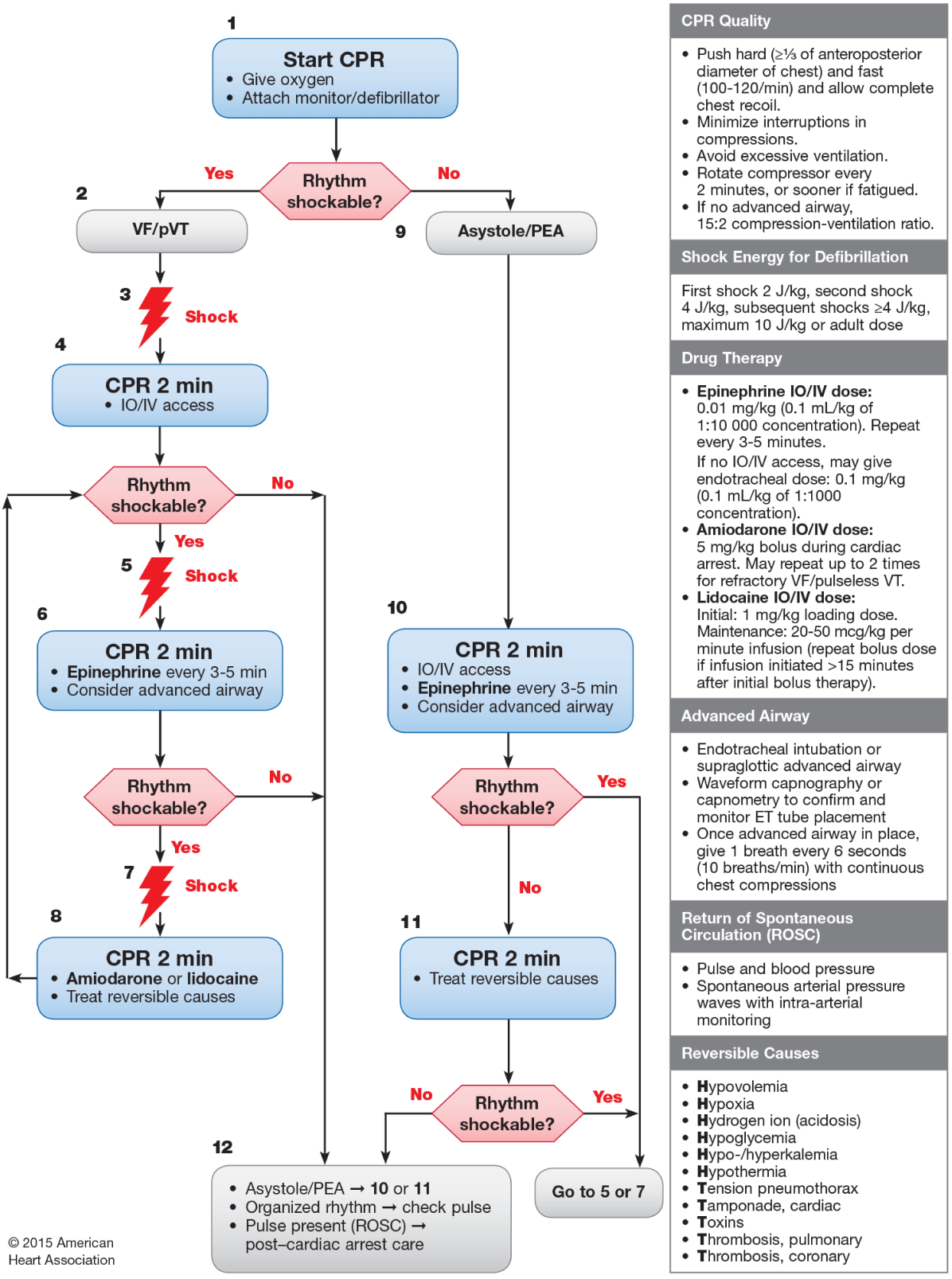
Any physician from the whole pediatric department actively training for a pediatric specialty (residents) was eligible. Residents with eye disorders were not included in the study. Shift-working residents were randomly recruited on the day of the study from a random alphabetical list. Written informed consent was obtained from all the participants before their voluntary involvement. Study participants were not involved in the study design, choice of outcome measures, or the execution of the study. No participants were asked for advice on interpretation or the writing of results. The results of the study were offered to the study participants after the completion of the study.

Setting and Resuscitation Scenario

The study was conducted in a pediatric emergency department (PED) of a tertiary hospital with approximately 28,000 visits per year. We created a standardized simulation scenario on a high-fidelity manikin (Laerdal SimJunior). The resuscitation team was composed of the resident participating in the study and 3 nurses to assist with resuscitation through drugs preparation, chest compressions, and bag-valve-mask ventilation, according to the resident's instructions. A certified technician (KH), who was not a member of the resuscitation team, operated the simulator. Except for participating residents, members of the resuscitation team remained unchanged across all scenarios and were the investigators in the study.

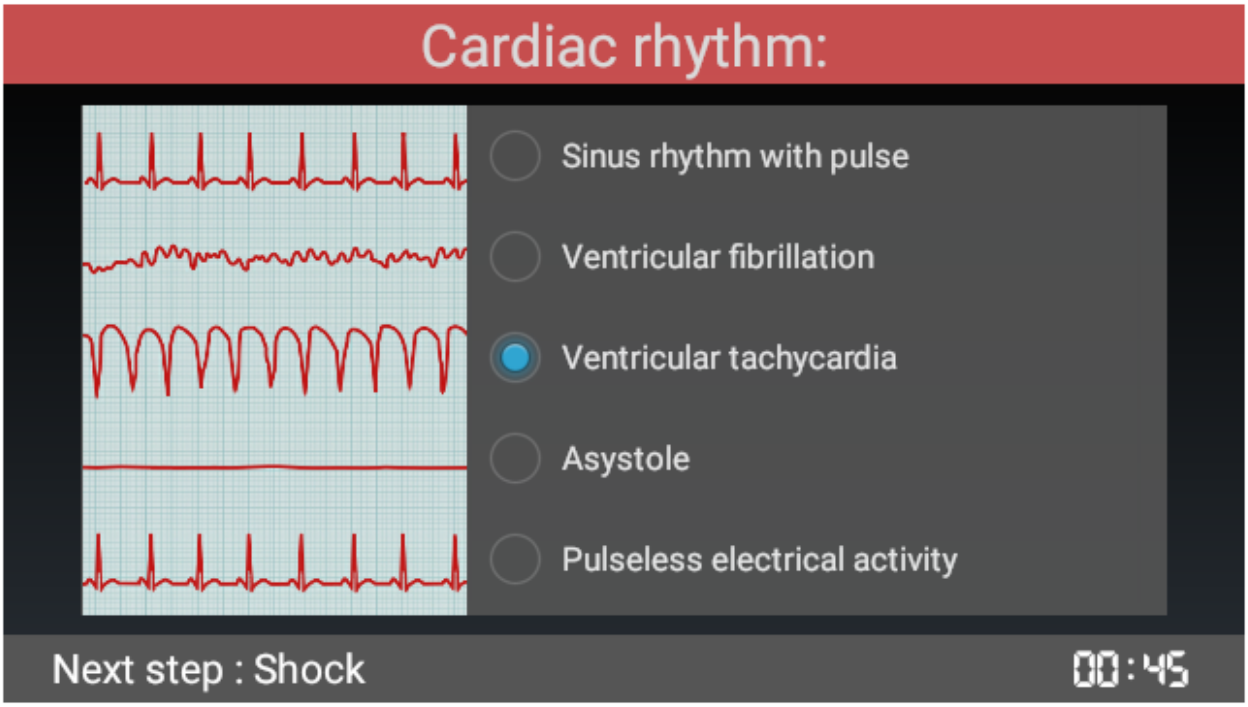
XSL•FO
RenderX

Figure 1. American Heart Association's pediatric cardiac arrest algorithm—2015 update.



XSL•FO
RenderX

Figure 2. Screenshot of a resuscitation step from the American Heart Association's Pediatric Advanced Life Support pulseless ventricular tachycardia algorithm as adapted in augmented reality glasses.



On the day of participation, residents were asked about their demographics. After random allocation, each participant allocated to the AR glasses group received a standardized 15-minute qualifying training session to familiarize them uniformly with the AR glasses. Then, all the participants were asked to perform a 15-minute highly realistic CPR scenario (Multimedia Appendices 2 and 3). Unlike adults, CA in children without prior cardiac disease is mainly due to asystole (40%) and pulseless electrical activity (24%) [28]. Ventricular fibrillation and pulseless ventricular tachycardia (pVT), namely shockable rhythms, are identified in 27% of pediatric IHCA [29]. We decided to study the pVT algorithm because, in our opinion, it offers a greater opportunity to assess the multiple-steps resuscitative skills set by AHA. The scenario was therefore standardized to strictly follow the 2015 AHA pediatric pVT algorithm (Figure 1) [30] and provided on the same manikin. It was conducted in situ in the pediatric resuscitation room of the PED to increase realism. No interactions occurred between participants and investigators. When entering the room, a short clinical statement to recognize the life-threatening condition of the patient, including his weight and age, was given to the resident. The resident was then asked to start the timed scenario and had to recognize by himself or herself the previously settled cardiac rhythm (pVT) with the AR glasses (allocation group A) or the PALS conventional pocket reference card (group B). All participants in allocation group B were required to have their PALS pocket card in their hands throughout the entire scenario. Whether they referred to it or not was left to them, as in real life. The scenario ran invariably until the manikin was defibrillated at the 4th shock and showed a subsequent return of spontaneous circulation (ROSC). In order to be consistent with the 2015 AHA pediatric CA algorithm [30] and to standardize our scenario, defibrillation doses of 2 Joules per kg for the first attempt, and 4 Joules per kg for subsequent 2nd, 3rd, and 4th attempts were expected (Figure 1).

Intervention

As previously described [27], the numerous steps of AHA PALS algorithms were wisely split into "cards." Each card transposed to the AR glasses paralleled the informational content of a resuscitation step from the original algorithm. However, the informational content was "augmented," thanks to the interactivity and display capability of the device. The AR algorithm thus obtained was set up in a manner similar to the PALS pocket references regarding the progression and sequence of actions along the original algorithm's sequences. For instance, the complete pulseless ventricular tachycardia (pVT) algorithm was adapted on 42 cards designed to be as concise as possible without hindering proper progression along the algorithm. The cards were tailored to the small size of the Google glass screen, following a user-centered and ergonomic-driven approach. Each card was structured on 4 zones: (1) a color-coded title allowing direct identification of each step in progress, (2) an image on the left helping with decision-making (such as distinctive illustration of cardiac rhythms), (3) a menu choice on the right helping to progress in the resuscitation steps, and (4) a footer to preview the next step (Figure 2). Interaction was also defined with end users. Tactile commands of the glasses were favored over voice commands due to the inability of AR glasses to distinguish between vocal orders in the noisy environment of a resuscitation room. Swiping up or down allowed navigating inside the card. Selection was done by a click and actions

XSL•FO
RenderX

cancelled by swiping back. Each cycle of chest compression–ventilation was timed, thanks to a countdown clock displayed on the screen. Treatment and defibrillation doses (Philips HeartStart MRx Biphasic Defibrillator) were automatically calculated on patients' weight or age.

Outcome Measures

The elapsed time in seconds in each allocation group from onset of pulseless shockable rhythm to the first defibrillation attempt was selected as the primary outcome, as it is the most important determinant of survival after CA [31]. Secondary outcomes were time elapsed to initiation of chest compression, time to subsequent defibrillation attempts, time to administration of epinephrine and amiodarone, and time interval (in seconds) between defibrillation attempts. AHA recommends 5 cycles of chest compression (about 2 minutes) between defibrillation attempts. The amount of time spent by participants to perform chest compressions by cycles of chest compression was defined as the hands-on time. It was measured in seconds with a chronometer. Drug doses, shock doses, and number of shocks were also assessed. All these outcomes were assessed for deviation from AHA guidelines. At the end of the scenario, a questionnaire using a 10-point Likert scale was submitted to the participant to measure the overall stress perceived during the scenario.

Methods of Measurement and Data Collection

All the actions (ie, the primary and secondary outcomes) performed by the resident during the scenario were automatically recorded and stored by the responsive simulator detectors and by several video cameras. The videos were embedded in a dedicated simulation software, allowing accurate assessment of timing and sequencing of actions. To avoid assessment bias, 2 evaluators then independently reviewed these video recordings. In case of disagreement, a third independent evaluator helped reach a consensus. Data were manually retrieved and entered into a Microsoft Excel spreadsheet (version 2011). Unaccomplished actions were left blank and not assigned any corresponding time. Only residents were assessed and their privacy preserved. Only the investigators of the study had access to the data. The statistical software GraphPad Prism version 6.0h (GraphPad Software, Inc) was used for all data analyses.

Sample Size

The primary objective of the study was to detect a difference in time to the first defibrillation attempt. The sample size was calculated to detect a 30-second decrease in time to first defibrillation between 2 independent groups with a power of 80% and a 2-sided risk alpha of .05. A previous study has shown a mean time to first defibrillation of 92 seconds [32] with a standard deviation (SD) of 23 seconds. Assuming a similar SD in each group in our study, 10 patients per group were required.

Randomization and Blinding

We randomly assigned residents in a 1:1 ratio with a Web-based software [33]. Blinding to the purpose of the study during recruitment was maintained to minimize preparation bias. Participants were unblinded after randomization. Allocation concealment was ensured with the same Web-based software [33] and was not released until the residents started the scenario.

Statistical Analysis

Primary Outcome

We first evaluated the time elapsed between onset of pVT and first defibrillation attempt. The Shapiro-Wilks test was used for normality analysis of the parameters. Means and standard deviations (SDs) with 95% CI were reported. Non-normally distributed variables were analyzed using the Mann-Whitney test. Frequencies were reported as percentages. *T* tests were used to compare independent groups. No paired data were compared. Kaplan–Meier curves for time elapsed between onset of pVT and 1) initiation of chest compression, and 2) first defibrillation, were estimated and compared using the log-rank (Mantel-Cox) test for bivariate survival analysis.

Secondary Outcomes

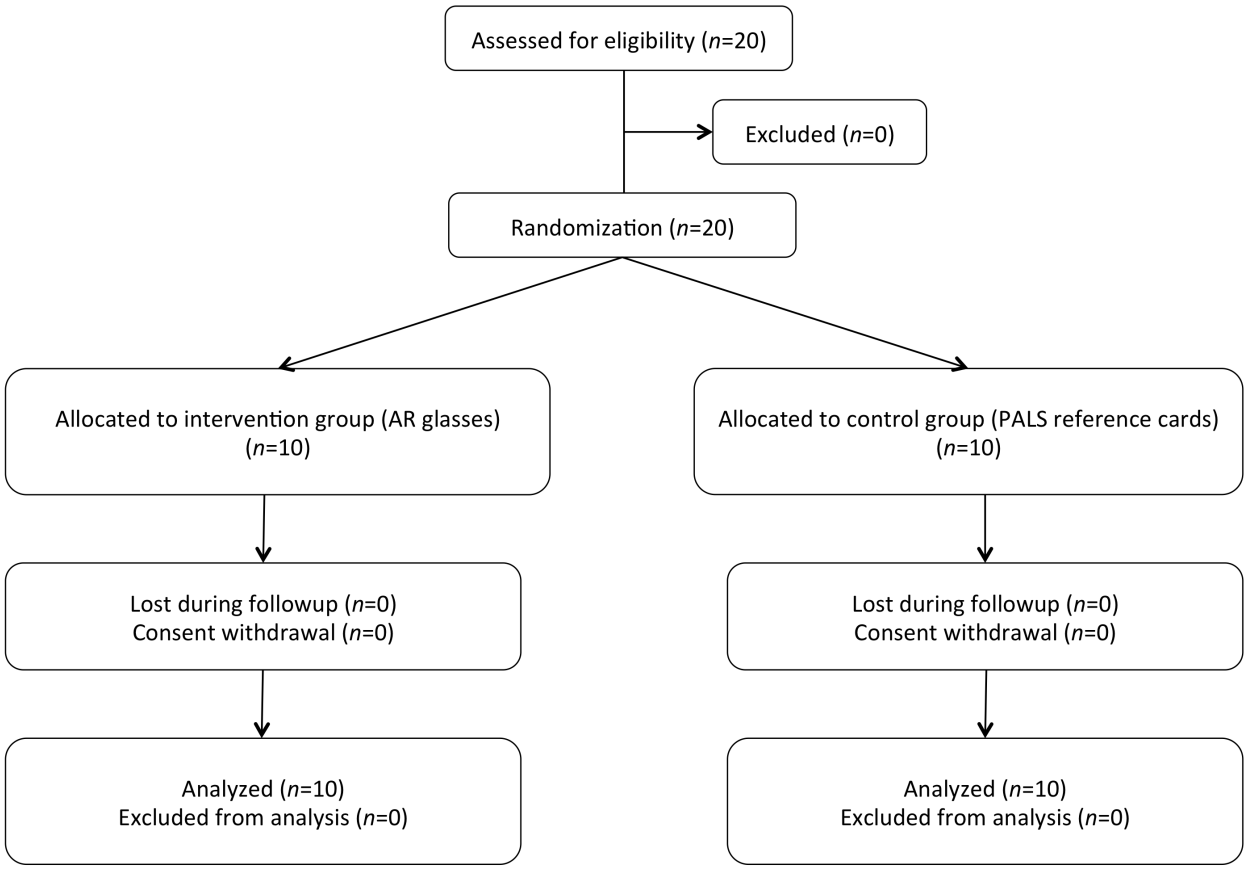
We evaluated the time elapsed between onset of pVT to subsequent defibrillation attempts and drug delivery. As most of the continuous variables were also normally distributed, means and SDs with 95% CI were reported. Non-normally distributed variables were analyzed using a Mann-Whitney test. Frequencies were reported as percentages. *T* tests were used to compare independent groups. No paired data were compared. Errors in cycles of chest compression-ventilation were measured as the deviation in percent from the experimental time spent in seconds compared with the 2-minute duration recommended by AHA guidelines. Wrong defibrillation or drug doses were measured as the deviation in percent from the amount of energy delivered in Joules or drug doses in milliliters compared with AHA recommendations. Wrong defibrillation mode was also measured. Absolute deviations were also analyzed. The mean (SD) difference in deviation obtained with each method was reported with 95% CI. A *t* test for unpaired data was used to compare interventions. Mean differences were reported by randomized group. We also determined if prior certification as a PALS provider before the study, prior resuscitations as a provider in real-life, or post-graduation years (PGY) had a significant impact on the above outcomes. Mean and SD were determined for stress scores of individuals in the questionnaire and reported with descriptive statistics. A *P* value less than .05 was considered significant.

Ethics and Informed Consent

The study was approved by the institutional ethics committee. According to the ICMJE, a registration number was not required for our trial, as the purpose of this study was to examine the effect of the intervention on the providers. Written informed consent was obtained from all participants before their voluntary involvement. The study was conducted in accordance with the principles of the Declaration of Helsinki, the standards of Good Clinical Practice, and Swiss regulatory requirements.

XSL•FO
RenderX

Figure 3. CONSORT flowchart of augmented reality glasses trial.



Results

Study Participants

In March 2016, 20 pediatric residents participated and completed the study with no dropout (Figure 3). The demographic results are summarized in Table 1.

Time to Resuscitation Critical Endpoints

The first defibrillation was delivered within 180 seconds by 70% of the residents in both groups (Figure 4). Mean times to resuscitation critical endpoints are summarized in Table 2. None of them were significantly different between allocation groups A and B. All participants (100%) correctly recognized the pVT rhythm on the monitor. In both allocation groups, 90% of the residents initiated chest compressions within 60 seconds from the onset of pVT, and half of them before 20 seconds. There were no statistically significant differences in hands-on time spent by cycles of chest compression between both groups. Due to a lack of some defibrillation attempts, 1 interval in allocation group A and 4 in group B were not measurable.

At the time of the study, 6 participants (60%) in allocation group A and 7 participants (70%) in allocation group B were PALS-certified providers. Eleven participants out of 20 (55%) were residents with more than 1 year of PGY. With regard to all outcomes measured, we observed in both allocation groups that PALS-certified residents or those with PGY>1 tended to defibrillate and deliver drugs more quickly than non-PALS residents or those with ≤1 PGY (Multimedia Appendix 4). We observed no difference with previous resuscitation experience (data not shown).

XSL•FO
RenderX

Table 1. Participants' demographics and clinical characteristics.

Demographics and clinical characteristics	Randomization Arm	
	AR[a] Glasses (n=10)	PALS[b] pocket cards (n=10)
Age in years, mean (SD)	27.9 (2.6)	29.2 (2.6)
Sex (female), n (%)	9 (90)	6 (60)
Years of residency, mean (SD)	2.4 (1.7)	2.4 (1.7)
Number of residents having been enrolled in > 5 resuscitations in the past, n (%)	4 (40)	3 (30)
Number of PALS providers among residents, n (%)	6 (60)	7 (70)
Number of BLS[c] providers among residents, n (%)	10 (100)	10 (100)
Level of self-confidence in following AHA guidelines (on a scale of 1 to 5), mean (SD)	1.8 (1.2)	2.4 (0.8)

[a]AR: Augmented Reality.

[b]PALS: Pediatric Advanced Life Support.

[c]BLS: Basic Life Support.

Figure 4. Kaplan–Meier curves of proportion of pediatric residents using augmented reality glasses or conventional Pediatric Advanced Life Support pocket cards who a) initiated chest compression, or b) delivered first defibrillation shock during a simulated pulseless ventricular tachycardiapVT scenario (Log-rank test statistic, P=.81 and P=.99).

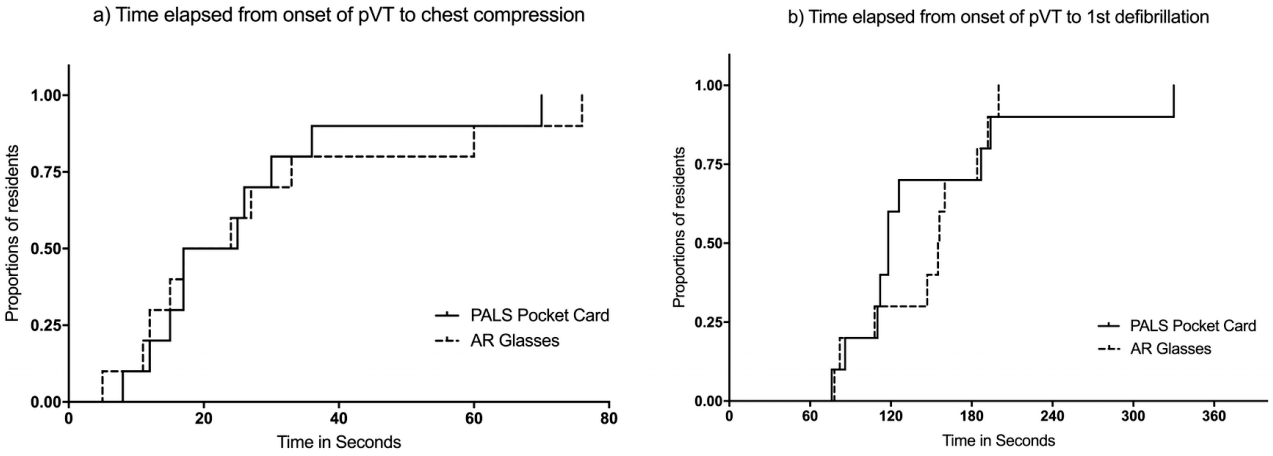


Table 2. Mean time to resuscitation critical endpoints.

Outcomes	AR[a] Glasses Mean (SD, 95% CI)	PALS[b] pocket cards Mean (SD, 95% CI)	Time difference[c]	*P* value
Time to initiation of CPR[d]	28.0 (22.9, 11.6-44.4)	25.6 (17.8, 12.9-38.3)	2.4	.80
Time to 1st defibrillation attempt	146.2 (43.5, 115.1-177.3)	145.7 (75.1, 92.0-199.4)	0.5	.99
Time to 2nd defibrillation attempt	264.0 (73.9, 211.1-316.9)	263.0 (74.2, 210.0-316.0)	1.0	.98
Time to epinephrine	317.3 (62.6, 265.0-369.5)	295.8 (97.7, 220.7-370.9)	21.5	.59
Time to 3rd defibrillation attempt	396.6 (93.6, 329.7-463.5)	389.0 (80.0, 314.8-462.9)	7.7	.86
Time to amiodarone	450.1 (53.6, 408.9-491.3)	492.7 (106.5, 416.5-568.9)	42.6	.28
Time to 4th defibrillation attempt	542.8 (83.3, 478.7-606.8)	526.8 (93.4, 455.0-598.6)	16.0	.71

[a]AR: Augmented Reality.

[b]PALS: Pediatric Advanced Life Support.

[c]Time difference represents absolute time difference between PALS pocket cards and AR Glasses.

[d]CPR: cardiopulmonary resuscitation.

XSL•FO
RenderX

Table 3. Errors and deviations from the American Heart Association's pulseless ventricular tachycardia algorithm.

Critical Resuscitation Endpoints (AHA pVT[a] algorithm)	AHA[b] recommended doses (Joules per kg), (mL per kg)	Randomization Arm			
		AR[c] Glasses Mean (SD, 95% CI)	% deviation (*P* value)	PALS[d] ref. cards Mean (SD, 95% CI)	% deviation (*P* value)[i]
1st defibrillation attempt	2.00	2.12 (0.71, 1.61-2.63)	6 (.60)	3.46 (2.23, 1.87-5.05)	73 (.068)
2nd defibrillation attempt	4.00	3.40 (0.97, 2.71-4.09)	15 (.081)	4.52 (1.53, 3.42-5.62)	13 (.31)
Epinephrine 1:10,000	0.1	0.1 (0, 0.1-0.1)	0 (1.00)	0.1 (0, 0.1-0.1)	0 (1.00)
3rd defibrillation attempt	4.00	4.00 (0, 4.00-4.00)	0 (1.00)	5.00 (1.51, 3.74-6.26)	25 (.10)
Amiodarone	0.1	0.1 (0, 0.1-0.1)	0 (1.00)	0.1 (0, 0.1-0.1)	0 (1.00)
4th defibrillation attempt	4.00	4.00 (0, 4.00-4.00)	0 (1.00)	5.68 (1.18, 4.84-6.52)	42 (.0014)
Cumulative defibrillation dose	14.0	13.52 (0.97, 12.32-14.72)	3.4 (.10)	18.66 (2.03, 13.87-23.45)[e]	33.3 (.025)
Correct AHA sequence		60%		40%[f]	
Correct number of shocks		80%		70%	
Shock overdoses (>100 J)		1/40 opportunities		21/40 opportunities[g]	
Total number of errors		5/40 opportunities		26/40 opportunities[h]	
Defibrillation errors at 1st attempt, n (details)		2 (40% lower, 200% higher)		5 (0.6 Joules per kg to 6 Joules per kg)	
Defibrillation errors at 2nd attempt, n (details)		3 (2 Joules per kg instead of 4 Joules per kg)		5 (1.2 Joules per kg to 6 Joules per kg)	
Defibrillation errors at 3rd attempt, n (details)		0		8 (2 Joules per kg to 6 Joules per kg)	
Defibrillation errors at 4th attempt, n (details)		0		8 (4.8 Joules per kg to 8 Joules per kg)	

[a]pVT: Pulseless ventricular tachycardia.

[b]AHA: American Heart Association.

[c]AR: Augmented Reality.

[d]PALS: Pediatric Advanced Life Support.

[e]Difference between AR glasses and PALS reference cards groups: *P*=.0010.

[f]Difference between AR glasses and PALS reference cards groups: *P*=.66.

[g]Difference between AR glasses and PALS reference cards groups: *P*<.001.

[h]Difference between AR glasses and PALS reference cards groups: *P*<.001.

[i]% deviation denotes percentage deviation from AHA recommended dose.

Errors and Deviations From AHA pVT Algorithm

Errors and deviations from the AHA pVT algorithm are summarized in Table 3. Out of 40 opportunities, 5 errors in defibrillation doses (12.5%) were committed during the whole scenario in allocation group A. This compares to 26 errors in defibrillation doses (65%) during the whole scenario in allocation group B (*P*<.001). Out of 10, one resident (10%) in allocation group B wrongly used a synchronized shock, 2/10 (20%) never delivered a third defibrillation, and 2/10 (20%) stopped the compressions after the 1st or 2nd defibrillation attempts.

The entire pVT algorithm was followed correctly in a stepwise fashion until ROSC by 60% of residents in allocation group A and 40% in allocation group B (*P*=.66).

Questionnaire About Perceived Stress and Satisfaction

The questionnaire was completed and returned by 100% of the participants. Participants in allocation group A and B rated the overall perceived stress to be 6.2 (95% CI 4.7-7.7) and 7.0 (95% CI 5.7-8.3), respectively on the Likert scale (*P*=.38). The usability, acceptance, and perception of the AR glasses were a major concern in our study and were assessed using a 17-item Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire [34]. The results will be published in a separate upcoming study.

Discussion

Principal Findings

To our knowledge, this is the first randomized controlled trial to investigate the benefit of a wearable technology to improve pediatric residents' performance and adherence with regard to AHA resuscitation guidelines. Using AR glasses, we found that

XSL•FO
RenderX

time to first defibrillation attempt, time to other critical resuscitation endpoints, and drug dose delivery were not improved in terms of adherence to AHA guidelines. However, errors and deviations from the pVT guideline in terms of defibrillation doses and cumulative defibrillation doses were significantly reduced when compared with the use of the PALS pocket reference cards.

During resuscitation, time is a decisive success criterion. During the first 15 minutes, survival and favorable neurological outcome decrease linearly by 2.1% and 1.2% per minute, respectively [35]. Delays in initiating CPR have a detrimental effect on patient outcome regardless of the quality of resuscitation [36]. AHA therefore recommends pulseless patients of any age to receive immediate CPR without delay starting with chest compressions followed by a defibrillation within 180 seconds of a shockable rhythm. However, management, procedural skills, and adherence to these guidelines have been shown to fade after a few months of initial training [37-39]. With a critical patient's condition and stress, physicians do not always have enough time to apply these guidelines and are prone to deviate from them [40]. PALS pocket cards are intended to resolve this problem by delivering fast and accurate summarized resuscitative knowledge and skills to providers. Nevertheless, Hunt et al have observed that despite availability of these recommendations, 66% of pediatric residents failed to start compressions within 60 seconds from the onset of a simulated pVT, 33% never started compressions, only 54% successfully defibrillated within 180 seconds, and 7% never discharged the defibrillator [40]. Similar results were obtained by Labrosse et al during a simulated pulseless shockable arrest scenario, where 25% of pediatric residents failed to start compressions and 4% never defibrillated a patient [41]. A more recent study among first-year pediatric residents showed a median time for initiation of CPR of 50 seconds and to first defibrillation of 282 seconds [42]. Pediatric residents performed better in our study with delays closer to AHA recommendations with a mean time to initiate chest compressions of 25 to 28 seconds and to first defibrillation of 146 seconds. In both allocation groups, 90% of residents started compressions within 60 seconds from the onset of pVT, and 70% defibrillated within 180 seconds. However, there were no advantages for residents to wear the AR glasses as they performed similarly with or without them regarding delay to critical resuscitation endpoints, whether they were PALS-certified or not. Despite an ergonomic-driven approach to adapt AHA resuscitation algorithms in AR glasses and a prior 15-minute training session for their use, our system failed to improve resuscitation efficiency in terms of time to major endpoints. An explanation might be that reducing further time to defibrillations and drug delivery was not achievable by residents training in emergency medicine. Indeed, our results in accordance with those from Hunt et al [40] showed a trend toward improvement over PGY of training and PALS certification in the mean time to all defibrillation attempts. It would be interesting in further studies to assess this assumption with certified emergency physicians.

Current AHA resuscitation guidelines emphasize 2 minutes of chest compressions between defibrillations attempts as optimal care for persistent pVT or VF in children [30,43]. In this study, residents performed similarly on average, with or without AR glasses and close to AHA recommendations.

Prompt defibrillation is crucial for termination of VF or pVT to achieve ROSC [43]. The AHA 2015 guidelines recommend treating pVT or VF in children with an initial dose of 2 J/kg [30]. For subsequent shocks, a dose of 4 J/kg is recommended, though higher energy levels may be considered up to adult dose, if not exceeding 10 J/kg (Figure 1). In this trial, residents using the PALS pocket cards were more prone to deviate from defibrillation doses than those using the AR glasses. On average, the shocks they provided were delivered with defibrillation doses 13% to 73% above AHA recommendations. "High" defibrillation doses concerned mostly the initial shock, with doses reaching up to 6 J/kg in 40% of cases when delivered by residents not wearing the glasses. In an observational study of 285 pediatric IHCA, a higher initial shock dose of more than 2 J/kg was not associated with superior termination of pVT or VF or improved survival rates [44]. In addition, children who were defibrillated with higher initial shock doses in the >3-5 J/kg range were significantly less likely to have termination of pVT or VF with ROSC or to survive the event. In our study, the final cumulative defibrillation dose delivered by the residents in the PALS pocket cards group was on average 33% significantly higher than the AHA expected value. This deviation was 10 times greater than that seen with residents wearing the AR glasses. In particular, 50% of residents using the PALS pocket cards used wrong energy doses whether it was for first or second shock delivery. In a total of 40 defibrillation attempts, they used wrong doses in 65% of cases for a cumulative defibrillation dose of 18.7 J/Kg. These errors were reduced by 53% and cumulative dose by 37% by using the AR glasses, suggesting a limited but worthwhile benefit of their use in simulated resuscitation. It would be interesting in further studies to determine whether this would translate into fewer errors in shock doses in real life.

Finally, in terms of drug dose concentrations, both groups in this study accurately administered epinephrine and amiodarone. The entire pVT algorithm was followed correctly until ROSC in a stepwise fashion by 60% of residents wearing AR glasses, compared with 40% of residents with the PALS pocket cards.

Limitations

This study has some limitations. First, it was conducted in a simulation-based resuscitation scenario. This choice was related to the ethical and organizational difficulties of conducting studies in real-life critical conditions. However, several studies have demonstrated the benefit of simulation as an investigative research methodology to answer research questions that otherwise could not be answered during CPR [45]. High-fidelity simulation is recognized as an essential tool to study resuscitation skills or technologies. Till date, none of the results obtained from simulation-based CPR studies disagreed with those obtained from studies in real life, confirming our study design choice. Realism was achieved, reflected by the stress levels experienced by the participants. They quoted the simulation as highly stressful when compared with real CPR situations.

XSL•FO
RenderX

Second, one might raise questions of our choice to choose Google Glass as the best AR glasses to display AHA guidelines. Indeed, despite its remarked 2012 commercial unveiling, Google Glass never reached its public audience and was discontinued in January 2015. However, there is a growing interest in recent literature toward its use in specialized medical fields [21,22,25,46,47]. We hypothesize that other AR glasses would not drastically change the results that we found with Google Glass, as their small size remains a major limitation to their use in displaying CPR algorithms. Further studies would be valuable to assess this assumption.

Conclusions

Taken together, our results support the interpretation that residents are not accurately following AHA recommendations during pediatric resuscitation, whether they are PALS certified or not. A wearable technology such as AR glasses might partially fill this gap and benefit patients by improving adherence and performance of residents to meet resuscitation requirements set by AHA, especially regarding delivery of defibrillation doses. In this sense, AR glasses appear as an interesting tool for emergency medicine and future studies are required to further examine this new paradigm.

Acknowledgments

We thank all the study participants for their contributions. All authors declare that there were no relationships or activities that could appear to have influenced the submitted work.

Authors' Contributions

JNS and SM did the literature search and reading, and JNS wrote the manuscript and prepared the figures and tables. JNS, FE, KH, PS, AS, and SM collected data, while JNS carried out the statistical analysis. JNS, FE, CL, AG, and SM carried out the development of the project software, and FE, LL, KH, CL, AG, SM, and JS were involved in the concept and design of the study and the critical review of manuscript content. All authors have contributed to, seen, and approved the final submitted version of the manuscript, had full access to all the data (including statistical reports and tables) in the study, and can take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author confirms that he had full access to the participants' data and endorsed the final responsibility for the submission. He further affirms that the manuscript is an honest, accurate, and transparent account of the study being reported, that no important aspects of the study have been omitted, and that any deviations from the study plan have been explained. Consent for publication was obtained from the participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH V1.6 checklist.

[PDF File (Adobe PDF File), 740KB - jmir_v19i5e183_app1.pdf]

Multimedia Appendix 2

Use of AR glasses to guide resuscitation.

[JPG File, 5MB - jmir_v19i5e183_app2.jpg]

Multimedia Appendix 3

Following the pVT pathway of the AHA algorithm when adapted for augmented reality glasses during a simulation-based pediatric cardiac arrest scenario.

[JPG File, 6MB - jmir_v19i5e183_app3.jpg]

Multimedia Appendix 4

The table details the mean time from onset of pVT to the study main endpoints in both PALS and non-PALS residents.

[PNG File, 512KB - jmir_v19i5e183_app4.png]

References

1. Thomas L, Cullum N, McColl E, Rousseau N, Soutter J, Steen N. Guidelines in professions allied to medicine. Cochrane Database Syst Rev 2000(2):CD000349. [doi: 10.1002/14651858.CD000349] [Medline: 10796531]

XSL•FO
RenderX

2. -. Standards and guidelines for Cardiopulmonary Resuscitation (CPR) and Emergency Cardiac Care (ECC). National Academy of Sciences - National Research Council. JAMA 1986 Jun 06;255(21):2905-2989. [Medline: 2871200]
3. Hazinski MF, Nolan JP, Aickin R, Bhanji F, Billi JE, Callaway CW, et al. Part 1: Executive Summary: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. Circulation 2015 Oct 20;132(16 Suppl 1):S2-39 [FREE Full text] [doi: 10.1161/CIR.0000000000000270] [Medline: 26472854]
4. Meaney PA, Bobrow BJ, Mancini ME, Christenson J, de Caen AR, Bhanji F, CPR Quality Summit Investigators, the American Heart Association Emergency Cardiovascular Care Committee, the Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation.. Cardiopulmonary resuscitation quality: [corrected] improving cardiac resuscitation outcomes both inside and outside the hospital: a consensus statement from the American Heart Association. Circulation 2013 Jul 23;128(4):417-435 [FREE Full text] [doi: 10.1161/CIR.0b013e31829d8654] [Medline: 23801105]
5. CDC. CDC. 2011. National hospital ambulatory medical care survey: 2011 emergency department summary tables URL: https://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf [accessed 2017-01-22] [WebCite Cache ID 6nh4kQv4M]
6. Knudson JD, Neish SR, Cabrera AG, Lowry AW, Shamszad P, Morales DL, et al. Prevalence and outcomes of pediatric in-hospital cardiopulmonary resuscitation in the United States: an analysis of the Kids' Inpatient Database*. Crit Care Med 2012 Nov;40(11):2940-2944. [doi: 10.1097/CCM.0b013e31825feb3f] [Medline: 22932398]
7. Topjian AA, Berg RA, Nadkarni VM. Pediatric cardiopulmonary resuscitation: advances in science, techniques, and outcomes. Pediatrics 2008 Nov;122(5):1086-1098 [FREE Full text] [doi: 10.1542/peds.2007-3313] [Medline: 18977991]
8. Gerein RB, Osmond MH, Stiell IG, Nesbitt LP, Burns S. What are the etiology and epidemiology of out-of-hospital pediatric cardiopulmonary arrest in Ontario, Canada? Acad Emerg Med 2006 Jun;13(6):653-658 [FREE Full text] [doi: 10.1197/j.aem.2005.12.025] [Medline: 16670256]
9. Kuisma M, Suominen P, Korpela R. Paediatric out-of-hospital cardiac arrests--epidemiology and outcome. Resuscitation 1995 Oct;30(2):141-150. [Medline: 8560103]
10. Sirbaugh PE, Pepe PE, Shook JE, Kimball KT, Goldman MJ, Ward MA, et al. A prospective, population-based study of the demographics, epidemiology, management, and outcome of out-of-hospital pediatric cardiopulmonary arrest. Ann Emerg Med 1999 Feb;33(2):174-184. [Medline: 9922413]
11. Atkins DL, Everson-Stewart S, Sears GK, Daya M, Osmond MH, Warden CR, Resuscitation Outcomes Consortium Investigators. Epidemiology and outcomes from out-of-hospital cardiac arrest in children: the resuscitation outcomes consortium epistry-cardiac arrest. Circulation 2009 Mar 24;119(11):1484-1491 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.802678] [Medline: 19273724]
12. McEvoy MD, Field LC, Moore HE, Smalley JC, Nietert PJ, Scarbrough SH. The effect of adherence to ACLS protocols on survival of event in the setting of in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):82-87 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.09.019] [Medline: 24103233]
13. Girotra S, Nallamothu BK, Spertus JA, Li Y, Krumholz HM, Chan PS, American Heart Association Get with the Guidelines–Resuscitation Investigators. Trends in survival after in-hospital cardiac arrest. N Engl J Med 2012 Nov 15;367(20):1912-1920 [FREE Full text] [doi: 10.1056/NEJMoa1109148] [Medline: 23150959]
14. Hunt EA, Duval-Arnould JM, Nelson-McMillan KL, Bradshaw JH, Diener-West M, Perretta JS, et al. Pediatric resident resuscitation skills improve after “rapid cycle deliberate practice” training. Resuscitation 2014 Jul;85(7):945-951. [doi: 10.1016/j.resuscitation.2014.02.025] [Medline: 24607871]
15. Kämäräinen A. Out-of-hospital cardiac arrests in children. J Emerg Trauma Shock 2010 Jul;3(3):273-276 [FREE Full text] [doi: 10.4103/0974-2700.66531] [Medline: 20930972]
16. Jayaram N, McNally B, Tang F, Chan PS. Survival after out-of-hospital cardiac arrest in children. J Am Heart Assoc 2015 Oct 08;4(10):e002122 [FREE Full text] [doi: 10.1161/JAHA.115.002122] [Medline: 26450118]
17. Field LC, McEvoy MD, Smalley JC, Clark CA, McEvoy MB, Rieke H, et al. Use of an electronic decision support tool improves management of simulated in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):138-142 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.09.013] [Medline: 24056391]
18. Low D, Clark N, Soar J, Padkin A, Stoneham A, Perkins GD, et al. A randomised control trial to determine if use of the iResus© application on a smart phone improves the performance of an advanced life support provider in a simulated medical emergency. Anaesthesia 2011 Apr;66(4):255-262 [FREE Full text] [doi: 10.1111/j.1365-2044.2011.06649.x] [Medline: 21401537]
19. Merchant RM, Abella BS, Abotsi EJ, Smith TM, Long JA, Trudeau ME, et al. Cell phone cardiopulmonary resuscitation: audio instructions when needed by lay rescuers: a randomized, controlled trial. Ann Emerg Med 2010 Jun;55(6):538-543.e1. [doi: 10.1016/j.annemergmed.2010.01.020] [Medline: 20202719]
20. Semeraro F, Frisoli A, Loconsole C, Bannò F, Tammaro G, Imbriaco G, et al. Motion detection technology as a tool for cardiopulmonary resuscitation (CPR) quality training: a randomised crossover mannequin pilot study. Resuscitation 2013 Apr;84(4):501-507. [doi: 10.1016/j.resuscitation.2012.12.006] [Medline: 23238423]

XSL•FO
RenderX

21. Chaballout B, Molloy M, Vaughn J, Brisson IR, Shaw R. Feasibility of augmented reality in clinical simulations: using google glass with manikins. JMIR Med Educ 2016 Mar 07;2(1):e2 [FREE Full text] [doi: 10.2196/mededu.5159] [Medline: 27731862]
22. Chai PR, Wu RY, Ranney ML, Bird J, Chai S, Zink B, et al. Feasibility and acceptability of google glass for emergency department dermatology consultations. JAMA Dermatol 2015 Jul;151(7):794-796. [doi: 10.1001/jamadermatol.2015.0248] [Medline: 25874898]
23. Iqbal MH, Aydin A, Lowdon A, Ahmed HI, Muir GH, Khan MS, et al. The effectiveness of google glass as a vital signs monitor in surgery: ac simulation study. Int J Surg 2016 Dec;36(Pt A):293-297. [doi: 10.1016/j.ijsu.2016.11.013] [Medline: 27833004]
24. Liebert CA, Zayed MA, Aalami O, Tran J, Lau JN. Novel use of google glass for procedural wireless vital sign monitoring. Surg Innov 2016 Aug;23(4):366-373. [doi: 10.1177/1553350616630142] [Medline: 26848138]
25. Mitrasinovic S, Camacho E, Trivedi N, Logan J, Campbell C, Zilinyi R, et al. Clinical and surgical applications of smart glasses. Technol Health Care 2015;23(4):381-401. [doi: 10.3233/THC-150910] [Medline: 26409906]
26. Stetler J, Resendes E, Martinez-Parachini JR, Patel K, Amsavelu S, Tarar MN, et al. Hands-free zoom and pan technology improves the accuracy of remote electrocardiogram interpretation using google glass. Int J Cardiol 2016 Feb 01;204:147-148. [doi: 10.1016/j.ijcard.2015.11.144] [Medline: 26657610]
27. Ehrler F, Siebert J, Haddad K, Sahin A, Schrurs P, Diener R, et al. Adapting guidelines for google glass: the case of pediatric CPR. Stud Health Technol Inform 2016;224:141-145. [Medline: 27225569]
28. Nadkarni VM, Larkin GL, Peberdy MA, Carey SM, Kaye W, Mancini ME, National Registry of Cardiopulmonary Resuscitation Investigators. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. JAMA 2006 Jan 4;295(1):50-57. [doi: 10.1001/jama.295.1.50] [Medline: 16391216]
29. Samson RA, Nadkarni VM, Meaney PA, Carey SM, Berg MD, Berg RA, American Heart Association National Registry of CPR Investigators. Outcomes of in-hospital ventricular fibrillation in children. N Engl J Med 2006 Jun 01;354(22):2328-2339. [doi: 10.1056/NEJMoa052917] [Medline: 16738269]
30. de Caen AR, Berg MD, Chameides L, Gooden CK, Hickey RW, Scott HF, et al. Part 12: pediatric advanced life support: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2015 Nov 3;132(18 Suppl 2):S526-S542 [FREE Full text] [doi: 10.1161/CIR.0000000000000266] [Medline: 26473000]
31. Chan PS, Krumholz HM, Nichol G, Nallamothu BK, American Heart Association National Registry of Cardiopulmonary Resuscitation Investigators. Delayed time to defibrillation after in-hospital cardiac arrest. N Engl J Med 2008 Jan 03;358(1):9-17. [doi: 10.1056/NEJMoa0706467] [Medline: 18172170]
32. Dilley SJ, Weiland TJ, O'Brien R, Cunningham NJ, Van Dijk JE, Mahoney RM, et al. Use of a checklist during observation of a simulated cardiac arrest scenario does not improve time to CPR and defibrillation over observation alone for subsequent scenarios. Teach Learn Med 2015;27(1):71-79. [doi: 10.1080/10401334.2014.979182] [Medline: 25584474]
33. Sealed envelope. URL: https://www.sealedenvelope.com [accessed 2017-01-21] [WebCite Cache ID 6ngSSUzOS]
34. JSTOR. MIS Quarterly URL: http://www.jstor.org/stable/30036540 [accessed 2017-03-29] [WebCite Cache ID 6pKNA7UCq]
35. Matos RI, Watson RS, Nadkarni VM, Huang H, Berg RA, Meaney PA, American Heart Association's Get With The Guidelines–Resuscitation (Formerly the National Registry of Cardiopulmonary Resuscitation) Investigators. Duration of cardiopulmonary resuscitation and illness category impact survival and neurologic outcomes for in-hospital pediatric cardiac arrests. Circulation 2013 Jan 29;127(4):442-451. [doi: 10.1161/CIRCULATIONAHA.112.125625] [Medline: 23339874]
36. Ross JC, Trainor JL, Eppich WJ, Adler MD. Impact of simulation training on time to initiation of cardiopulmonary resuscitation for first-year pediatrics residents. J Grad Med Educ 2013 Dec;5(4):613-619 [FREE Full text] [doi: 10.4300/JGME-D-12-00343.1] [Medline: 24455010]
37. Smith KK, Gilcreast D, Pierce K. Evaluation of staff's retention of ACLS and BLS skills. Resuscitation 2008 Jul;78(1):59-65. [doi: 10.1016/j.resuscitation.2008.02.007] [Medline: 18406037]
38. Tarantinos K, Chalkias A, Giotakis E, Athanasopoulou P, Xanthos T. Retention of knowledge and skills after advanced cardiovascular life support courses. Am J Emerg Med 2014 Sep;32(9):1143-1147. [doi: 10.1016/j.ajem.2014.06.017] [Medline: 25053263]
39. Yang C, Yen Z, McGowan JE, Chen HC, Chiang W, Mancini ME, et al. A systematic review of retention of adult advanced life support knowledge and skills in healthcare providers. Resuscitation 2012 Sep;83(9):1055-1060. [doi: 10.1016/j.resuscitation.2012.02.027] [Medline: 22391016]
40. Hunt EA, Vera K, Diener-West M, Haggerty JA, Nelson KL, Shaffner DH, et al. Delays and errors in cardiopulmonary resuscitation and defibrillation by pediatric residents during simulated cardiopulmonary arrests. Resuscitation 2009 Jul;80(7):819-825. [doi: 10.1016/j.resuscitation.2009.03.020] [Medline: 19423210]
41. Labrosse M, Levy A, Donoghue A, Gravel J. Delays and errors among pediatric residents during simulated resuscitation scenarios using pediatric advanced life support (PALS) algorithms. Am J Emerg Med 2015 Oct;33(10):1516-1518. [doi: 10.1016/j.ajem.2015.07.049] [Medline: 26298054]

XSL•FO
RenderX

42. Jeffers J, Eppich W, Trainor J, Mobley B, Adler M. Development and evaluation of a learning intervention targeting first-year resident defibrillation skills. Pediatr Emerg Care 2016 Apr;32(4):210-216. [doi: 10.1097/PEC.0000000000000765] [Medline: 26990849]
43. Kleinman ME, Chameides L, Schexnayder SM, Samson RA, Hazinski MF, Atkins DL, et al. Part 14: pediatric advanced life support: 2010 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2010 Nov 02;122(18 Suppl 3):S876-S908 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.110.971101] [Medline: 20956230]
44. Meaney PA, Nadkarni VM, Atkins DL, Berg MD, Samson RA, Hazinski MF, American Heart Association National Registry of Cardiopulmonary Resuscitation Investigators. Effect of defibrillation energy dose during in-hospital pediatric cardiac arrest. Pediatrics 2011 Jan;127(1):e16-e23. [doi: 10.1542/peds.2010-1617] [Medline: 21172997]
45. Cheng A, Auerbach M, Hunt EA, Chang TP, Pusic M, Nadkarni V, et al. Designing and conducting simulation-based research. Pediatrics 2014 Jun;133(6):1091-1101 [FREE Full text] [doi: 10.1542/peds.2013-3267] [Medline: 24819576]
46. Chang JY, Tsui LY, Yeung KS, Yip SW, Leung GK. Surgical vision: google glass and surgery. Surg Innov 2016 Aug;23(4):422-426. [doi: 10.1177/1553350616646477] [Medline: 27146972]
47. Aungst TD, Lewis TL. Potential uses of wearable technology in medicine: lessons learnt from Google Glass. Int J Clin Pract 2015 Oct;69(10):1179-1183. [doi: 10.1111/ijcp.12688] [Medline: 26412708]

Abbreviations

AHA: American Heart Association
AR: Augmented Reality
CA: Cardiac Arrest
CPR: Cardiopulmonary Resuscitation
IHCA: In-Hospital Cardiac Arrest
OHCA: Out-of-Hospital Cardiac Arrest
PALS: Pediatric Advanced Life Support
PED: Pediatric Emergency Department
PGY: Post-Graduation Years
pVT: pulseless ventricular tachycardia
ROSC: Return Of Spontaneous Circulation
VF: Ventricular Fibrillation

Edited by G Eysenbach; submitted 21.01.17; peer-reviewed by K Morawski, M Leary, A Elmessiry; comments to author 26.03.17; revised version received 03.04.17; accepted 28.04.17; published 29.05.17.

Please cite as:
Siebert JN, Ehrler F, Gervaix A, Haddad K, Lacroix L, Schrurs P, Sahin A, Lovis C, Manzano S
Adherence to AHA Guidelines When Adapted for Augmented Reality Glasses for Assisted Pediatric Cardiopulmonary Resuscitation: A Randomized Controlled Trial
J Med Internet Res 2017;19(5):e183
URL: http://www.jmir.org/2017/5/e183/
doi:10.2196/jmir.7379
PMID:28554878

©Johan N Siebert, Frederic Ehrler, Alain Gervaix, Kevin Haddad, Laurence Lacroix, Philippe Schrurs, Ayhan Sahin, Christian Lovis, Sergio Manzano. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identifying and Understanding the Health Information Experiences and Preferences of Caregivers of Individuals With Either Traumatic Brain Injury, Spinal Cord Injury, or Burn Injury: A Qualitative Investigation

Nathan T Coffey[1], MPH; James Cassese[1], MS; Xinsheng Cai[2], PhD; Steven Garfinkel[2], PhD; Drasti Patel[1], MPH; Rebecca Jones[2], MPH; Dahlia Shaewitz[2], MA; Ali A Weinstein[1], PhD

[1]Center for the Study of Chronic Illness and Disability, George Mason University, Fairfax, VA, United States

[2]American Institutes for Research, Washington, DC, United States

Corresponding Author:
Ali A Weinstein, PhD
Center for the Study of Chronic Illness and Disability
George Mason University
4400 University Drive
MSN 5B7
Fairfax, VA, 22030
United States
Phone: 1 703 993 9632
Fax: 1 703 993 2695
Email: aweinst2@gmu.edu

Abstract

Background: In order to meet the challenges of caring for an injured person, caregivers need access to health information. However, caregivers often feel that they lack adequate information. Previous studies of caregivers have primarily focused on either their time and emotional burdens or their health outcomes, but the information needs of caregivers have not been thoroughly investigated.

Objective: The purpose of this investigation was to identify the preferred sources of health information for caregivers supporting individuals with injuries and to explore how access to this information could be improved.

Methods: A total of 32 caregivers participated in semistructured interviews, which were used in order to develop a more in-depth understanding of these caregivers' information needs. Digital audio recordings of the interviews were used for analysis purposes. These audio recordings were analyzed using a thematic analysis or qualitative content analysis. All of participant's interviews were then coded using the qualitative analysis program, Nvivo 10 for Mac (QSR International).

Results: The caregivers endorsed similar behaviors and preferences when seeking and accessing health information. Medical professionals were the preferred source of information, while ease of access made the Internet the most common avenue to obtain information. The challenges faced by participants were frequently a result of limited support. In describing an ideal health system, participants expressed interest in a comprehensive care website offering support network resources, instructive services about the injury and caregiving, and injury-specific materials.

Conclusions: According to the participants, an ideal health information system would include a comprehensive care website that offered supportive network resources, instructive services about the injury and caregiving, and materials specific to the type of patient injury.

(J Med Internet Res 2017;19(5):e159) doi:10.2196/jmir.7027

KEYWORDS

traumatic brain injury; burns; spinal cord injuries; caregivers; health information, consumer; qualitative research

XSL•FO
RenderX

Introduction

More than 52 million US residents act as a caregiver to an adult aged 18 years or older [1]. That number increases to over 65.7 million for the US adult population that care for patients of all age groups [2]. Almost two-thirds of those receiving care have a long-term physical condition [3]. The caregivers supporting these individuals are commonly unpaid persons, including family, friends, and neighbors of the injured or ill [2]. For this investigation, participants are self-identified as the primary caregiver, either in a full-time or part-time capacity.

Caregivers assume responsibilities that affect their personal lives, financial status, and mental well-being [4]. They often provide housing, coordinate rehabilitative services, and communicate with health care providers [5]. An injury, by its very nature, is the result of an acute event. Unlike chronic conditions that can be slow moving, those caring for an injured person are thrust into the position with no time to prepare. Previous examinations of those providing care to individuals with traumatic brain injury (TBI) found that they frequently feel overwhelmed [6,7]. Low health-related quality of life among TBI caregivers has been associated with decreased mental health status [8]. Caregivers of individuals with spinal cord injury (SCI) can experience depressive symptoms at similar rates as the injured person [9], whereas caregivers of individuals with burn injuries experience more pronounced symptoms of depression than caregivers of individuals who are hospitalized for other reasons [10].

Many caregivers experience unmet support needs, including insufficient care-related information, unsatisfactory emotional support, and difficulty maintaining relationships [11-15]. The majority of caregivers lack the time or energy to care for themselves and often experience declining health after assuming that responsibility [6,16]. Among parents caring for children with a burn injury, perceived lack of support was significantly associated with symptoms of general anxiety and depression [15].

In order to meet the challenges of caring for an injured person, caregivers need access to health information. However, caregivers often feel that they lack adequate information and support [11,14]. Previous examination of TBI caregivers found that health information was one of their most important needs [12]. Caregivers also seek methods of accessing services and support [14]. Information about managing physical and emotional stress is highly sought as well [17].

Health information is critical for caregivers to provide the most effective care for the patients [5]. Family caregivers commonly perform medical or nursing tasks [18]. Consequently, caregivers seek Web-based reviews of particular drugs or medical treatments more frequently than noncaregivers [19]. A lack of information can lead to frustration and anxiety about the short- and long-term effects of the injury [4]. Common sources of health information for caregivers include medical professionals, written sources, family or friends, nonprint media, and the Internet [20]. Those with higher health literacy are more likely to obtain information from multiple sources [20]. However, many potentially beneficial activities, including independent research regarding care, may be hindered due to unmet support needs such as managing concerns and reducing stress [21].

TBI, SCI, and burn injuries are prevalent throughout the United States and are severe in nature [22-24]. According to the Centers for Disease Control and Prevention (CDC), over 2 million patients were admitted to the hospital for a TBI in 2010 [22]. Another 40,000 were hospitalized for a burn injury [24]. There are about 12,500 new cases of SCI every year according to the National Spinal Cord Injury Statistical Center (NSCISC) [23]. Due to their severity, all three injuries can require long-term care and rehabilitation; thus the need for health information among caregivers for persons with these injuries is pronounced. Previous studies of caregivers have primarily focused on either their burdens or health outcomes. Additionally, many prior studies focused on the challenges associated with either pediatric or elder care. This qualitative study seeks to gain insight into the preferred sources of health information for caregivers supporting individuals with TBI, SCI, or burn injuries to ensure a greater depth of understanding in this area than in previous work. Additionally, the investigators sought to assess the obstacles faced by those caring for persons with these injuries when accessing services and information through the health care system, and how access to these items could be improved.

Methods

Subject Recruitment and Selection

Subjects were recruited through a number of different outreach methods. First, National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR)–funded SCI, TBI, and Burn Model Systems grantees were contacted to recruit participants. The grantees were asked to identify potential participants and direct candidates to the recruitment website. Subjects were also solicited through advertisements placed in printed materials, websites, as well as social media sites. Finally, participants were reached by sending recruitment emails provided by TBI, SCI, and burn consumer advocacy groups.

Subjects were selected to participate if they self-identified as the primary caregiver of a person with either a TBI, SCI, or burn injury. The number of hours per week spent on caregiving was not an inclusion/exclusion criterion. All caregivers needed to be at least 18 years of age and could care for a recipient of any age. Informed consent was obtained (orally) from all of the participants and assurances of anonymity/confidentiality given. Participant distress and safeguarding protocols were established. The Institutional Review Boards of both the American Institutes of Research and George Mason University approved this investigation.

Data Collection

Semistructured interviews were used in order to develop a more in-depth understanding of caregivers' information needs. A psychologist (AAW) trained all interviewers on the proper techniques in conducting semistructured interviews. These trained interviewers were an occupational therapist and Masters of Public Health graduate students. The interview consisted of three sections. In the first section of the interview, caregivers were asked to discuss the subject's care background as well as

XSL•FO
RenderX

the activities of daily living and instrumental activities of daily living of the person receiving care. The second section used a questionnaire to assess the various types of needs caregivers may have. The final open-ended section of the interview examined the individual's medical, rehabilitation, and continuing care experience, specifically focusing on information needs assessment (Textbox 1). For this qualitative investigation, only the open-ended portion of the interviews is reported. The investigators used a semistructured interview guide that allowed consistency in questions that were asked across interviews. The interviews were conducted and recorded over the phone for participants' convenience and to allow for geographic variability in the location of the caregivers' residences.

Textbox 1. Open-ended question prompts.

- Thinking back over the time since the injury, what has been most difficult for you?
- What has been the most helpful for you since the injury?
- We had discussed your health earlier in the interview, and I was wondering if you feel that your health has changed since becoming a caregiver?
- Now, we are going to spend some time focusing on the informational component of your life as a caregiver. So, how do you currently receive information related to (TBI, SCI, and burn injury)?
- When you do receive or access information, how, if at all, do you use the information?
- Has the information that you received or accessed been useful? Pertinent to your needs? Accurate? Understandable? Trustworthy?
- What are the biggest difficulties you face in accessing the kind of information you want about the patient's (or loved one if appropriate) condition?
- I want you to imagine that we can start from scratch and develop a new approach for getting information to caregivers. In your dream world, what would this new model look like? Don't worry about the money required, or whether your idea is logistically feasible.

Analysis

Digital audio recordings of the interviews were used for analysis purposes. These audio recordings were analyzed using the framework approach [25]. This approach is within the broad family of analysis methods often termed "thematic analysis" or "qualitative content analysis." This type of analysis was used to draw descriptive or explanatory conclusions clustered around themes, while allowing the ability to analyze the data across individuals as well as within individuals. The research team, before setting up a thematic structure, familiarized themselves with the interviews [26] by listening to the recordings and reviewing notes taken by the interviewers during the interviews. Then, the analysis team listened to a random selection of interviews that included all injury types (TBI, SCI, and burn) to create a coding system. These researchers then met to discuss the coding and establish a set of codes that would be applied to all of the interviews. Once the coding system had been reviewed and refined, the two researchers who would code all of the interviews, coded the same interview to determine interrater reliability (IRR) and the systematic application of codes. Nvivo's coding comparison query showed a 96.4% agreement (kappa coefficient, κ=.65). Percentage agreement refers to the percentage where the two users coded the data in the same way. Since audio files were coded, the unit of measurement is duration in seconds. Average percentage agreement and kappa values indicated consistent interpretation of codes by the different interview coders.

All of participant's interviews were then coded using Nvivo. In some cases, a participant would respond to a question in a way that fell beyond the scope of the developed coding system. In those cases, a new code was added and the remaining research team members were notified of the change. In total, 5 codes were modified during the coding process.

The coding system was then used to develop an analytical framework. The codes were grouped together into categories. In order to facilitate this process, the tree map feature was used to visually represent potential categories. The researchers then defined these groupings of codes. The last step was the interpretation of the data. The researchers met many times to discuss the characteristics of and differences between the data.

Results

Participant Characteristics

A total of 40 caregivers provided consent and were interviewed. However, 8 of the interview audio files were unable to be read by the Nvivo program. Thus, the analysis was based on the remaining 32 caregiver interviews, of which 16 providing care to a person with a TBI, 10 providing care to a person with a SCI, and 6 providing care to a person with a burn injury. The open-ended response section of interviews lasted 34 min on average, ranging from 12 min to 115 min.

Study participants had provided care for 7.0 years (SD 8.4) on average. Details of the study population's ethnicity, socioeconomic status, and injury type can be found in Table 1.

XSL•FO
RenderX

Table 1. Demographics of the care recipients and the caregivers.

Characteristics of the sample	Total (N=32)	TBI[a] (n=16)	SCI[b] (n=10)	Burn (n=6)
Recipient age[c] (in years)				
Under 18	3	1	1	1
18-29	10	6	2	2
30-39	3	3	0	0
40-49	5	1	3	1
50 and over	10	4	4	2
Recipient gender				
Male	27	14	9	4
Female	5	2	1	2
Recipient insurance status				
Insured	29	14	10	5
Uninsured	3	2	0	1
Caregiver age[d,e] (in years)				
18-29	0	0	0	0
30-39	6	2	2	2
40-49	10	4	3	3
50 and over	13	8	4	1
Caregiver gender				
Male	2	1	0	1
Female	30	15	10	5
Caregiver education level[f]				
High school	7	3	2	2
Associates degree	2	1	0	1
1-3 years of college	10	6	4	0
Bachelor's degree	7	3	1	3
Graduate school	5	2	3	0
Length of time caring for patient (in years), mean (SD[g])	7.0 (8.4)	7.1 (6.7)	10.1 (11.7)	1.5 (1.5)
Hours per week spent caring for patient, mean (SD)	67.9 (59.1)	73.6 (64.2)	62.3 (53.0)	61.8 (63.6)

[a]TBI: traumatic brain injury.

[b]SCI: spinal cord injury.

[c]For TBI, n=15.

[d]For TBI, n=14.

[e]For SCI, n=9.

[f]For TBI, n=15.

[g]SD: standard deviation.

Preferred Sources of Health Information

The majority of subjects (n=30) received injury-related information via the Internet. The most common ways caregivers accessed Web-based information was through sites found using search engines (n=20) and medical websites (n=20). Not all types of Internet resources were widely used; a smaller proportion of caregivers visited government (n=10), support group (n=9), or discussion board (n=5) websites regularly.

Health Information Access and Use

Interference With Accessing Information

Caregivers described difficulties that interfered with their ability to access information about treatment, caregiving, and health conditions. When asked to describe their greatest obstacle to accessing information, caregivers most commonly cited challenges simply finding information (n=10), followed by medical/insurance bureaucracy (n=9), unsupportive medical

XSL•FO
RenderX

staff (n=7), and lack of support from sources other than medical staff (n=4). The majority of caregivers (n=28) believed that "access to the Internet" did not hinder their ability to access health information. Similarly, difficulty in finding "culturally appropriate" information did not hinder access (n=26). In contrast, caregivers believed that the "accessibility to information (other than the Internet)" (n=15), "time to search" (n=14), the "understandability of information" (n=12), and obtaining "credible information" (n=10) at least partially interfered with their ability to access to injury-related information. Only 8 caregivers indicated that they did not face any difficulties when accessing information. However, finding concrete information still proved to be a challenging task due to the inaccessibility of specialized care or gaps in medical research.

> *There are just so few people that are knowledgeable in our area, and probably in most geographic areas, about the specialized needs of people with spinal cord injury; especially quadriplegics. There are so many systems of the body affected and there aren't too many (specialists) in any one given area who really understand all the ramifications of how the injury's affecting all the body systems.* [Interview 2003; SCI]

> *I think just finding (health information) was the biggest challenge. There's not a lot known about burns, relatively speaking. There's a lot of anecdotal stuff from survivors but there's not a lot publishedresearch on burns and burn recovery.* [Interview 1002; burn]

> *We've chosen doctors who are, who will listen. Who are open to, if they don't know the answer, taking what we say, and checking it. But last time we looked for a doctor, I actually went and interviewed doctors...because there aren't a lot of people that know about (patient's) exact condition.* [Interview 2011; SCI]

Information-Seeking Behaviors

Themes surrounding the process of gaining health information emerged seeking different types of information, utilizing multiple sources, and making comparisons. The process of locating information involved extensive preparation, persistence, and due diligence.

Although there were differences in the types of health information individual caregivers sought, common themes emerged. Caregivers researched information that was directly related to supporting the individual receiving care. "Treatment" (n=14), "rehabilitation" (n=13), and "medication" (n=11) emerged as dominant themes. Due to the inherent chronicity of these serious injuries, many caregivers sought information on "long-term care" (n=15), that is, research on the long-term effects of patient conditions and sequela, and adjustments to disabilities.

For caregivers, an important source of health information was seeking information through individuals who had similar experiences. These opportunities for seeking information could be from various outlets including organized support groups designed to share experiences with those in similar situations, informal sharing, development of coping strategies with professional counselors, doctor relationships, or rehabilitation settings with clinical specialists and organizations centered on TBI, SCI, or burn survivors.

> *When I have actually sought counseling...about... trying to adapt and cope with this kind of new person that I'm dealing with. That's been very helpful to me to have some support person that I can call.* [Interview 2009; SCI]

> *we call it the working support group, because it was headed by a neuropsychologist, and he would actually have the survivors worktrying to improve themselves. So that has been very useful. The support group that we go to here in our town is more of a social kind of thing* [Interview 3015; TBI]

Caregivers utilized a number of different sources to acquire health information. Common sources included "Web-based research" (n=26), "doctor instructions" (n=16), "medical journals" (n=14), and "injury organizations" (n=13). Web-based communication, library visits, clinical trials, and other persons with similar injury experiences were referenced, albeit less frequently. Caregivers often compared information from multiple sources (n=21) to evaluate its quality and corroborate its accuracy.

> *I spent a lot of time, really the first 4 monthsdoing PubMed searches, and reading literature, and reading textbooks on burns, and all that kind of stuff. That was really the first thing I did, was kind of educate myself about burn and the burn process and recovery and healing.* [Interview 1008; burn]

> *Everything and anything I could get my hands on to understand (the injury). I have visited WebMD, the traumatic brain injury model centers...the National Traumatic Brain Injury Association. I've been out there on all the websites.* [Interview 3001; TBI]

Information Utility

Personal experience dealing with an injury also served as a reference point for caregivers. Caregivers retrospectively compared their patient's outcomes with Web-based information they received, enabling them to determine its usefulness, accuracy, and understandability. However, the caregivers faced a steep learning curve when acclimating themselves with injury-related health information early in their transition to a caregiver role.

> *When I did read about the burns and different things like that, we may not have been at the stage that it talked about, but we did get there. And I could remember times where it was like "oh, that is what they said was going to happen.*" [Interview 1004; burn]

> *I did understand it, but after, going, advocating and learning more, and then going back to some of the older documents that I had researched, and rereading it, I understood it more, because I guess I had actually experienced some of it.* [Interview 3001; TBI]

XSL•FO
RenderX

Knowledge of various treatment options also influences caregiver actions and their respective patient's rehabilitative outcomes. Caregivers reported instances of reliance on their own judgment based on independent research. This sometimes led to negotiation or collaboration with medical professionals.

> *...there's very little research on the subject, but there's some, about the link between things like fish oil and brain injury. And, I had to try that, and noticed a really big improvement in both areas, in terms of clarity of thinking and his skin condition. It's anecdotal, I have no proof, but I really feel like it's made a difference. It's not something I would have considered without looking at websites.* [Interview 3003; TBI]

> *a few years ago I said to his physician, shouldn't he be tested for osteoporosisI had learned online that quads are at high-risk for osteoporosis. So he agreed to test him, and did test him, and indeed he did have osteoporosis.* [Interview 2011; SCI]

> *If we're talking about medications or treatments, or, you know things that I may have read about but no one has suggested. So I kind of throw that out there like neurofeedback, you know, as something that I ran across, you know on the Internet, and benefits for survivors of traumatic brain injury. And so I brought that up to one of his neuropsychologists and we have adopted...getting him involved with neurofeedback.* [Interview 3017; TBI]

Improvements to Health Information Access

The final open-ended interview question asked participants to describe their ideal system for getting health information to caregivers without having to consider practical limitations such as cost. In doing so, respondents offered a variety of potential modifications to health information systems that they felt would be beneficial to caregivers. In conceptualizing model systems, participants often focused on health information, support services, and modes of delivery most applicable to them. As a result, responses diverged from one another, but maintained commonalities in information sought, types of services, and access to information.

Health Information

Support network information (n=26) was the most common response, though again employing a multitudinous definition of support, including organized support groups, information for and from people with shared experiences, resources to access services related to rehabilitation, ongoing care, and daily living.

> *I would want the information probably right after whatever had happened and I wouldwant some positive reinforcement, and just general information, and maybe somesupport right off the bat from other people; other people that have gone through thiswho have experienced the same thing. And it needs to be positive interaction, not negative.* [Interview 3009; TBI]

Caregivers also expressed a desire for access to updated publications (n=11) as a method of staying well-informed with current research. Guides covering treatment, recovery, and rehabilitation, via the Internet (n=5) or in hard-copy form (n=9), were suggested as well. Those guides would serve as an educational tool to help prepare caregivers and thus mitigate their risk of feeling overwhelmed following the injury event.

> *I think it would be great to have a manual on burn injury and recovery that is categorized by subject matter and is in both print and online form and easily available to all patients and family members. That has put together in one place, all known and accurate current research and has weeded out anything that has either old and has been proven inaccurate, or irrelevant, or whatever. So it's done all the librarian work for us, so that we have it all in one place.* [Interview 1002; burn]

> *There's no good, central clearinghouse. It'd be great if there was one place on the Web you could go that had links to all these places. That had the government links, that had care-cure forums, that had Avery biomedical, every place that, that can help you. Every place that can help you, like with durable medical supplies, a place that can help you with...hooking up with caregivers. Places that could help you connect with other people in your situation.* [Interview 2011; SCI]

Support Services

As respondents discussed the services that they believed should be available in a model health system, two themes emerged. The first was caregiver support services. Specifically, resources such as "support groups" (n=22), "care coordinators" (n=15), "caregiver training" (n=14), and "classes on injury condition, treatment, and life" (n=14) were frequently suggested to prepare caregivers and their families for the rigors of caring for someone with TBI, SCI, or burn.

> *Everybody needs to be professionally trainedwhether it be education, legal, medical, or whatever, that the proper people are trained and are able to explain, or tell the caregivers and the survivor, what to expect, where they can go, and so forth. So I think the training is very much needed.* [Interview 3001; TBI]

Several respondents recommended support group sessions with people that had similar experiences for patients and their caregivers, while they are still in the hospital.

> *A support group led by a professional, but run by, and taken over by the people who are the caregivers, or the family membersthe one paid person would actually get the real information and give it to the family members, and the family members in turn could be providing things that they learned, which in turn could be coordinated and disseminated as information.* [Interview 2010; SCI]

> *I wish there was somebody like myselfthat could come into a burn survivor and their family's room and make that connection and answer their personal questionsif I would have had somebody like me or my other SOAR (Survivors Offering Assistance in Recovery) volunteer*

XSL•FO
RenderX

> *members come into my room, even that first week, I feel like my life would have been so much easier. I felt so lost.* [Interview 3016; burn]

The second theme that emerged was medical care services. Respondents expressed their desire for "rehabilitation services" (n=19), "medical specialists" (n=13), and "financial services" (n=6). In many cases, respondents expressed their desire for their local health care provider to recreate the services offered at state-of-the-art facilities. Physical access to specialized treatment facilities can be a substantial barrier; therefore, caregivers prioritized proximity and transportation to health care facilities.

Information Access

Caregivers most frequently recommended that an ideal health information system be accessed through in-person communication (n=20) or via the Internet (n=18). A combination of these two methods, such as a "livechat" feature (n=7), was cited to allow for interactive question and answer sessions. Printed materials (n=9), phone systems (n=8), and television programs (n=3) were also recommended as resources for sharing information or to serve as learning materials for caregivers. In the context of comprehensive Web-based information, the user-friendliness of the Web page and an effective search function emerged as critical features.

> *At least if you have that human, that person-to-person dialogue going, you can sometimes get answers to things that are specific to your own situation, instead of reading through a Q&A or something that's posted.* [Interview 1004; burn]

> *The Internet, of course, is the easiest and fastest way for me to get information besides talking to actual doctors in the hospital.* [Interview 3001; TBI]

> *It would be nice if there was a hotline you could call if you were in a situation where you needed somebody to talk to right away. There are a lot of urgent situations when you are dealing with disability and I wish there was some kind of caregiver hotline that would not only support you emotionally, but maybe help alleviate some of the stress when you feel like you're dealing with something completely alone. [Interview 2003; SCI]* [Interview 2003; SCI]

Discussion

Principal Findings

This study offers insights into the challenges faced by individuals caring for persons with TBI, SCI, and burn injuries. Participants revealed the obstacles that they faced, including gaps in formal and informal support systems, emotional strain, and a limited ability to access health services and health information resources. The interviews further revealed patterns and preferences for seeking health information among those caring for individuals with TBI, SCI, and burn injuries.

Although previous research has examined the burden experienced by individuals caring for elderly persons or degenerative conditions associated with aging, such as dementia, relatively few studies have compared the burden experienced by those caring for life-long injuries. This study finds that medical professionals are the preferred method of receiving health information among TBI, SCI, and burn caregivers, similar to the preference of those caring for other conditions [27]. Nevertheless, the Internet was an integral health information resource for caregivers of persons with all three injuries. It is important to remember that one of our recruitment techniques was Web-based, therefore, our results may be positively biased toward Internet resources. These findings suggest that caregivers of these injuries rely on the Internet in a similar capacity as those caring for persons with chronic or long-term diseases [17,20,28]. Furthermore, our study participants sought health information from multiple sources beyond the Internet, much like caregivers of persons with other conditions [20,29,30]. These findings indicate that the methods and preferences of accessing health information among caregivers are similar regardless of the conditions being cared for.

When TBI, SCI, and burn caregivers sought health information, it was commonly related to medical care for the injured person. Caregivers are often required to perform medical tasks despite rarely having clinical training [3,18]. This underscores caregiver's need for increased access to medical personnel. It also highlights the need for formal caregiver training, a desire strongly expressed by caregivers of every injury. Caregiver training has been demonstrated to improve caregiver self-efficacy and stress management [31-33]. Comprehensive health care systems should therefore incorporate structured training programs to reduce caregiver burnout and dependence on medical personnel.

The obstacles and preferences that caregivers discussed indicate a pronounced need for access to Web-based resources that provide synthesized care information and access to support resources. Features that combine the utility of the Internet with the expertise of medical professionals, including care-support hotlines, have been shown to reduce caregiver burden and therefore may improve their ability to provide care [34]. Thus, use of technology to improve communication with clinicians and support organizations should be prioritized by health care systems. Internet-based support for caregiving has been a valued model for TBI patients [35-37]. In addition, training caregivers on how to evaluate health care information on the Internet could be an excellent way to improve the caregivers' confidence in accessing quality information.

Strengths and Limitations

The in-depth, semistructured interviews allowed each participant to provide thorough insights into the health information issues related to their experience. Common experiences emerged, denoting the transferability of responses. Still, there were limitations. This study focused on individual interviews. A focus group study involving caregivers of individuals with these injury types may provide additional information because of the group interaction that is inherent in focus groups. Given the retrospective nature of the questions, there is a risk of recall bias from participants. In addition, this study had a small sample size and an uneven distribution of caregivers for each injury type. Study participants were predominantly recruited using the Internet which may have biased our sample in terms of

XSL•FO
RenderX

preferences for Internet resources. The number of caregivers for each injury type was different, impacting the relative weight of responses from each study participant. Due to technical limitations, eight of the interview audio files had to be excluded from analysis. Caregivers were not asked if they were the only individuals who were caring for the recipient, which could have impacted their information needs.

The codebook itself could have been simplified by combining codes. Interrater reliability measures, percentage agreement, and kappa coefficient, indicated that both coders applied codes consistently, most likely due to the shared creation of the codebook, and follow-ups on interpretation of each code. However, IRR was not 100%, therefore minor discrepancies in code placement may be assumed.

Conclusions

The detailed interviews conducted with TBI, SCI, and burn caregivers revealed the methods and resources that were used to acquire health information. Medical professionals were the preferred source of information, while ease of access made the Internet the most common. The challenges faced by participants were frequently a result of limited professional and social support. In describing an ideal health system, study participants expressed interest in a comprehensive care website that offered support network resources, instructive services about the injury and caregiving, and materials specific to their injury. An ideal health information system should incorporate methods of communicating directly with health professionals.

Acknowledgments

This work was funded by a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR).

Conflicts of Interest

None declared.

References

1. Coughlin J. Estimating the impact of caregiving and employment on well-being. Outcomes Insights Health Manag 2010 May;2(1):1-7 [FREE Full text]
2. National Alliance for Caregiving. Caregiving. 2009. Caregiving in the U.S., 2009 URL: http://www.caregiving.org/pdf/research/Caregiving_in_the_US_2009_full_report.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmTles2U]
3. National Alliance for Caregiving and the AARP Public Policy Institute. Caregiving. 2015 Jun. Caregiving in the U.S, 2015 URL: http://www.caregiving.org/wp-content/uploads/2015/05/2015_CaregivingintheUS_Final-Report-June-4_WEB.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmTy6A1Q]
4. Greenwood N, Smith R. Barriers and facilitators for male carers in accessing formal and informal support: a systematic review. Maturitas 2015 Oct;82(2):162-169 [FREE Full text] [doi: 10.1016/j.maturitas.2015.07.013] [Medline: 26271710]
5. Wharewera-Mika J, Cooper E, Kool B, Pereira S, Kelly P. Caregivers' voices: the experiences of caregivers of children who sustained serious accidental and non-accidental head injury in early childhood. Clin Child Psychol Psychiatry 2016 Apr;21(2):268-286. [doi: 10.1177/1359104515589636] [Medline: 26088208]
6. Kratz AL, Sander AM, Brickell TA, Lange RT, Carlozzi NE. Traumatic brain injury caregivers: a qualitative analysis of spouse and parent perspectives on quality of life. Neuropsychol Rehabil 2017 Jan;27(1):16-37. [doi: 10.1080/09602011.2015.1051056] [Medline: 26052805]
7. Saban KL, Hogan NS, Hogan TP, Pape TL. He looks normal but … challenges of family caregivers of veterans diagnosed with a traumatic brain injury. Rehabil Nurs 2015;40(5):277-285. [doi: 10.1002/rnj.182] [Medline: 25424609]
8. Gulin SL, Perrin PB, Stevens LF, Villaseñor-Cabrera TJ, Jiménez-Maldonado M, Martínez-Cortes ML, et al. Health-related quality of life and mental health outcomes in Mexican TBI caregivers. Fam Syst Health 2014 Mar;32(1):53-66. [doi: 10.1037/a0032623] [Medline: 24417652]
9. Dreer LE, Elliott TR, Shewchuk R, Berry JW, Rivera P. Family caregivers of persons with spinal cord injury: predicting caregivers at risk for probable depression. Rehabil Psychol 2007;52(3):351-357 [FREE Full text] [doi: 10.1037/0090-5550.52.3.351] [Medline: 18716667]
10. Cella DF, Perry SW, Poag ME, Amand R, Goodwin C. Depression and stress responses in parents of burned children. J Pediatr Psychol 1988 Mar;13(1):87-99. [Medline: 3385579]
11. Sklenarova H, Krümpelmann A, Haun MW, Friederich H, Huber J, Thomas M, et al. When do we need to care about the caregiver? supportive care needs, anxiety, and depression among informal caregivers of patients with cancer and cancer survivors. Cancer 2015 May 01;121(9):1513-1519 [FREE Full text] [doi: 10.1002/cncr.29223] [Medline: 25677095]
12. Arango-Lasprilla JC, Quijano MC, Aponte M, Cuervo MT, Nicholls E, Rogers HL, et al. Family needs in caregivers of individuals with traumatic brain injury from Colombia, South America. Brain Inj 2010;24(7-8):1017-1026. [doi: 10.3109/02699052.2010.490516] [Medline: 20545455]

XSL•FO
RenderX

13. Dillahunt-Aspillaga C, Jorgensen-Smith T, Ehlke S, Sosinski M, Monroe D, Thor J. Traumatic brain injury: unmet support needs of caregivers and families in Florida. PLoS One 2013;8(12):e82896 [FREE Full text] [doi: 10.1371/journal.pone.0082896] [Medline: 24358236]
14. Kirk S, Fallon D, Fraser C, Robinson G, Vassallo G. Supporting parents following childhood traumatic brain injury: a qualitative study to examine information and emotional support needs across key care transitions. Child Care Health Dev 2015 Mar;41(2):303-313. [doi: 10.1111/cch.12173] [Medline: 25039833]
15. Willebrand M, Sveen J. Perceived support in parents of children with burns. Gen Hosp Psychiatry 2016;38:105-108. [doi: 10.1016/j.genhosppsych.2015.10.004] [Medline: 26596191]
16. Evercare, National Alliance for Caregiving. Caregiving. 2006 Sep. Caregivers in decline: a close-up look at the health risks of caring for a loved one internet URL: http://www.caregiving.org/data/Caregivers%20in%20Decline%20Study-FINAL-lowres.pdfwww.caregiving.org/data/Caregivers%20in%20Decline%20Study-FINAL-lowres.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmUSpTn0]
17. Longacre ML. Cancer caregivers information needs and resource preferences. J Cancer Educ 2013 Jun;28(2):297-305 [FREE Full text] [doi: 10.1007/s13187-013-0472-2] [Medline: 23553000]
18. Reinhard SC, Levine C, Samis S. AARP. 2012 Oct. Home alone: family caregivers providing complex chronic care internet URL: http://www.aarp.org/content/dam/aarp/research/public_policy_institute/health/home-alone-family-caregivers-providing-complex-chronic-care-rev-AARP-ppi-health.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmUiJaXD]
19. Fox S, Brenner J. Pew Research Center. 2012 Jul. Family caregivers online URL: http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_Family_Caregivers_Online.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmUtVKYC]
20. Fagnano M, Halterman JS, Conn KM, Shone LP. Health literacy and sources of health information for caregivers of urban children with asthma. Clin Pediatr (Phila) 2012 Mar;51(3):267-273 [FREE Full text] [doi: 10.1177/0009922811421413] [Medline: 21911409]
21. Girgis A, Lambert SD, McElduff P, Bonevski B, Lecathelinais C, Boyes A, et al. Some things change, some things stay the same: a longitudinal analysis of cancer caregivers' unmet supportive care needs. Psychooncology 2013 Jul;22(7):1557-1564. [doi: 10.1002/pon.3166] [Medline: 22941765]
22. Centers for Disease Control and Prevention. 2015. The report to congress on traumatic brain injury in the united states: epidemiology and rehabilitation URL: https://www.cdc.gov/traumaticbraininjury/pdf/tbi_report_to_congress_epi_and_rehab-a.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmVDZiXz]
23. National Spinal Cord Injury Statistical Center (NSCISC). 2014. Complete public version of the 2014 annual statistical report for the spinal cord injury model systems URL: https://www.nscisc.uab.edu/PublicDocuments/reports/pdf/2014%20NSCISC%20Annual%20Statistical%20Report%20Complete%20Public%20Version.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmVXwjPP]
24. Ameriburn. 2016. Burn incidence and treatment in the United States URL: http://www.ameriburn.org/resources_factsheet.php [accessed 2017-04-16] [WebCite Cache ID 6pmVgcVEo]
25. Ritchie J, Lewis J, editors. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Thousand Oaks, CA: SAGE Publications Ltd; 2003.
26. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
27. Longacre ML, Galloway TJ, Parvanta CF, Fang CY. Medical communication-related informational need and resource preferences among family caregivers for head and neck cancer patients. J Cancer Educ 2015 Dec;30(4):786-791. [doi: 10.1007/s13187-015-0814-3] [Medline: 25893922]
28. Dolce MC. The Internet as a source of health information: experiences of cancer survivors and caregivers with healthcare providers. Oncol Nurs Forum 2011 May;38(3):353-359. [doi: 10.1188/11.ONF.353-359] [Medline: 21531685]
29. National Alliance for Caregiving. Caregiving. 2012 Mar. Multiple sclerosis caregivers URL: http://www.caregiving.org/pdf/research/MSCaregivers2012_FINAL.pdf [accessed 2017-04-16] [WebCite Cache ID 6pmW3ZcZw]
30. Stajduhar KI, Funk L, Outcalt L. Family caregiver learning--how family caregivers learn to provide care at the end of life: a qualitative secondary analysis of four datasets. Palliat Med 2013 Jul;27(7):657-664. [doi: 10.1177/0269216313487765] [Medline: 23695826]
31. Hendrix CC, Bailey DE, Steinhauser KE, Olsen MK, Stechuchak KM, Lowman SG, et al. Effects of enhanced caregiver training program on cancer caregiver's self-efficacy, preparedness, and psychological well-being. Support Care Cancer 2016 Jan;24(1):327-336. [doi: 10.1007/s00520-015-2797-3] [Medline: 26062925]
32. Hendrix CC, Landerman R, Abernethy AP. Effects of an individualized caregiver training intervention on self-efficacy of cancer caregivers. West J Nurs Res 2013 May;35(5):590-610. [doi: 10.1177/0193945911420742] [Medline: 21949091]
33. Takizawa T, Takahashi M, Takai M, Ikeda T, Miyaoka H. Changes in job stress and coping skills among caregivers after dementia care practitioner training. Psychogeriatrics 2017 Jan;17(1):52-60. [doi: 10.1111/psyg.12188] [Medline: 26786941]

XSL•FO
RenderX

34. Tremont G, Davis JD, Papandonatos GD, Ott BR, Fortinsky RH, Gozalo P, et al. Psychosocial telephone intervention for dementia caregivers: a randomized, controlled trial. Alzheimers Dement 2015 May;11(5):541-548 [FREE Full text] [doi: 10.1016/j.jalz.2014.05.1752] [Medline: 25074341]
35. Rotondi AJ, Sinkule J, Spring M. An interactive web-based intervention for persons with TBI and their families: use and evaluation by female significant others. J Head Trauma Rehabil 2005;20(2):173-185. [Medline: 15803040]
36. Sander AM, Clark AN, Atchison TB, Rueda M. A web-based videoconferencing approach to training caregivers in rural areas to compensate for problems related to traumatic brain injury. J Head Trauma Rehabil 2009;24(4):248-261. [doi: 10.1097/HTR.0b013e3181ad593a] [Medline: 19625864]
37. Damianakis T, Tough A, Marziali E, Dawson DR. Therapy online: a web-based video support group for family caregivers of survivors with traumatic brain injury. J Head Trauma Rehabil 2016;31(4):E12-E20. [doi: 10.1097/HTR.0000000000000178] [Medline: 26291634]

Abbreviations

CDC: Centers for Disease Control and Prevention
IRR: interrater reliability
NIDILRR: National Institute on Disability, Independent Living, and Rehabilitation Research
NSCISC: National Spinal Cord Injury Statistical Center
SCI: spinal cord injury
TBI: traumatic brain injury

Edited by G Eysenbach; submitted 21.11.16; peer-reviewed by L Olabarrieta, J Ash; comments to author 12.01.17; revised version received 20.02.17; accepted 08.03.17; published 10.05.17.

Please cite as:
Coffey NT, Cassese J, Cai X, Garfinkel S, Patel D, Jones R, Shaewitz D, Weinstein AA
Identifying and Understanding the Health Information Experiences and Preferences of Caregivers of Individuals With Either Traumatic Brain Injury, Spinal Cord Injury, or Burn Injury: A Qualitative Investigation
J Med Internet Res 2017;19(5):e159
URL: http://www.jmir.org/2017/5/e159/
doi:10.2196/jmir.7027
PMID:28490418

©Nathan T Coffey, James Cassese, Xinsheng Cai, Steven Garfinkel, Drasti Patel, Rebecca Jones, Dahlia Shaewitz, Ali A Weinstein. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of eHealth Literacy on General Practitioner Consultations: A Mediation Analysis

Peter Johannes Schulz[1], PhD; Mary Anne Fitzpatrick[2], PhD; Alexandra Hess[3], PhD; Lynn Sudbury-Riley[4], PhD; Uwe Hartung[1], PhD

[1]Institute of Communication and Health, Faculty of Communication Science, Università della Svizzera italiana, Lugano, Switzerland

[2]Waikato Management School, University of Waikato, Hamilton, New Zealand

[3]School of Communication, Journalism and Marketing, Massey University, Auckland, New Zealand

[4]University of Liverpool Management School, Liverpool, United Kingdom

Corresponding Author:
Peter Johannes Schulz, PhD
Institute of Communication and Health
Faculty of Communication Science
Università della Svizzera italiana
Via Giuseppe Buffi 13
Lugano, 6900
Switzerland
Phone: 41 58 666 4724
Fax: 41 58 666 4647
Email: peter.schulz@usi.ch

Abstract

Background: Most evidence (not all) points in the direction that individuals with a higher level of health literacy will less frequently utilize the health care system than individuals with lower levels of health literacy. The underlying reasons of this effect are largely unclear, though people's ability to seek health information independently at the time of wide availability of such information on the Internet has been cited in this context.

Objective: We propose and test two potential mediators of the negative effect of eHealth literacy on health care utilization: (1) health information seeking and (2) gain in empowerment by information seeking.

Methods: Data were collected in New Zealand, the United Kingdom, and the United States using a Web-based survey administered by a company specialized on providing online panels. Combined, the three samples resulted in a total of 996 baby boomers born between 1946 and 1965 who had used the Internet to search for and share health information in the previous 6 months. Measured variables include eHealth literacy, Internet health information seeking, the self-perceived gain in empowerment by that information, and the number of consultations with one's general practitioner (GP). Path analysis was employed for data analysis.

Results: We found a bundle of indirect effect paths showing a positive relationship between health literacy and health care utilization: via health information seeking (Path 1), via gain in empowerment (Path 2), and via both (Path 3). In addition to the emergence of these indirect effects, the direct effect of health literacy on health care utilization disappeared.

Conclusions: The indirect paths from health literacy via information seeking and empowerment to GP consultations can be interpreted as a dynamic process and an expression of the ability to find, process, and understand relevant information when that is necessary.

(J Med Internet Res 2017;19(5):e166) doi:10.2196/jmir.6317

KEYWORDS

health literacy; eHealth literacy; health empowerment; information seeking; health care utilization; baby boomers

XSL•FO
RenderX

Introduction

Health Literacy and the Utilization of the Health Care System

The high attention health communication research pays to health literacy is grounded in a core group of factors positively related to health literacy. Among these factors are self-reported health status [1-3], participation in prevention and screening [4,5], frequency of exercise [6,7], a healthy diet [8,9], better management of chronic diseases [10,11], and even lower mortality for individuals with serious conditions [6,12].

Another key factor on this list is the utilization of the health care system; that is, visits to one's physician or general practitioner (GP) and other health care professionals, visits to accident and emergency facilities, admissions to hospitals, and various health care treatments. The usual implication is that high levels of health literacy are associated with low levels of using the services of the health care system, assuming that lower levels of utilization are a good thing. This assumption is based on the fact that higher utilization of the health care system is generally associated with higher costs [9,13-16].

Studies provide strong empirical evidence for the inverse relationship of health literacy and health care system utilization in different countries [17-21], across different age-groups [21,22], as well as across different patient groups [23-26]. Most researchers seem to agree that health literacy affects health care system utilization; however, there are several explanations proposed for this effect. For instance, people with low health literacy skills are understood to have limited abilities to access and understand health information, and to make decisions about their health. According to Baker et al [22], these people delay seeking care for serious health problems and have poor self-management skills; therefore, they eventually have higher rates of hospitalization and more visits to emergency clinics, both for treatment of serious conditions and for chronic care or conditions that could be more effectively managed by visiting a regular GP and early intervention [18]. Those with inadequate health literacy might not be aware of health services such as screening tests and their role in disease prevention and early diagnosis, and consequently use less preventive care [20,27] and rely more on prescription therapy than prevention [28]. Other researchers contend that patients with low health literacy can feel ashamed or distrustful of the health care system [29,30] and use more emergency medical care because they do not have a regular GP or health professional [31].

Several researchers explain the inverse correlation by findings that individuals with high health literacy know more about diseases and self-care, engage in positive health behaviors, use more preventive health care, and have better compliance with medication regimens [20,32]. Rasu et al [28] agree that high literacy likely goes along with higher use of preventive care and thus, in the long run, with lower utilization; however, upward they point out that the correlation can also be explained by the increasing use of electronic media such as the Internet to help manage symptoms and conditions.

We note that as more studies investigate the relationship between health literacy and health care system utilization in different contexts, important inconsistencies are also being reported. Some studies find no statistically significant evidence of a relationship. For instance, health literacy was not independently associated with utilization of the health care system by adults in Iran [33]. Similarly, researchers did not find health literacy was a barrier to service utilization for adults with addiction [34] or for caregivers of children with asthma [35]. Finally, some results focus specifically on the association between health literacy and utilization of a specific service within the health care system. For example, health literacy in adult patients presenting to emergency care clinics (Atlanta, United States) was related to hospitalization but not to visits with one's physician [1], whereas another study of respondents with heart failure showed no association between health literacy and hospitalization specifically [36]. Considered together, such findings highlight the urgent need for more research that investigates the nature of the health literacy-utilization correlation.

This article aimed at contributing to meet that need for a special subset of the two concepts: eHealth literacy rather than general health literacy and number of visits to one's GP as one aspect of utilization of the health system. We chose to study eHealth literacy because the Internet has become one of the primary sources for health information [37], and versatility in using this source can be expected to affect health decisions, such as the decision between seeking consultation and trying to help yourself.

The Relationship between Health Literacy and eHealth Literacy

We use the broad definition of health literacy cited frequently in the literature: "Health literacy is the degree to which individuals have the basic capacity to obtain, process, and understand basic information and services needed to make appropriate health decisions" [38]. Historically, the ability to read, comprehend, and act on health-related information related primarily to material provided by one's GP or health care organizations. However, broader contemporary definitions of health literacy acknowledge patients initiating their own searches for information, then processing that information, and applying it in their interactions with health care practitioners and with the health care system [2,37,39,40]. A consumer behavior with important implications for patient empowerment, patient responsibility, and self-care, patient-initiated information search has been facilitated by the Internet, which continues to make increasing amounts of health information available through an expanding range of Web-based channels and communication technologies. Today, the Internet is a major source of health information [41] and the impact of Web-based information on patient health behaviors is increasing [42].

The corresponding abilities and skills needed by people to educate themselves on health matters using the Internet are brought together in the concept of eHealth literacy: the "foundational skill set that underpins the use of information and communication technologies (ICT) for health" [43]. Regarded as a "metaliteracy" [44], eHealth literacy combines both specific

XSL•FO
RenderX

and general forms of literacy. Health literacy is one of the three specific sets of skills comprising eHealth literacy, along with computer literacy and science literacy. The general forms include traditional, information, and media literacies. Therefore, whereas health literacy certainly features prominently in eHealth literacy, nevertheless it is just one of six essential literacies. The eHealth literacy concept thus reflects the complexity inherent in the use of Internet information and technologies for health compared with the use of offline resources.

Finally, with health literacy and eHealth literacy presented as distinct but related and also correlated [45,46] concepts, researchers are now turning their attention to the relationships between eHealth literacy, health-related behaviors, and health outcomes [47,48]. Please note that, because health literacy is a key constituent of eHealth literacy and has a substantial literature, in this paper we draw on relevant research on health literacy as well as more recent studies of eHealth literacy to provide insights and inform our discussion.

In summary, most evidence (but by far not all) points in the direction that there is a negative relationship between the broader concept of health literacy and utilization of the health system, indicating that individuals with a higher level of health literacy will less frequently utilize the health care system than persons with lower levels of health literacy. Moreover, it is largely unclear what the underlying reasons of this relationship are no matter whether positive or negative. We assume these general findings and limitations also apply to eHealth literacy. The situation is equivocal enough to treat the relationship between eHealth literacy and utilization of one's GP services and a possible network of causal relationships behind it as open questions. Consequently, the magnitude and direction of that relationship is, as research question 1 (RQ1), "How is eHealth literacy related to the number of GP consultations a person seeks?," the starting point of our analyses.

Mediators of the eHealth Literacy-Utilization Association

Causality between 2 variables x and y can follow 4 fundamental models: x affects y, y affects x, s affects x and y, and x affects y via m (x affects m and m affects y). The first 2 models capture direct causation in either direction. The third model is a spurious correlation that traces back the original correlation between x and y to a common cause s or any number of common causes s1, s2, s3, and so on. The fourth model divides a possible causation between x and y into 2 or more steps defined by mediating variables or mediators m1, m2, m3, and so on. All explanations summarized thus far posit mediating variables to explain the correlation. This article, therefore, is also concerned with the question of whether mediators can be found that provide a possible explanation for the correlation between eHealth literacy and number of GP visits. It does not stop at positing the role of mediators; it tests some of them.

Numerous studies report links between the general "usage" of Internet health information and various health outcomes such as improved self-care, change of decision about how to treat a condition, asking for a second opinion, improved medication compliance, and less inpatient care [49]. In relation to eHealth literacy specifically, scholars find that eHealth literacy is associated with health outcomes for the individual [50-52]. However, little research has investigated the exact nature of the relationships between eHealth literacy and particular health outcomes, especially those at the wider public health level such as utilization of the health care system.

Consequently, in this study we aim to investigate the correlation of eHealth literacy and the number of GP consultations as an aspect of utilization of the health care system. We propose and test two potential mediators: (1) Internet health information seeking and (2) gain in empowerment. Previous research explains a possible effect of health literacy on utilization of the health care system according to people's ability to seek health information independently. For example, recent studies found strong evidence that a higher level of eHealth literacy was associated with an increase in Internet health information seeking behavior [53,54]. More importantly, people with low health literacy skills were less likely to properly evaluate health information presented on the Web [55-57]. We expect therefore that higher eHealth literacy skills are associated with an extended Web-based search for health information.

It is safe to assume that individuals with high eHealth literacy more often make use of this ability and draw more benefit from the information found. Research supports this assumption [58]; for example, Tennant et al [52] found baby boomers with high level of eHealth literacy use the Internet and social media (Web 2.0) for health-related purposes more than those with lower levels. That is to say, on commonsense grounds as well as corroborating research, we expect a positive relationship between eHealth literacy and intensity of Internet health information search. This expectation is our hypothesis (H1): Individuals who show higher levels of eHealth literacy will more often search for health information on the Internet than individuals with lower levels of eHealth literacy.

If persons with high levels of eHealth literacy do practice more and better self-education in health matters using the Internet, it remains less clear how that affects the utilization of the health care system. One could argue a negative relationship, assuming that health self-education by the Internet may spare a visit to the doctor because a good website might provide the help or advice the individual might have hoped to get from the GP. Self-education may also put a person in a position to make better judgments on the need to consult the doctor. If relatively inconsequential situations in which the individual considers a consultation are more frequent than more serious situations in which she or he does not consult the doctor, the better judgment might reduce the number of visits to one's doctor.

In contrast, a positive relationship could also be argued, implying that successful self-educators turn to the health care system more often. One reason for this could be that their Web-based information seeking behavior makes them aware of medical conditions that are treatable and which would otherwise have just been considered a nuisance. The second reason for a positive relationship could be the explicit advice given by many medical websites to see a doctor in case of doubt whereas explicit advice not to see the doctor occurs less frequently. Hence, the better judgment through self-education might also work in direction of a positive relationship, if not consulting

XSL•FO
RenderX

one's doctor in serious conditions is the more frequent error, than seeking consultation when it is not necessary.

As both directions are possible and to some degree plausible, we formulate the relationship as research question 2 (RQ2): How does the frequency of the Internet information seeking behavior affect the utilization of the health care system?

Although there is plenty of evidence that the Internet, with its extensive availability of health information, provides many opportunities for people with high literacy skills, less is known about what the precise consequences of the additional information might be, other than being better informed. One possible effect of having Internet access to health information is that consumers are enabled to participate in decisions regarding their health [59]. Following insights from previous studies [37,60,61], we expect that consumers looking on the Web for health information will also be more empowered in the sense that they will consider themselves to be more capable of taking the proper action once they have found useful information on the health condition.

Empowerment is usually defined as the state of having or the process of acquiring mastery over one's own life. If pertaining to health, empowerment is mastery over one's health or the health care decisions one has to face. It can be understood as an objective state but is used in the context of health care most often in a psychological sense as the person's subjective perception of mastery. Self-education might have objective consequences on empowerment, but as most health decisions in acute situations are made consciously, the subjective impression of such consequences becomes important. A person might, through the use of Web-based health content, become more enabled to describe their symptoms, but if she or he is also aware of that improved ability, it might be expected to be translated more easily into behavior.

As with the expectations relating to Internet information search behavior as a mediating factor, we formulate a hypothesis for the effect of eHealth literacy on a mediator, self-perceived gain in empowerment, and a research question for the relationship between the mediator and the ultimate dependent variable, utilization of the health care system. Again, the hypothesis (H2) primarily rests on plausibility: Persons who engage in more Internet health information seeking feel more empowered than persons who seek health information less often on the Internet.

Awareness of an empowering effect of information seeking has different components, among them perceived communicative abilities in dealing with GPs or other health care providers, and a form of self-assuredness in making health decisions by oneself and taking responsibility for one's health. Better communicative abilities can be expected to result in more benefits from consultation with the GP. This suggests a positive association between self-perceived gain in empowerment and utilization of the health care system more generally. In contrast, self-assuredness in taking responsibility for one's health rather suggests the opposite; namely, to stay away from one's doctors. In other words, the direction of the possible association is unclear, resulting in this research question (RQ3): How does self-perceived gain in empowerment from using Internet health information affect the utilization of the health care system?

To complete the rundown of expectations, we also consider the possibility that perceived gain in empowerment is directly affected by eHealth literacy. It can be assumed that persons with a high ability to find, process, and understand information on health matters will be able to draw the benefit without necessarily reaching high values for actual information seeking. There is some support in the literature for an association between eHealth literacy and concepts related to empowerment, such as self-perceived competence in finding Web-based health information [62], aptitude, and sophistication in using such information [51]. To consider this possibility we formulate another hypothesis (H3): Persons with a high level of eHealth literacy will report higher gain in empowerment than persons with lower levels of eHealth literacy.

In summary, we investigate three paths of mediating variables that could potentially explain a relationship between eHealth literacy and utilization of the health care system: (1) via intensity of Internet health information seeking, (2) via self-perceived gain in empowerment, and (3) via both, by a path leading to intensity of Internet health information seeking and then to self-perceived gain in empowerment.

Baby Boomers as Sampling Frame

Hypotheses and research questions were tested on baby boomers; that is, the age cohort born between the end of the Second World War and the advent of pharmaceutical contraceptives in the mid-1960s. This generation is particularly well suited for studying eHealth communication as they were relatively young at the time personal computers began to make their mark on our daily lives and they are now approaching the years when age-related troubles set in, making health a salient subject. Policy makers and health care service providers are particularly concerned at the costs and adequate provision of health care to baby boomers [63,64]. With advances in behavioral health and medicine, as the lifespan of baby boomers increases, so too will their lifetime health care costs and the pressure they exert as a cohort of significant size on the health care system. Moreover, there are mounting concerns about meeting the costs of health conditions such as diabetes and heart disease expected with the escalating number of overweight and obese baby boomers [64].

Baby boomers are increasingly using the Internet to search for and share health information [52]. For example, more than 88% of US baby boomers use a variety of digital devices to search for relevant Web-based health-related information and services, especially for increasing their knowledge of the prognosis, symptoms, and treatment options for personal health issues [65]. These emerging patterns of Web-based information seeking and sharing behaviors have immediate implications for health-related concepts such as health literacy and eHealth literacy, patient empowerment, patient autonomy, self-management, patient responsibility, and health outcomes [37,39]. Therefore, there is an urgent need to investigate baby boomers' Web-based health information behaviors to provide a sound empirical base for designing more effective eHealth communication with them as heavy users of the health care system.

XSL•FO
RenderX

Methods

Sample

Data for a cross-sectional study were collected over a 4-month period in New Zealand, the United Kingdom, and the United States. The questionnaire was administered on the Web-based survey platform Qualtrics. The questionnaire was designed purposefully so there would be no "missing data;" a "not applicable" response was provided for suitable questions, and incomplete questionnaires could not be submitted. Baby boomers in each country were selected using the inclusion criteria that they were born between 1946 and 1964, and that they had used the Internet to search for and share health information in the previous 6 months. The link to the questionnaire was distributed to a representative sample stratified in terms of gender, ethnicity, education, income, and location. Approval for the research was obtained by the relevant university ethics committees before the questionnaire was pretested (6 respondents) and pilot-tested (64 respondents). The operative sample included a total of 996 persons (New Zealand, n=276; United Kingdom, n=407; United States, n=313).

Average age was 59.29 (SD 5.43). A little over half of the participants (50.1%, 499/996) were females. The modal educational level was secondary school (41.5%, 413/996); 1.2% (12/996) of the participants had a lower education, whereas 32.3% (322/996) had attended university. As to ethnicity, by far most British respondents were white (95.6%, 389/407), which was also the case for 84.8% (234/276) of the New Zealand and 79.6% (249/313) of the US sample. Almost half of the respondents (44.9%, 447/996) were employed fulltime or part-time, whereas almost a third (31.2%, 311/996) were already retired. The median income was slightly below GDP 20,000 in Britain, slightly above USD 40,000 in the United States, and between NZD 35,000-40,000 in New Zealand.

Measures

As the measure of eHealth literacy, we used the eHealth Literacy Scale (eHEALS) as devised by Norman and Skinner [66]. It consists of 8 self-reported items that formulate self-perceived ability and confidence in gathering health information from the Internet. Items are scored on 5-point Likert scales with high scores indicating high agreement with the items and thus high eHealth literacy. The application of the measure produced reliable data (Cronbach alpha=.92, mean=3.69, SD=0.640, N=996).

Internet health information seeking behavior was measured by 4 items formulating different activities that are examples of Web-based information seeking. The items were: "I've looked online to try to diagnose a health condition," "I've researched a health-related product or service online," "I've read or watched someone else's commentary or experience online about health-related issues," and "I've read online reviews or rankings of health care services or treatments." The corresponding question asked about the frequency of these behaviors. Items were measured on 5-point scales ranging from "never" to "very often." The 4-item scales were averaged to achieve our measure of Internet health information seeking behavior. The scale was found to be reliable (Cronbach alpha=.80, mean=2.49, SD=0.874, N=996).

Self-perceived gain in empowerment was measured by 7 self-designed items formulating self-perceived changes attributed to the use of the Internet. The items were: "I am more aware of my health," "I feel more in control of my health," "I have a better understanding of the condition or disease I have," "I feel more connected to others with a similar problem," "I can communicate more effectively with my health professional(s)," "The quality of the relationship with my health professional(s) has improved," and "I can make better choices about the treatment of health issues." Items were scored on 5-point Likert scales with high scores indicating high agreement with the items and thus high self-perceived gain in empowerment. Using the measure produced reliable data (Cronbach alpha=.88, mean=3.59, SD=0.647, N=996).

Utilization of the health care system, our dependent variable was measured (as mentioned) by a single item inquiring about the number of medical consultations with one's GP in the past year coded as 0, 1, 2, 3, 4, 5 to 9, and 10 or more. Presence of chronic disease, recoded as a binary variable from a question inquiring about 10 different chronic diseases, was used as a control variable in the ensuing analyses.

Data Analysis

First we conducted a set of preliminary analyses including descriptive data examination, outliers, and nonnormality checks. Given this was a Web-based survey, no variables showed missing data. Second, descriptive statistics and Pearson product-moment correlation analyses were computed to determine univariate and bivariate relations among the variables in our study.

A serial mediation analysis, sometimes described as multiple-step multiple mediation [67], was conducted using the SPSS macro PROCESS [68] (model 6) with the 2 variables, Internet health information seeking behavior and self-perceived gain in empowerment, as mediators in the analysis. We used bootstrapping in the analysis to obtain bias-corrected 95% CIs for the total direct and indirect effect (ie, total mediated effect) and the specific indirect effects.

Mediation analysis used ordinary least squares path analysis. Three paths were included by which eHealth literacy may indirectly influence people to visit a GP. The first leads from eHealth literacy to GP visits via Internet information seeking behavior: those people who show higher levels of eHealth literacy are assumed to look more often for health information on the Internet and the Web-based search behavior is associated with the number of GP visits. Second, people with higher eHealth literacy levels also felt more empowered to make good health-related decisions, which in turn is assumed to affect the number of visits of the GP. Third, more eHealth literate people look for more health information on the Web, which in turn led them report a higher level of empowerment, which also is assumed to affect the number of GP visits.

XSL•FO
RenderX

Results

Significant bivariate relationships between the main study variables—eHealth literacy, Internet health information seeking behavior, self-perceived gain in empowerment, and number of GP consultations— were detected. The strongest relationships were found between Internet health information seeking behavior and perceived gain in empowerment (r=.55), eHealth literacy and perceived gain in empowerment (r=.49), and eHealth literacy and Internet health information seeking behavior (r=.40), whereas the weakest association was between eHealth literacy and utilization of the health care system (r=.09).

The paths for the full model are represented in Figures 1 and 2, the corresponding coefficients in Table 1. In the first model, number of GP consultations was predicted by eHealth literacy and the covariate chronic disease (illustrated in Figure 1), whereas the 2 mediator variables, Internet health information seeking (M_1), and self-perceived gain in empowerment (M_2), are excluded. The total association, c, between eHealth literacy and health system utilization is 0.009 (beta=.246, t_{994}=2.926, P=.004). So, levels of eHealth literacy are very weakly but statistically significantly associated positively with health system utilization. In the second step, the 2 mediators were included in the model (illustrated in Figure 2). The first indirect path of eHealth literacy through Internet health information seeking to number of GP consultations was significant and positive (a_1b_1=0.1806; 95% CI=0.0993-0.2658). The second indirect path connects eHealth literacy to GP visits through the second mediator perceived empowerment; it was also significant and positive (a_2b_2=0.1595; 95% CI=0.0898-0.2398). Finally, the third specific path runs from eHealth literacy to Internet health information seeking to perceived gain in empowerment and to GP visits and is 0.0844 (95% CI=0.0481-0.1289); that is, significant and positive also. Taking together all specific indirect effects, that are paths a_1b_1, a_2b_2, and a_1, d_{21}, b_2, the sum of all three specific indirect effects modeled amounted to 0.4245, which was different from zero as determined by the bootstrap CI which does not contain a zero (95% CI=0.3297-0.5325; Table 2). When adding the mediators to our model, the direct path from eHealth literacy to health system utilization, c, acquired a negative sign but became statistically indistinguishable from zero (P=.12)

With regard to our hypotheses and research questions, H1, H2, and H3 were all supported, meaning that higher eHealth literacy went along with more Internet health information seeking and self-perceived gain in empowerment, and that Internet health information seeking was associated with empowerment. RQ2 and RQ3 are answered in the positive: both Internet health information seeking and enhanced empowerment were associated with increased utilization of the health care system. RQ1 will be discussed in the following section.

Figure 1. Simple regression model of GP visits on eHealth literacy.

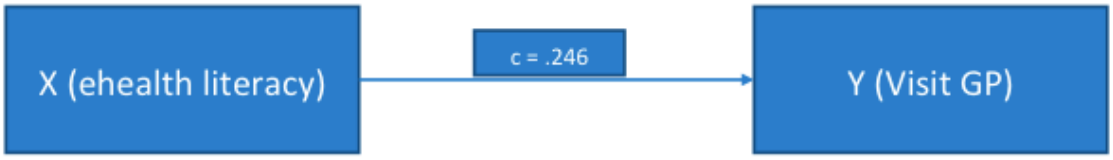
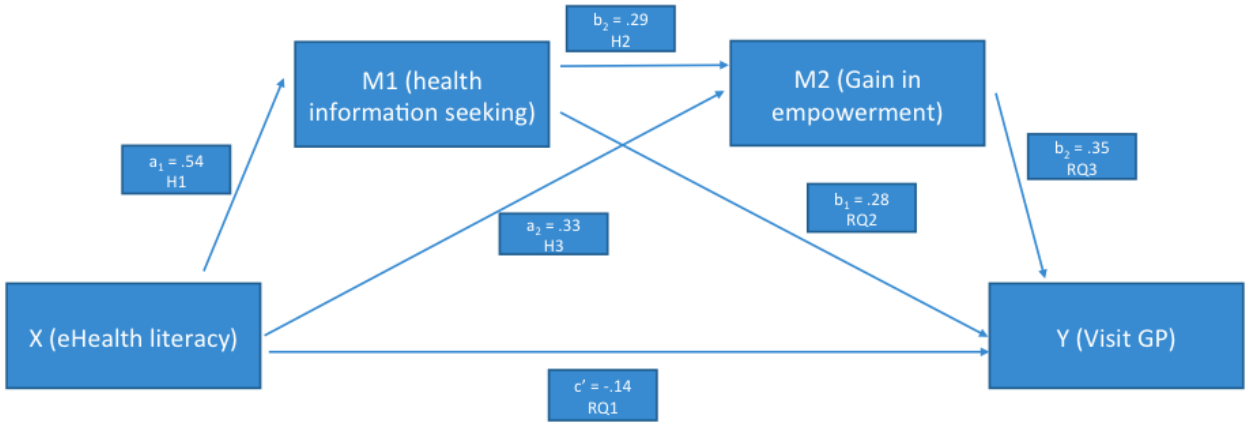


Figure 2. eHealth literacy serial multiple mediator model.



XSL•FO
RenderX

Table 1. Regression coefficients, standard errors, and model summary information for the baby boomer serial multiple mediator model.

Antecedent	Consequence								
	M_1(search behavior)			M_2(gain in empowerment)			Y (general practitioner visits)		
	Coefficient	Standard error	*P* value	Coefficient	Standard error	*P* value	Coefficient	Standard error	*P* value
X (eHealth-literacy)	.558	0.041	<.001	.324	0.027	<.001	−0.139	0.091	.12
M_1(search behavior)	–	–	–	.307	0.020	<.001	.324	0.067	<.001
M_2(gain in empowerment)	–	–	–	–	–	–	.492	0.098	<.001
Covariate health	.036	0.025	.16	−0.004	0.016	.81	.480	0.048	<.001
Constant	.435	0.173	.01	1.606	0.107	<.001	−0.223	0.365	.54
	R^2=0.1591			R^2=0.391			R^2=0.171		
	$F_{2, 993}$=93.973			$F_{3, 992}$=212.159			$F_{4, 991}$=51.009		
	P≤.001			*P*≤.001			*P*≤.001		

Table 2. Total, direct and indirect effects of eHealth literacy on general practitioner visits.

Type of effect	Effect size	95% CI
Total effect of eHealth literacy on GP[a]visits	0.2859	0.1281-0.4437
Direct effect of eHealth literacy on GP visits	−0.1386	−0.3149 to 0.0377
Total indirect effects of eHealth literacy on GP visits	0.4245	0.3297-0.5325
Path 1: eHealth literacy≥search behavior≥GP visits	0.1806	0.0993-0.2658
Path 2: eHealth literacy≥search behavior≥gain in empowerment≥GP visits	0.0844	0.0481-0.1289
Path 3: eHealth literacy≥gain in empowerment≥GP visits	0.1595	0.0898-0.2398

[a]GP: general practitioner.

Discussion

Principal Findings

Mediation analysis usually aims at explaining some covariance between 2 variables by the influence of third variables that function as mediators. One usually finds a part of the original covariance being explained by the mediator or mediators. Our analysis is unusual in that not only was a part of the positive relationship between eHealth literacy and utilization of GP health care services explained by an indirect path, but that the indirect paths explain much more variance and eliminated the direct relationship between health literacy and utilization.

To be more precise: we investigated the way how eHealth literacy might be related with utilization of the health care system. We argued that people with higher eHealth literacy skills will turn to the Internet more often for important health information relative to those who score lower on the eHealth literacy measure, which would in turn increase their gain in empowerment, which in turn will translate into a higher number of visits of the GP. That is to say, health literacy is modeled to exert an effect on utilization of the health care system indirectly through 2 mediators: Internet health information seeking, and gain in empowerment.

The first indirect effect (path 1) is the one of eHealth literacy on Internet health information seeking. Those who show higher eHealth literacy skills were also more likely to look for health information on the Internet, and the increased search was associated with more GP visits independent of gain in empowerment. Second, another indirect effect (path 2) describes how higher level of eHealth literacy leads to a higher number of GP visits through increased search for health information on the Internet, which in turn is associated with a gain in empowerment. A third indirect effect is presented by path 3, that is, the impact of higher levels of eHealth literacy on GP visits via increased empowerment. This effect is independent of Internet health information seeking. Relative to those patients who show low levels of eHealth literacy, skillful people are more likely to consider themselves also as capable to judge whether they need to have an interaction with a GP or not, which in turn is associated with more visits.

In addition to this bundle of indirect effects, we did not find a direct effect of eHealth literacy on health care system utilization. This suggests that those who are less eHealth literate but do search for health information on the Internet and consider themselves as much empowered as those with higher levels of eHealth literacy will visit their GPs as often as those who are more literate. This effect, though, is not statistically significant if one controls also for the health status (chronic disease) of people. In other words, keeping constant Internet health information seeking behavior and empowerment, the number of GP consultations is independent of eHealth literacy.

The complete model up go posits that persons with high eHealth literacy tend to see their GP more often in as much as they seek

XSL•FO
RenderX

health information on the Internet on their own or feel empowered by their Web-based information search (either directly or as a consequence of increased search behavior). This finding can be interpreted as an expression of a more dynamic element in eHealth literacy involved in the use of information and communication technologies to find relevant health information on the Web when that is necessary. It might lead to more GP visits by persons with high eHealth literacy because they have enabled themselves to make that decision.

Establishing link between eHealth literacy and utilization of health care services is the unique contribution of this research. In research into health literacy that featured an analysis structurally very similar to ours (though using performance-based measures for health literacy), Cho et al [32] found that, contrary to their expectation, neither disease knowledge, nor health behaviors, nor prevention behaviors, nor health status mediated the negative relationship between broad health literacy and utilization of the health care system. In addition to the fundamentally different results, Cho and colleagues used a different measure for utilization: hospitalization and emergency help seeking. Thus, their research does not help us to assess whether our mediators play a similar role in other studies. However, it does suggest that finding mediators of the strength we detected is unusual.

The indirect positive paths can be interpreted as indicating an improved capability in people with high eHealth literacy to distinguish serious illnesses from less consequential conditions.

In a somewhat serious health situation, people with high eHealth literacy will understand they need to act; for example, they will try to educate themselves on the Internet and consult their doctor. A heightened sense of empowerment, which is an integral part of the system of paths our results show, fits well with this kind of behavior. It might even be that, in contrast to the model presented here, a functional rather than a causal explanation of the correlation between Internet health information seeking and visits to the GP might be at work: those people with high eHealth literacy in a more serious condition already know they have to see their doctor and they use the Internet's potential for self-education as preparation, in order to optimize the consultation.

Our results also highlight the relationship between eHealth literacy and empowerment. Schulz and Nakamoto [37] have conceptualized a model of this relationship for general health literacy and empowerment that holds that the two are not necessarily linked, as is often assumed explicitly or implicitly. Applied to eHealth literacy, the empowerment model clearly addresses the concern that high levels of empowerment in combination with lower levels of eHealth literacy mean patients are likely to make decisions that are harmful to themselves. Empowerment gives them the will to make their own decisions and low eHealth literacy prevents them from choosing the right alternative. The relationship between empowerment and eHealth literacy in this study eases that concern to some degree, as does the positive association of sense of empowerment and GP visits. Both findings emphasize a much more encouraging combination in the empowerment model: high empowerment and high eHealth literacy.

Limitations

It might be considered a weakness of this study that it appears to treat eHealth literacy as a stand-in for the broader concept of health literacy in general. In addition to the arguments already mentioned in the text before, there is one more justification for referring to both health literacy and eHealth literacy. The measurement of health literacy is increasingly dominated by self-report measures that reach beyond the simple concept of functional health literacy but still correlate with the respective established measures covering that simple concept, such as the Short Test of Functional Health Literacy in Adults (S-TOFHLA). The measure of eHealth literacy is also based on self-reports, which establishes yet another link to the broader concept of health literacy.

Another limitation is that the utilization of the health care system is measured only with one item, number of GP consultations. A broader operationalization is to be achieved in the future.

Suggestions for Further Research

First, further research should try and identify more variables that might mediate the relationship between health literacy and health care utilization. Second, the interpretations forwarded in the discussion (as related to the dynamic nature of health literacy and the severity of the medical condition) should be put to empirical test.

Conflicts of Interest

None declared.

References

1. Baker DW, Parker RM, Williams MV, Clark WS, Nurss J. The relationship of patient reading ability to self-reported health and use of health services. Am J Public Health 1997 Jun;87(6):1027-1030. [Medline: 9224190]
2. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
3. Ownby RL, Acevedo A, Jacobs RJ, Caballero J, Waldrop-Valverde D. Quality of life, health status, and health service utilization related to a new measure of health literacy: FLIGHT/VIDAS. Patient Educ Couns 2014 Sep;96(3):404-410 [FREE Full text] [doi: 10.1016/j.pec.2014.05.005] [Medline: 24856447]

XSL•FO
RenderX

4. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, 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]
5. von Wagner C, Semmler C, Good A, Wardle J. Health literacy and self-efficacy for participating in colorectal cancer screening: the role of information processing. Patient Educ Couns 2009 Jun;75(3):352-357. [doi: 10.1016/j.pec.2009.03.015] [Medline: 19386461]
6. 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]
7. von Wagner C, Knight K, Steptoe A, Wardle J. Functional health literacy and health-promoting behaviour in a national sample of British adults. J Epidemiol Community Health 2007 Dec;61(12):1086-1090 [FREE Full text] [doi: 10.1136/jech.2006.053967] [Medline: 18000132]
8. Velardo S. The nuances of health literacy, nutrition literacy, and food literacy. J Nutr Educ Behav 2015;47(4):385-9.e1. [doi: 10.1016/j.jneb.2015.04.328] [Medline: 26026651]
9. World Health Organization (WHO). Health literacy: the solid facts URL: http://www.euro.who.int/__data/assets/pdf_file/0008/190655/e96854.pdf [accessed 2017-04-21] [WebCite Cache ID 6pt5f1QPd]
10. Heijmans M, Waverijn G, Rademakers J, van der Vaart R, Rijken M. Functional, communicative and critical health literacy of chronic disease patients and their importance for self-management. Patient Educ Couns 2015 Jan;98(1):41-48. [doi: 10.1016/j.pec.2014.10.006] [Medline: 25455794]
11. Lambert K, Mullan J, Mansfield K, Lonergan M. A cross-sectional comparison of health literacy deficits among patients with chronic kidney disease. J Health Commun 2015;20 Suppl 2:16-23. [doi: 10.1080/10810730.2015.1080329] [Medline: 26513027]
12. Moser DK, Robinson S, Biddle MJ, Pelter MM, Nesbitt TS, Southard J, et al. Health literacy predicts morbidity and mortality in rural patients with heart failure. J Card Fail 2015 Aug;21(8):612-618. [doi: 10.1016/j.cardfail.2015.04.004] [Medline: 25908018]
13. Lee SY, Arozullah AM, Cho YI. Health literacy, social support, and health: a research agenda. Soc Sci Med 2004 Apr;58(7):1309-1321. [doi: 10.1016/S0277-9536(03)00329-0] [Medline: 14759678]
14. Weiss BD, Palmer R. Relationship between health care costs and very low literacy skills in a medically needy and indigent Medicaid population. J Am Board Fam Pract 2004;17(1):44-47 [FREE Full text] [Medline: 15014052]
15. Eichler K, Wieser S, Brügger U. The costs of limited health literacy: a systematic review. Int J Public Health 2009;54(5):313-324 [FREE Full text] [doi: 10.1007/s00038-009-0058-2] [Medline: 19644651]
16. Mantwill S, Schulz PJ. Low health literacy associated with higher medication costs in patients with type 2 diabetes mellitus: evidence from matched survey and health insurance data. Patient Educ Couns 2015;98:1625-1630. [doi: 10.1016/j.pec.2015.07.006] [Medline: 26198546]
17. Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. J Gen Intern Med 1998 Dec;13(12):791-798 [FREE Full text] [Medline: 9844076]
18. Hardie NA, Kyanko K, Busch S, Losasso AT, Levin RA. Health literacy and health care spending and utilization in a consumer-driven health plan. J Health Commun 2011;16 Suppl 3:308-321. [doi: 10.1080/10810730.2011.604703] [Medline: 21951260]
19. Wolf MS, Gazmararian JA, Baker DW. Health literacy and health risk behaviors among older adults. Am J Prev Med 2007 Jan;32(1):19-24. [doi: 10.1016/j.amepre.2006.08.024] [Medline: 17184964]
20. Chen JZ, Hsu HC, Tung HJ, Pan LY. Effects of health literacy to self-efficacy and preventive care utilization among older adults. Geriatr Gerontol Int 2013 Jan;13(1):70-76. [doi: 10.1111/j.1447-0594.2012.00862.x] [Medline: 22672704]
21. Javadzade SH, Sharifirad G, Radjati F, Mostafavi F, Reisi M, Hasanzade A. Relationship between health literacy, health status, and healthy behaviors among older adults in Isfahan, Iran. J Educ Health Promot 2012;1:31 [FREE Full text] [doi: 10.4103/2277-9531.100160] [Medline: 23555134]
22. Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al. Health literacy and use of outpatient physician services by Medicare managed care enrollees. J Gen Intern Med 2004 Mar;19(3):215-220 [FREE Full text] [Medline: 15009775]
23. Haun JN, Patel NR, French DD, Campbell RR, Bradham DD, Lapcevic WA. Association between health literacy and medical care costs in an integrated healthcare system: a regional population based study. BMC Health Serv Res 2015;15:249 [FREE Full text] [doi: 10.1186/s12913-015-0887-z] [Medline: 26113118]
24. DeWalt DA, Dilling MH, Rosenthal MS, Pignone MP. Low parental literacy is associated with worse asthma care measures in children. Ambul Pediatr 2007;7(1):25-31 [FREE Full text] [doi: 10.1016/j.ambp.2006.10.001] [Medline: 17261479]
25. Omachi TA, Sarkar U, Yelin EH, Blanc PD, Katz PP. Lower health literacy is associated with poorer health status and outcomes in chronic obstructive pulmonary disease. J Gen Intern Med 2013 Jan;28(1):74-81 [FREE Full text] [doi: 10.1007/s11606-012-2177-3] [Medline: 22890622]
26. Marrie RA, Salter A, Tyry T, Fox RJ, Cutter GR. Health literacy association with health behaviors and health care utilization in multiple sclerosis: a cross-sectional study. Interact J Med Res 2014;3(1):e3 [FREE Full text] [doi: 10.2196/ijmr.2993] [Medline: 24513479]

XSL•FO
RenderX

27. Murray MD, Tu W, Wu J, Morrow D, Smith F, Brater DC. Factors associated with exacerbation of heart failure include treatment adherence and health literacy skills. Clin Pharmacol Ther 2009 Jun;85(6):651-658 [FREE Full text] [doi: 10.1038/clpt.2009.7] [Medline: 19262464]
28. Rasu RS, Bawa WA, Suminski R, Snella K, Warady B. Health literacy impact on national healthcare utilization and expenditure. Int J Health Policy Manag 2015;4(11):747-755 [FREE Full text] [doi: 10.15171/ijhpm.2015.151] [Medline: 26673335]
29. Baker DW, Parker RM, Williams MV, Pitkin K, Parikh NS, Coates W, et al. The health care experience of patients with low literacy. Arch Fam Med 1996 Jun;5(6):329-334. [Medline: 8640322]
30. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]
31. Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health 2002 Aug;92(8):1278-1283. [Medline: 12144984]
32. Cho YI, Lee SY, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. Soc Sci Med 2008 Apr;66(8):1809-1816. [doi: 10.1016/j.socscimed.2008.01.003] [Medline: 18295949]
33. Karimi S, Keyvanara M, Hosseini M, Jazi MJ, Khorasani E. The relationship between health literacy with health status and healthcare utilization in 18-64 years old people in Isfahan. J Educ Health Promot 2014 Jun;3:75 [FREE Full text] [doi: 10.4103/2277-9531.134910] [Medline: 25077168]
34. Lincoln A, Paasche-Orlow MK, Cheng DM, Lloyd-Travaglini C, Caruso C, Saitz R, et al. Impact of health literacy on depressive symptoms and mental health-related: quality of life among adults with addiction. J Gen Intern Med 2006 Aug;21(8):818-822 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00533.x] [Medline: 16881940]
35. Wood MR, Price JH, Dake JA, Telljohann SK, Khuder SA. African American parents'/guardians' health literacy and self-efficacy and their child's level of asthma control. J Pediatr Nurs 2010 Oct;25(5):418-427. [doi: 10.1016/j.pedn.2009.05.003] [Medline: 20816565]
36. Peterson PN, Shetterly SM, Clarke CL, Bekelman DB, Chan PS, Allen LA, et al. Health literacy and outcomes among patients with heart failure. JAMA 2011 Apr 27;305(16):1695-1701 [FREE Full text] [doi: 10.1001/jama.2011.512] [Medline: 21521851]
37. Schulz PJ, Nakamoto K. Health literacy and patient empowerment in health communication: The importance of separating conjoined twins. Patient Educ Couns 2013 Jan;90(1):4-11. [doi: 10.1016/j.pec.2012.09.006]
38. National Network of Libraries of Medicine. Health Literacy URL: https://nnlm.gov/outreach/consumer/hlthlit.html [accessed 2016-07-01] [WebCite Cache ID 6ifsgflMP]
39. Schulz PJ, Nakamoto K. Patient behavior and the benefits of artificial intelligence: the perils of “dangerous” literacy and illusory patient empowerment. Patient Educ Couns 2013 Aug;92(2):223-228. [doi: 10.1016/j.pec.2013.05.002]
40. Schulz P, Nakamoto K. Thematic section on enhancing health literacy through communication. Stud Commun Sci 2006;5(2):1-10.
41. Connolly KK, Crosby ME. Examining e-Health literacy and the digital divide in an underserved population in Hawai'i. Hawaii J Med Public Health 2014 Feb;73(2):44-48 [FREE Full text] [Medline: 24567867]
42. Manafo E, Wong S. Exploring older adults' health information seeking behaviors. J Nutr Educ Behav 2012 Jan;44(1):85-89. [doi: 10.1016/j.jneb.2011.05.018]
43. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
44. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011 Dec;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
45. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
46. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. The European Health Psychologist 2015;17(6):275-280.
47. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Jul;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
48. Yang S, Luo Y, Chiang C. The associations among individual factors, eHealth literacy, and health-promoting lifestyles among college students. J Med Internet Res 2017 Jan 10;19(1):e15 [FREE Full text] [doi: 10.2196/jmir.5964] [Medline: 28073739]
49. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19. [doi: 10.2196/jmir.1549]
50. Choi NG, Dinitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. J Med Internet Res 2013 May;15(5):e97 [FREE Full text] [doi: 10.2196/jmir.2333] [Medline: 23681083]
51. 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]

XSL•FO
RenderX

52. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
53. Chang FC, Chiu CH, Chen PH, Miao NF, Lee CM, Chiang JT, et al. Relationship between parental and adolescent eHealth literacy and online health information seeking in Taiwan. Cyberpsychol Behav Soc Netw 2015 Oct;18(10):618-624. [doi: 10.1089/cyber.2015.0110] [Medline: 26375050]
54. Lee YJ, Boden-Albala B, Jia H, Wilcox A, Bakken S. The association between online health information-seeking behaviors and health behaviors among Hispanics in New York City: a community-based cross-sectional study. J Med Internet Res 2015 Nov;17(11):e261 [FREE Full text] [doi: 10.2196/jmir.4368] [Medline: 26611438]
55. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DL. “It's got to be on this page”: age and cognitive style in a study of online health information seeking. J Med Internet Res 2015;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
56. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
57. Diviani N, van den Putte B, Meppelink CS, van Weert JC. 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]
58. Suri VR, Majid S, Chang YK, Foo S. Assessing the influence of health literacy on health information behaviors: a multi-domain skills-based approach. Patient Educ Couns 2016;99(6):1038-1045. [doi: 10.1016/j.pec.2015.12.017]]
59. Car J, Lang B, Colledge A, Ung C, Majeed A. Interventions for enhancing consumers' online health literacy. Cochrane Database Syst Rev 2011(6):CD007092. [doi: 10.1002/14651858.CD007092.pub2] [Medline: 21678364]
60. Camerini AL, Schulz PJ. Health literacy and patient empowerment: separating con-joined twins in the context of chronic low back pain. PLoS ONE 2015 Feb 13;10(2):e0118032. [doi: 10.1371/journal.pone.0118032]
61. Riva S, Camerini AL, Allam A, Schulz PJ. Interactive sections of an Internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014 Aug;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
62. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement: German translation of the eHEALS. J Media Psychol 2014;26(1):29-38 [FREE Full text]
63. Humpel N, O'Loughlin K, Wells Y, Kendig H. The health of Australian baby boomers. Australas J Ageing 2010 Mar;29(1):8-13. [doi: 10.1111/j.1741-6612.2010.00412.x] [Medline: 20398080]
64. Quine S, Carter S. Australian baby boomers' expectations and plans for their old age. Australas J Ageing 2006;25(1):3-8.
65. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health online 2013 URL: http://pewinternet.org/Reports/2013/Health-online.aspx [accessed 2016-07-01] [WebCite Cache ID 6ifuQZgKC]
66. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e2. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
67. Hayes AF, Preacher KJ, Myers TA. Mediation and the estimation of indirect effects in political communication research. In: Bucy E, Holbert RL, editors. Sourcebook for political communication research: Methods, measures, and analytical techniques. New York: Routledge; 2011:434-465.
68. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (Methodology in the Social Sciences). New York, London: The Guilford Press; 2012.

Abbreviations

GP: general practitioner
ICT: information and communication technologies

Edited by G Eysenbach; submitted 07.07.16; peer-reviewed by R Jacobs, YJ Lee, E Neter, E Chavarria; comments to author 17.11.16; revised version received 06.12.16; accepted 15.03.17; published 16.05.17.

Please cite as:
Schulz PJ, Fitzpatrick MA, Hess A, Sudbury-Riley L, Hartung U
Effects of eHealth Literacy on General Practitioner Consultations: A Mediation Analysis
J Med Internet Res 2017;19(5):e166
URL: http://www.jmir.org/2017/5/e166/
doi:10.2196/jmir.6317
PMID:28512081

XSL•FO
RenderX

©Peter Johannes Schulz, Mary Anne Fitzpatrick, Alexandra Hess, Lynn Sudbury-Riley, Uwe Hartung. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 16.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Physician Rating Websites: What Aspects Are Important to Identify a Good Doctor, and Are Patients Capable of Assessing Them? A Mixed-Methods Approach Including Physicians' and Health Care Consumers' Perspectives

Fabia Rothenfluh[1], MSc; Peter J Schulz[1], PhD

Institute of Communication and Health, Department of Communication Sciences, Università della Svizzera italiana, Lugano, Switzerland

Corresponding Author:
Fabia Rothenfluh, MSc
Institute of Communication and Health
Department of Communication Sciences
Università della Svizzera italiana
Via Giuseppe Buffi 13
Lugano, 6900
Switzerland
Phone: 41 58 666 4485
Fax: 41 58 666 4647
Email: fabia.rothenfluh@usi.ch

Abstract

Background: Physician rating websites (PRWs) offer health care consumers the opportunity to evaluate their doctor anonymously. However, physicians' professional training and experience create a vast knowledge gap in medical matters between physicians and patients. This raises ethical concerns about the relevance and significance of health care consumers' evaluation of physicians' performance.

Objective: To identify the aspects physician rating websites should offer for evaluation, this study investigated the aspects of physicians and their practice relevant for identifying a good doctor, and whether health care consumers are capable of evaluating these aspects.

Methods: In a first step, a Delphi study with physicians from 4 specializations was conducted, testing various indicators to identify a good physician. These indicators were theoretically derived from Donabedian, who classifies quality in health care into pillars of structure, process, and outcome. In a second step, a cross-sectional survey with health care consumers in Switzerland (N=211) was launched based on the indicators developed in the Delphi study. Participants were asked to rate the importance of these indicators to identify a good physician and whether they would feel capable to evaluate those aspects after the first visit to a physician. All indicators were ordered into a 4×4 grid based on evaluation and importance, as judged by the physicians and health care consumers. Agreement between the physicians and health care consumers was calculated applying Holsti's method.

Results: In the majority of aspects, physicians and health care consumers agreed on what facets of care were important and not important to identify a good physician and whether patients were able to evaluate them, yielding a level of agreement of 74.3%. The two parties agreed that the infrastructure, staff, organization, and interpersonal skills are both important for a good physician and can be evaluated by health care consumers. Technical skills of a doctor and outcomes of care were also judged to be very important, but both parties agreed that they would not be evaluable by health care consumers.

Conclusions: Health care consumers in Switzerland show a high appraisal of the importance of physician-approved criteria for assessing health care performance and a moderate self-perception of how capable they are of assessing the quality and performance of a physician. This study supports that health care consumers are differentiating between aspects they perceive they would be able to evaluate after a visit to a physician (such as attributes of structure and the interpersonal skills of a doctor), and others that lay beyond their ability to make an accurate judgment about (such as technical skills of a physician and outcome of care).

(J Med Internet Res 2017;19(5):e127) doi:10.2196/jmir.6875

XSL•FO
RenderX

KEYWORDS

physician rating websites; quality of care; health care consumers; judgment; physicians; doctors; Delphi technique; cross-sectional study; ethics; assessment; electronic word of mouth

Introduction

Electronic Word of Mouth in Health Care

The Internet has created a space for interaction where anyone with a Web connection can freely voice his or her opinion on any conceivable subject. One special feature of this public sphere is rating platforms that allow consumers to review goods and services purchased on and offline [1]. This has put power into consumers' hands who can instantly issue their commendation or crushing verdict to a mass audience. Using electronic word of mouth (eWOM) when purchasing a new good or service, comparing features, prices, and reviews has become commonplace for many consumers [2]. They especially appreciate recommendations from former clients to inform their choice of the so-called experience or credence goods and services [3,4], where consumers lack expertise or experience [5]. For certain industries such as tourism or gastronomy, where this development appears to be largely accepted, receiving positive Web-based recommendations through eWOM has become crucial to stay in business (see [6]).

This development has also found its way into health care. The so-called physician rating websites (PRWs) are numerously available either as separate rating websites exclusively for physicians, as part of a larger review platform where also other goods and services can be appraised, or even as part of a hospital's staff directory website. This culture of reviewing in health care developed in parallel to a change in the relationship between patients and physicians. The traditionally paternalistic rapport between the doctor and patient has transformed into a more patient- or client-centered approach, whereby passive patients have become active health care consumers [7,8].

This change in the doctor-patient–relationship was also backed by initiatives and corresponding research efforts. Patient empowerment [9] has become a globally fostered goal through charters founded by patient organizations [10] or conferences held by the WHO [11]. Simultaneously, concepts to foster patients' involvement in the health care process, such as shared decision-making with the physician [12,13], patient self-efficacy [14], and autonomy [15] have been promoted in research and practice to demolish the traditionally hierarchical relationship between a physician and a patient.

The movement away from a paternalistic model to a more egalitarian collaborative encounter has changed today's relationship between a physician and a patient [16]. PRWs take this movement to the next level: no longer are physicians and patients collaborative entities, but the latter is becoming a consumer of the former's service with the power to publicly review and assess the health care service received. The traditional paternalistic relationship between the provider and the consumer [8,12,17] and a vast knowledge gap between the two parties created by professional training and experience [18] led health care providers to voice skepticism toward ratings of hospitals and physicians [19].

Health care providers worry about unjustified negative reviews [20-22] because physicians' efforts to get negative reviews on PRWs deleted are rarely successful. Suing review platforms may even gain more public attention than keeping a negative review on one's profile [23]. Hence, some doctors who were reviewed negatively even breached their oath to treat patient information with confidentiality by leaking private details about the patients' care when replying to comments on PRWs [24,25].

Despite the debatable nature of physician rating websites due to the anonymity of the posted reviews, the low number of reviews per physician, and the lacking legal regulations [20,23,26], patients keep reviewing their doctors. Simultaneously, PRW users continue to read reviews and consider them in their choice of a health care provider [27,28]. The access to PRW reviews is a two-edged sword; it may reveal transgressions and ineptitude of health care providers that may have stayed in the dark otherwise, but could also make unjustified malpractice allegations to providers just because the consumer lacks knowledge to evaluate the health care encounter accurately.

Experts and researchers pose a different perspective, arguing that PRWs could aid the creation of more transparency and surveillance of the health care system [29,30]. PRWs could, for example, provide instant feedback if quality of care is alarmingly deteriorating in a certain care facility [31]. A study with German physicians on their use of PRWs showed that this may bear some fruit; more than half of the participating doctors used PRWs for the improvement of their quality of care, particularly to ameliorate their communication with the patient and the scheduling of appointments [32]. Given the tension between advocates and opponents of PRWs, research is needed to address from both physicians' and health care consumers' perspectives how the quality of a physician and the treatment he or she provides could be operationalized into sensible indicators for physician rating websites.

Theoretical Background

Defining and measuring quality of care has a long history due to its complexity. The World Health Organization describes quality in health care as effective, efficient, accessible, acceptable, patient-centered, equitable, and safe [33]. Due to its latent qualities, scholars have developed various models to operationalize the concept of quality care. Donabedian, a prominent scholar in the field, divides quality of care into pillars of structure, process, and outcome [34-36]. According to Donabedian, "structure denotes the attributes of the setting in which care occurs" (p. 1745), such as material and human resources and the organizational structure [35]. The process of care is concerned with the way care is provided, such as how a diagnosis is made or a treatment is executed by the health care provider [35]. The third pillar, outcome of treatment or care encounter, denotes the effects that the treatment has had on the patient such as the improvement in health status or a change in patient's behavior toward a cure-inducing lifestyle [35].

XSL•FO
RenderX

The Donabedian model is analogous to the division Maribito suggested in his categorization of health care into aspects of search, experience, and credence properties [37]. In health care, search aspects could be translated into features of infrastructure of the practice facilities as they are transparently comparable even before a physician visit. The treatment process and outcomes of the received care could either be categorized as experience or credence aspects; the experience of a treatment encounter may allow health care consumers to assess some aspects such as the interpersonal demeanor of a physician. Other aspects of more technical nature may not even be assessable after a health care consumer experienced them, making them credence traits. The application of a search, experience, and credence model to health care is not novel [37]. However, how the model on search, experience, and credence could concretely be translated into the assessment of physicians has not yet been answered. Hence, the question emerges which aspects of care can be evaluated and whether physicians and health care consumers agree on that. In order to answer these questions, samples of experts (physicians) and health care consumers were studied. This study's objectives are summarized in the following research questions: What aspects do physicians consider to be important to identify a good physician? To what degree do health care consumers agree with the indicators suggested by physicians to identify a good physician? and In which aspects or dimensions and to what extent do physicians and health care consumers agree that these aspects can be evaluated by patients after a physician visit?

Methods

Step 1: Delphi Study With Physicians

This study was divided into two parts: a Delphi study with physicians and an electronic questionnaire with adult health care consumers (see Multimedia Appendices 1 and 2). Before the study launch, ethical approval was obtained by the ethics committee of the Università della Svizzera italiana (CE 2015-8).

The Delphi study was conducted over two rounds consulting guidelines published by von der Gracht [38]. The sample comprised Swiss physicians from 4 different specializations (general physicians, pediatricians, orthopedic surgeons, and dentists). These specializations were chosen by the research team due to the diversity of services and treatments they provide, the distinct skills needed in these specializations, the different audiences served, and because some of the treatments these doctors offered are not covered by basic health insurance in Switzerland (dentist visits are not covered). First, indicators of quality of care were theoretically derived from the Donabedian quality of care model and work that extended his classification [35,39], pretested with 6 physicians and adapted where necessary. This yielded 43 indicators on the basis of the quality of care dimensions (1) structure (infrastructure and staff at the hospital or practice), (2) process (technical and interpersonal skills of the physician), and (3) outcome of care.

In the first round of the Delphi study, doctors were asked to rate each of these indicators twice: (1) how important that indicator was to identify a good physician of his or her specialization (1=not at all important; 5=very important), and (2) how well a patient could evaluate this quality of care indicator after a first physician visit (1=not at all; 5=very well). After each section of the survey, doctors were provided with a blank space to comment on the questioned indicators and to add new indicators that were not included in the questionnaire yet. Data for the first round was collected from October to December 2015.

A total of 120 physicians were invited via email (with information about the study and a link to the survey) or through a collaboration partner at the Central Switzerland Physician Association to participate in the Delphi study. Of the total, 29 physicians consented to participate and all joined in the first Delphi round. Nineteen physicians completed the second round as well. The sample in the first round consisted of 9 general practitioners (33.3%), 5 pediatricians (17.9%), 6 orthopedic surgeons (21.4%), 7 dentists (25.0%), and 1 internist (3.6%). Participants were predominantly male (n=19, 66.6%), aged between 29 and 63 years (mean 47.6 years, SD 9.6 years), and had on average 19.6 years of work experience (range 3-37 years, SD 10.1 years).

The results from the first round of the Delphi questionnaire were analyzed using SPSS statistical software 23.0 (IBM Corp). We consulted guidelines published by von der Gracht [38] to define decision criteria to establish whether participants reached agreement. The decision criteria yielded the following categorization: (1) indicators to which a two-third majority (66.66%) or more of physicians agreed that they were important (scored 4 or 5 on the 5-point scale); (2) indicators, which at least two-thirds (66.66%) scored as unimportant (ie, scores of 1 or 2); and (3) remaining indicators that were either scored mainly on a middle value (ie, 3) or highly scattered. Indicators categorized in the second category were excluded from the second Delphi round (n=5), whereas the ones categorized into groups 1 and 3 were adjusted or rephrased where necessary and presented again (category 1 for confirmation and stability and category 3 for clarification). For each of the debated indicators, the modes and corresponding percentages of votes of the first round's results were presented to the participants. Physicians were asked to choose the one value among the results from round 1 that they most agreed with. They were given free text space to explain their responses. The same logic and classification criteria were applied to the answers about whether these indicators can be evaluated by patients after a first visit. Data for the second round of the Delphi study was collected between March and April 2016. The results of the second Delphi round were analyzed using the same criteria and classification listed above (see Multimedia Appendices 1 and 2). Physicians reached agreement on 35 indicators applying the same criteria as after round 1. Stability after two rounds was reached for 32 indicators in terms of importance, and for 27 indicators in terms of evaluation by health care consumers.

Step 2: Cross-Sectional Survey With Health Care Consumers

Subsequently, a cross-sectional e-questionnaire with health care consumers was launched surveying the same indicators that were retained for the second round Delphi questionnaire (see Multimedia Appendices 1 and 2). Participants were asked to rate on 5-point scales (1) how important the listed indicators

XSL•FO
RenderX

were for them to identify a good physician and (2) to what extent they would feel capable of evaluating an indicator of care after the first visit with a physician. The questionnaire was pretested over 3 rounds with 5-10 participants each. Survey design and layout features were adjusted where necessary.

The survey was launched via Qualtrics between April and June 2016 via snowball sampling through email and social media. Eligible for participation were individuals who were (1) 18+ years old, (2) residing in Switzerland, (3) and had Internet access. A total of 211 participants completed the survey. Participants were predominantly female (63%), on average 38.24 years old (range 19-74 years, SD 15.51), and the vast majority was of Swiss nationality (188/211, 89.1%). The sample was well-educated including 53.9% respondents with some form of tertiary education (PhD: 12/211, 5.7%; university degree: 74/211, 35.1%; applied science university degree: 29/211, 13.7%; high school degree: 33/211, 15.6%; apprenticeship: 55/211, 26%; secondary school: 3/211, 1.4%; and other: 5/211, 2.4%).

The vast majority of participants had already used review websites (82.9%), with Tripadvisor being the best known and the most used (66.4%), followed by Booking (58.8%), Amazon (49.8%), and Ricardo (48.8%). The majority of health care consumers had a neutral opinion toward physician rating websites (48.8%), whereas 36.0% were in favor of being able to write and access physician ratings online and 15.2% were opposed. Only about a tenth of participants worked in a medical profession (19/211, 9.0%).

Step 3: Comparison Between Physicians' and Health Care Consumers' Perceptions

In a third step, agreement between physicians and health care consumers was calculated. The results of the second round of the Delphi study and the electronic health care consumer survey were ordered and categorized for (1) importance and (2) health care consumers' evaluation capability. First, the health care consumer data were cleaned. Means, medians, and standard deviations were computed. Then, for the purpose of categorization, all indicators were recoded from 5- to 3-point scales (1 and 2=not important or not evaluable; 3=unsure or debatable; 4 and 5=important or evaluable). If the following decision criteria were met, the indicator was categorized to be important: (1) if at least two-thirds (66.66%) of the sample voted for an indicator to be important or very important, (2) if criterion 1 was not met, the mean of the indicator on the 5-point scale had to be above 3.5 and the standard deviation less than 1. The same decision rules were applied to assess whether an indicator was evaluable by health care consumers. The first criterion was adapted from the Delphi study, whereas the second criterion was constructed to assert that the opinions did not diverge much.

On the basis of those results, indicators were ordered into 1 of 4 categories for both the physician sample and the health care consumer sample: (1) indicator is important and can be evaluated, (2) indicator is important but cannot be evaluated, (3) indicator is not important but can be evaluated, and (4) indicator is not important and cannot be evaluated by health care consumers. Indicators were ordered into a 4×4 matrix with physicians' assessment represented on the horizontal and health care consumers' on the vertical dimension (see Table 1). We then calculated the agreement on importance and evaluation capability of health care consumers based on the physicians' and the health care consumers' ratings.

Results

Delphi Study and Indicator Development

The Delphi study yielded 35 indicators that fulfilled the consensus criteria derived from the literature (see [38]). The comments from physicians were worked into the analysis and provided additional insight into physicians' perceptions, some of which are listed later in this section. Physicians unanimously agreed that all indicators of interpersonal competence were very important. The technical aspects of care and the outcome indicators were also rated highly, but the overall agreement was lower.

Infrastructure however was overall assessed to be slightly less important; especially aspects of practice management and competence of the medical assistants were deemed less relevant. The physicians who scored the management and staff less important commented that from their point of view, practice organization and particularly the quality of the staff and medical practice assistants did not provide information about the quality of a physician. They argued that the quality of the physician, the quality of his staff, and the practice management should not be mixed up. From their point of view, the quality of a physician was independent of the other listed aspects. Hence, they scored these indicators to be less relevant than some of their colleagues who perceived components of management, staff, and infrastructure as integral aspects to recognize a good physician.

Four indicators, namely, efficiency, hygiene standards, patient satisfaction, and the presentation of an appropriate number of treatment options were not unanimously voted to be important. Specifically mentioned was that TARMED, the mandatory Swiss cost calculation system, specifies the rate of ambulatory treatments to standardized levels in Switzerland (see [40]) and thus regulates efficiency by law. Also, respecting hygiene standards was not unanimously accepted as a quality standard from doctors' perspective because hygiene is a prerequisite for any practice in Switzerland as they have to pass quality assessments (such as EQUAM [41] or QBM [42]) to stay open. Also patient satisfaction, an outcome measure, did not achieve undivided agreement among doctors. Physicians voiced concerns that this measure could not necessarily be an end goal, especially if the patient requested a treatment that was not justifiable by best practice guidelines. Hence, they deemed patient satisfaction a double-edged sword: important but not at the cost of a correct treatment. Furthermore, doctors stated that offering a patient an appropriate number of treatment options may overwhelm some health care consumers. If a patient was overstrained with many options, the quality of the decision would be lower.

Agreement Between Physicians and Health Care Consumers on Importance and Evaluation Capability of Care Aspects

Health care consumers' answers on which indicators are important to identify a good physician and on whether these are

XSL•FO
RenderX

evaluable after the first visit were compared with the answers of the experts (physicians). The results for both dimensions are listed in detail for both health care consumers and physicians: Multimedia Appendix 1 shows how important the indicators were scored by the two samples, and Multimedia Appendix 2 presents the results on health care consumers' capability to assess those aspects after the first care encounter. Indicators are listed according to the Donabedian model: Structure (1 infrastructure, 2 staff, 3 organization), process (4 technical skill, 5 interpersonal skill), and outcome of care (6 outcome). All indicators were classified into a 4×4 matrix in order to visualize agreement between physicians and health care consumers (see Table 1).

On the basis of our coding, 26 out of 35 indicators were assessed the same way by health care consumers and physicians (Table 1). Agreement between the two parties according to Holsti's method [43] was calculated at 0.7429 if the assessments of importance and evaluability are both considered together. As this number is larger than 0.70, the agreement between physicians and health care consumers is fairly high.

Looking at the data more closely, the aspects of physician performance on which both doctors and health care consumers agreed that they were both important and evaluable by health care consumers included: infrastructure, organization, and management of the practice, quality, education of, and collaboration among the staff, and interpersonal demeanor of the doctor. Physicians and health care consumers also agreed on 8 aspects that were assessed by both parties as important, but not evaluable by health care consumers after the first visit. Mainly, the technical ability and skills of the physician (such as the way the physician made the diagnosis, whether treatment steps were correctly executed, and whether the doctor follows hygiene guidelines) and outcome measures (for example, whether the treatment was efficient) were assessed important, yet not assessable by health care consumers.

Disagreement between physicians' and health care consumers' judgments primarily emerged concerning health care consumers' capability to assess certain medical components of care. Health care consumers differed from physicians in thinking that they were not able to tell whether they were diagnosed correctly, whether the diagnosis was initiated timely and treatment started swiftly, and whether their concerns were treated confidentially. Physicians however judged health care consumers as incapable of assessing whether the necessary diagnostic instruments were available in the practice and if the doctor already had a lot of work experience.

Investigating the different dimensions of care by calculating the overall means, the results of the health care consumers show that they judged the technical skills of the physician to be the most important (mean 4.70, SD 0.37), followed by the interpersonal skills (mean 4.64, SD 0.38), and outcome of care (mean 4.46, SD 0.48). Infrastructure (mean 4.20, SD 0.56), organization (mean 4.00, SD 0.72), and the quality of the staff (mean 3.89, SD 0.57) were judged to be less important.

In terms of health care consumers' self-perceived capability to assess a physician, they attributed themselves the highest competence to assess the organization and management of the practice (mean 4.17, SD 0.79), followed by a physician's interpersonal skills (mean 4.09, SD 0.66), and the infrastructure and accessibility (mean 4.07, SD 0.67). The outcome of care (mean 3.85, SD 0.74), the quality of the staff (mean 3.70, SD 0.63), and technical skills of the doctor (mean 3.46, SD 0.79) were perceived to be more difficult to judge.

Table 1. A 4×4 matrix classification of indicators denoting importance to identify quality of health care and health care consumers' evaluation capability.

Indicators[a]		Physicians			
		Important and evaluable	Important but not evaluable	Not important but evaluable	Not important and not evaluable
Health care consumers	Important and evaluable	1.4, 2.2, 2.5, 3.1, 3.2, 3.3, 3.4, 4.5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.7, 5.8, 6.1, 6.2, 6.3	1.2, 2.4		
	Important but not evaluable	4.1, 4.9, 5.6	2.6, 4.2, 4.3, 4.4, 4.6, 4.7, 4.8, 6.4		
	Not important but evaluable	1.1, 1.3, 2.1			
	Not important and not evaluable		2.3		

[a]The indicators that were categorized into the graph above are numerically listed in Multimedia Appendices 1 and 2.

XSL•FO
RenderX

Patients' Ratings of Their Ability to Evaluate Aspects of Health Care by Gender and Educational Level

Independent sample means *t* tests [44] were conducted to check whether there were differences in the health care consumer sample based on the sociodemographic variables. As we conducted multiple tests, we applied the Holm-Bonferroni method and report sequential corrected *P* values (at alpha=.05) to control for type-1 error [45,46]. Only significant results were reported (see Tables 2-5). In terms of gender, women scored 7 aspects of care higher or more important to identify a good physician than men (Table 2). Women scored aspects such as privacy, cleanliness and hygiene, information provision, the presentation of treatment options, and empathy to be more important than men did. Also, in terms of self-perceived capability to assess aspects of care, women scored higher than men in one aspect of care. Namely, women perceived to be better capable of assessing cleanliness and hygiene. Overall, this leads to the conclusion that women generally assign higher scores than men both in terms of importance and tend to have a slightly higher self-perceived capability to assess aspects of care.

Table 2. Gender differences in aspects that were deemed important by health care consumers (independent samples *t* tests). Standard deviations appear in parentheses below means.

Indicator		Gender			
		Male	Female	*t*	*df*
1.3	The patient's privacy is guaranteed	4.24 (0.91)	4.67 (0.68)	−3.569[a]	127.56
1.4	The practice is clean and hygienic	4.58 (0.64)	4.86 (0.39)	−3.630[b]	110.78
4.4	The physician presents an appropriate and complete number of treatment options to the patient	4.31 (0.84)	4.69 (0.54)	−3.616[b]	114.59
4.5	The physician assesses the patient's handicaps correctly and presents him or her with appropriate information and treatment options	4.35 (0.77)	4.62 (0.60)	−2.740[a]	131.51
4.7	The physician and his team adhere to hygiene guidelines	4.63 (0.65)	4.89 (0.34)	−3.280[a]	102.49
5.3	The physician comprehensibly communicates all important information about the diagnosis and treatment	4.71 (0.51)	4.87 (0.36)	−2.644[a]	126.01
5.7	The physician treats the patient empathically	4.29 (0.80)	4.63 (0.60)	−3.255[a]	128.52

[a]*P*<.05.

[b]*P*<.001.

To assess if individuals with differing educational levels have divergent opinions on what aspects of care are important to identify a good doctor and whether they are assessable by health care consumers, the sample was subdivided into 2 groups. Individuals who completed tertiary education (ie, PhD, university, and applied science university) were denoted as "high education," whereas individuals without tertiary education (ie, high school, apprenticeship, and secondary school) were classified as "low education."

Table 3. Gender differences in health care consumers' self-perceived capability to evaluate a doctor (independent samples *t* tests). Standard deviations appear in parentheses below means.

Indicator		Gender			
		Male	Female	*t*	*df*
4.7	The physician and his team adhere to hygiene guidelines.	3.38 (1.13)	3.90 (1.09)	−3.279[a]	209

[a]*P*<.05.

The results of independent samples *t* tests show that individuals with lower education scored 9 items as significantly more important than individuals with higher education (eg, experience and friendliness of staff, correct execution of treatment steps, empathy, and patient involvement in the treatment process). When asked about their self-perceived ability to assess aspects of care after the first visit, individuals with lower education perceived themselves to be significantly better capable of assessing whether a timely diagnosis was made and the treatment swiftly initiated, as well as whether their concerns were treated confidentially. Participants with a higher educational background only scored significantly higher than individuals with low education in terms of capability to assess if decisions about the course of action were made in collaboration between the physician and the patient.

XSL•FO
RenderX

Table 4. Educational differences in terms of importance (independent samples *t* tests). Standard deviations appear in parentheses below means.

Indicator		Educational level			
		Low	High	*t*	*df*
2.2	The medical practice assistants are helpful and friendly	4.54 (0.56)	4.23 (0.95)	2.941[a]	190.624
2.3	The medical practice assistants are experienced in their work	3.55 (1.00)	3.22 (1.01)	2.355[a]	204
2.5	The physician collaborates well with his team	4.51 (0.57)	4.18 (0.80)	3.256[a]	204
3.4	The physician is available for phone consultations (before or after the appointment)	4.23 (0.79)	3.86 (0.99)	2.906[a]	204
4.6	The physician and his team execute the treatment steps correctly	4.82 (0.41)	4.11 (0.61)	2.533[a]	199.250
5.5	The physician motivates the patient to actively take part in the treatment process	4.52 (0.64)	4.22 (0.88)	2.831[a]	202.718
5.7	The physician treats the patient empathically	4.66 (0.62)	4.38 (0.73)	2.938[a]	203.110
6.2	The patient is satisfied with the treatment.	4.74 (0.491)	4.49 (0.754)	2.863[a]	197.191
6.3	The patient returns to the same physician for check-ups, etc (patient loyalty).	4.62 (0.491)	4.49 (0.754)	3.266[b]	203.841

[a]*P*<.05.
[b]*P*<.001.

Table 5. Educational differences in terms of evaluation capability (independent samples *t* tests). Standard deviations appear in parentheses below means.

Indicator		Educational level			
		Low	High	*t*	*df*
4.9	The physician makes the correct diagnosis timely and initiates the treatment swiftly	3.58 (1.08)	3.22 (1.02)	2.485[a]	204
5.6	The patient's concerns are treated confidentially	3.22 (1.26)	2.81 (1.40)	2.183[a]	204
5.8	Decisions about the course of action are made together with the patient	4.14 (1.03)	4.43 (0.76)	−2.261[a]	161.254

[a]*P*<.05.
[b]*P*<.001.

XSL•FO
RenderX

Table 6. Pearson's correlation coefficient on importance of indicators to identify a good physician and age (N=211).

Indicator		Age
2.3	The medical practice assistants are experienced in their work	.241[b]
2.4	The physician has a lot of work experience and practices already for a longer time	.214[b]
3.1	The waiting time until the next available appointment is short	.238[b]
3.2	The patients are notified in case of appointment delays or cancellations	.151[a]
3.3	It is easy to schedule an appointment with the physician	.154[a]
5.5	The physician motivates the patient to actively take part in the treatment process	.150[a]
5.8	Decisions about the course of action are made together with the patient.	.148[a]

[a]*P*<.05.

[b]*P*<.001 (2-tailed).

Health Care Consumers' Assessment of Importance and Perceived Capability to Evaluate a Physician Based on Age

The health care consumer data were further analyzed to check whether there is a relationship between age and the importance health care consumers' attribute to certain aspects of care and their self-perceived capability to assess health care providers. The results from the correlation analysis reveal that with older age health care consumers perceive aspects of organization and shared decision making as more important (Table 6). At the same time, the older health care consumers are, the more they perceive aspects of care assessable, particularly the organization of the practice, the physicians' technical competency (eg, correct diagnosis and treatment execution, timely diagnosis, etc), as well as the efficiency of the treatment (Table 7).

Table 7. Pearson's correlation coefficient on health care consumers' perceived capability to assess aspects of health care and age (N=211).

Indicator		Age
1.1	The physician's practice is easily accessible and reachable by public transport and car	−.149[a]
2.3	The medical practice assistants are experienced in their work	.257[b]
2.4	The physician has a lot of work experience and practices already for a longer time	.179[b]
2.5	The physician collaborates well with his team	.223[b]
4.1	The patient is diagnosed correctly	.202[b]
4.2	The physician asks the relevant questions and orders the correct tests to reach the correct diagnosis	.189[b]
4.3	The physician proceeds systematically and competently to reach the correct diagnosis	.168[a]
4.7	The physician and his team execute the treatment steps correctly	.224[b]
4.9	The physician makes the correct diagnosis timely and initiates the treatment swiftly	.173[a]
5.8	Decisions about the course of action are made together with the patient	.171[a]
6.4	The treatment was efficient (ie, cost-benefit ratio was accurate)	.142[a]

[a]*P*<.05.

[b]*P*<.001 (2-tailed).

Discussion

Principal Findings

Our results show that a majority of health care consumers and physicians categorized 26 out of 35 indicators similarly in terms of both importance to identify a good doctor and patients' perceived competence to evaluate them after the first visit. The data show that the majority of indicators were assessed to be both important and able to be evaluated by health care consumers, thereby creating limited variance. This occurred because indicators that physicians agreed were not important in the first round of the Delphi study and were excluded in the second round. Also, these items were not presented to health care consumers, which may explain some of this lacking variance. Nevertheless, the data provide an initial indication that health care consumers have a moderate self-perceived ability to assess the quality and skill of a medical doctor.

Looking more closely into the care aspects that are deemed assessable, we found that health care consumers and physicians judged the formers' ability to evaluate the infrastructure,

XSL•FO
RenderX

organization, and physician's interpersonal behavior to be high. These aspects of quality of care lend themselves for evaluation by health care consumers because they do not require medical expertise to be assessed. In terms of aspects that cannot be assessed by health care consumers, our results showed that health care consumers and physicians had reservations toward patients' ability to assess a doctor's technical skills or the outcome of care. Specifically, the quality of a physician's education, the process of reaching the correct diagnosis, the execution of the treatment (competence, hygiene, and efficiency), and the presentation of treatment options was mutually accepted to be crucial but not assessable by health care consumers.

These results suggest that health care consumers acknowledge and are aware of the gap in expertise between doctor and patient that arises based on doctors' medical education [18], even though the numerous ethical concerns previously voiced in literature have suggested otherwise [20,21,47]. However, whether health care consumers would be cautious or even refuse to assess a physician's competence or technical skill when asked to review a doctor in a real-life remains questionable. An analysis of open-ended textual reviews on a German PRW found that in 63% of the 3000 analyzed cases, PRW users assessed physicians' competence, a technical aspect of care [48]. Explaining this contradiction, and whether the gap between the intention to review technical aspects of care and the actual reviewing behavior could provide further insights into these incongruent findings, may be the subject of further research.

Physician selection research shows that technical aspects are often identified as the paramount criteria when health care consumers have to select a doctor [49-51]. The results in this study confirmed that health care consumers perceive technical skills of a physician to be the most important to recognize a good physician. Nonetheless, it has been confirmed that most health care consumers are not capable of using and accurately interpreting medical or technical quality of care reports to inform their physician selection [52-54]. Hence, data presentation formats that take into account that health care consumers have difficulty to assess and interpret technical aspects of care, and hence translate quality of care data in an understandable manner, are needed [52,53]. This study suggests that health care consumers are not a good source to provide this kind of information.

The results further show that for 6 of these 26 indicators, which mainly concern the physicians' competence to reach and execute a diagnosis, should not be reviewed by health care consumers because they lack competence to do so. Mixed information sources to report different aspects of care quality, combining patient reviews as a complement to customary quality reports have already been suggested by Verhoef and colleagues [26]. A large-scale experiment by Schlesinger and colleagues attempted to do that. They presented PRWs featuring quality of care data in combination with written reviews. However, the combined format did not yield better physician selection results, especially if choices grew more complex with larger choice sets and more indicators and information present [55]. Hence, finding PRW formats in which health care consumers can voice their opinion on aspects that are deemed assessable, while condensing and summarizing technical quality of care information in a format that is understandable by health care consumers (as suggested by Hibbard et al [54]) should be the subject for further research.

In the analysis on differences on the perceived importance and evaluation capability based on gender and educational level, two patterns emerged: women and lower educated individuals rate indicators higher or more important and perceive aspects of care as better evaluable than men and individuals with higher educational background. Given that women are more affine toward health issues and search for health information more eagerly [56] and are more aware and likely to use PRWs than males [28], this experience with health information on the web may lead to a perception of expertise. This hypothesis should be tested in future research.

Furthermore, our results suggest that individuals without tertiary education attribute themselves a higher capability to evaluate aspects of health care than individuals with some university degree. This finding is alarming as lower educational levels have been associated with low health literacy [57], more difficulty in processing quality of care information, and less optimal health care choices [55,58,59]. These findings suggest that individuals who are most in need for tools that guide them toward a better search and assessment of Web-based information [60], attribute themselves higher expertise than they probably have from an objective point of view. As the Internet is a resource that may lead to or even encourage dangerous outcomes if guidance is lacking [60], more effort should be invested in fostering individuals' critical judgment of health information on the Internet in general, and on PRWs in particular.

In addition to education level and gender, age plays a significant role in individuals' judgment of what is important to identify a good physician and can be judged after a doctoral visit. Overall, with an increase in age, individuals perceive aspects of decision making with the physician and convenience (reachability, scheduling, etc) more important. Furthermore, older individuals attribute themselves a higher ability to assess physicians' technical skills. Most likely, older individuals have throughout their lifetime collected hands-on health care experiences that make them more comfortable about their skill to assess physicians. The interpretation of these results nevertheless calls for caution, as potential confounding factors have not been included in the analysis.

Limitations

The study has a number of limitations. First, there are limitations caused by the recruitment and composition of the two samples. The Delphi study faced a participant dropout rate from n=29 to n=19 from the first to the final round of the study. An additional limitation is posed by the health care consumer sample recruited in this study. The data were collected via snowball sampling on the Internet. Hence, the results cannot claim representativeness. Because the sample had a large share of younger, highly educated females who filled in the questionnaire, it would be recommendable to replicate this study in a different context or country with a more balanced sample.

XSL•FO
RenderX

Also, the results of the comparison between physicians and health care consumers are limited because the Delphi study sample was substantially smaller than the health care consumer sample due to the design (Delphi vs cross-sectional survey). Hence, it was not possible to conduct parametric tests to identify whether the assessments by health care consumers and physicians was statistically significant due to the diverse samples. In order to adjust for this limitation, strict grouping criteria and thresholds, as explained in the methods section, were applied to classify the indicators into the above-listed 4 categories. Furthermore, the cross-sectional survey consisted only of items that were retained after the first round of the Delphi study in order to shorten the survey for health care consumers. Hence, only items that physicians had identified as "important" in the first round of the Delphi study were presented to health care consumers, thereby potentially limiting the variance in the findings.

In addition, the way we assessed individuals' self-perceived capability to assess a physician's skill does not allow for definite answers about how well individuals would in reality be able to assess the quality of their health care. Also, it is debatable whether individuals would refrain from judging a physician's technical ability if they had the chance to do so, even if they indicated previously that they did not think they could assess this aspect of health care. Rather than asking individuals about their perceived ability to assess certain aspects of health care, showing them cases or video samples of treatment encounters and asking them to evaluate and review them could provide additional answers about individuals' perceived ability to assess their physicians' performance.

Four specializations with varying skill requirements, client groups, surgical involvement, and payment schemes (pediatricians, orthopedic surgeons, general practitioners, dentists) were invited to participate. In future research, other medical specialists should be invited to develop indicators to identify a good physician separately and only for that particular medical expertise. Indicators of how a good physician can be identified may vary depending on the specialization of physicians studied.

Conclusions

Physicians' and health care consumers' moderate agreement on important and assessable aspects of health care quality suggests that PRWs may profit from presenting information and word of mouth about the quality of a doctor in selected ways. Patients and physicians agreed that health care consumers' assessment of their provider should be constrained to matters of infrastructure, organization, staff, and his or her interpersonal skills. Technical expertise and outcomes of care were also identified to be important but both physicians and health care providers did not attribute to patients the capability to accurately assess them. Furthermore, our results show that sociodemographic characteristics (age, gender, educational level) play a role in health care consumers' assessment of what is important to identify a good doctor and what can be evaluated after the first visit.

Our findings suggest that health care consumers may consent to a mixed model in which search and experience aspects of care could be assessed by health care consumers, whereas technical care information could be provided by a source or committee that is competent to asses a physicians' medical skill (eg, an external expert committee). This could yield a hybrid model in which both health care consumers and experts may contribute information that is adjusted to their level of expertise. How a mixed format of health care consumers' evaluation of physicians and expert information could best be implemented, and to what extent PRW users would support and use such mixed PRW formats should be the subject of future research.

Acknowledgments

The authors would like to thank both the physicians as well as the cross-sectional study participants for their valuable time and partaking in the study. A special thanks to Ueli Zihlmann from the Central Switzerland Physician Association for his help in the recruitment of physicians participating in the Delphi study, as well as to Kilian Schaudig for his technical support in the setup and pretesting of the questionnaire.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Indicator comparison in terms of importance of the indicators as assessed by health care consumers and physicians.

[PDF File (Adobe PDF File), 249KB - jmir_v19i5e127_app1.pdf]

Multimedia Appendix 2

Indicator comparison between health care consumers and physicians in terms of health consumers' evaluation capability.

[PDF File (Adobe PDF File), 245KB - jmir_v19i5e127_app2.pdf]

References

XSL•FO
RenderX

1. Papacharissi Z. The virtual sphere: the internet as a public sphere. New Media Soc 2002 Feb 01;4(1):9-27. [doi: 10.1177/14614440222226244]
2. Cheung CMK, Lee MKO. What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. Decis Support Syst 2012 Apr;53(1):218-225. [doi: 10.1016/j.dss.2012.01.015]
3. Nelson P. Information and consumer behavior. J Polit Econ 1970;78(2):311-329.
4. Hsieh YC, Chiu HC, Chiang MY. Maintaining a committed online customer: a study across search-experience-credence products. J Retail 2005 Jan;81(1):75-82. [doi: 10.1016/j.jretai.2005.01.006]
5. Mitra K, Reiss MC, Capella LM. An examination of perceived risk, information search and behavioral intentions in search, experience and credence services. J Serv Mark 1999 Jun;13(3):208-228. [doi: 10.1108/08876049910273763]
6. Hennig-Thurau T, Walsh G. Electronic word-of-mouth: motives for and consequences of reading customer articulations on the Internet. Int J Electron Commer 2004;8(2):51-74.
7. Rosenthal M, Schlesinger M. Not afraid to blame: the neglected role of blame attribution in medical consumerism and some implications for health policy. Milbank Q 2002 Mar;80(1):41-95. [doi: 10.1111/1468-0009.00003]
8. Baron-Epel O, Dushenat M, Friedman N. Evaluation of the consumer model: relationship between patients? Expectations, perceptions and satisfaction with care. Int J Qual Heal Care 2001;13(4):317-323. [doi: 10.1093/intqhc/13.4.317]
9. Funnell MM, Anderson RM, Arnold MS, Barr PA, Donnelly M, Johnson PD, et al. Empowerment: an idea whose time has come in diabetes education. Diabetes Educ 1991;17(1):37-41. [doi: 10.1177/014572179101700108] [Medline: 1986902]
10. European Patients Forum. Charter on patient empowerment URL: http://www.eu-patient.eu/campaign/PatientsprescribE/charter-on-patient-empowerment/ [accessed 2016-08-08] [WebCite Cache ID 6jbzOdFxs]
11. World Health Organization (WHO) Regional Office for Europe. WHO. Empowering patients URL: http://www.euro.who.int/en/health-topics/noncommunicable-diseases/diabetes/news/news/2012/4/empowering-patients [accessed 2017-03-17] [WebCite Cache ID 6p1j70nc4]
12. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med 1999 Sep;49(5):651-661. [doi: 10.1016/S0277-9536(99)00145-8] [Medline: 10452420]
13. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/S0277-9536(96)00221-3]
14. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004;31(2):143-164. [doi: 10.1177/1090198104263660]
15. Ende J, Kazis L, Ash A, Moskowitz MA. Measuring patients' desire for autonomy: decision making and information-seeking preferences among medical patients. J Gen Intern Med 1989;4(1):23-30. [Medline: 2644407]
16. Roter D, Hall J. Doctors talking with patients/patients talking with doctors: improving communication in medical visits. Westport, CT: Praeger; 2006.
17. Moreau A, Carol L, Dedianne MC, Dupraz C, Perdrix C, Lainé X, et al. What perceptions do patients have of decision making (DM)? Toward an integrative patient-centered care model. A qualitative study using focus-group interviews. Patient Educ Couns 2012 May;87(2):206-211. [doi: 10.1016/j.pec.2011.08.010] [Medline: 21903355]
18. Suziedelyte A. How does searching for health information on the internet affect individuals' demand for health care services? Soc Sci Med 2012 Nov;75(10):1828-1835. [doi: 10.1016/j.socscimed.2012.07.022] [Medline: 22884947]
19. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 07;362(1):6-7. [doi: 10.1056/NEJMp0903473] [Medline: 20054044]
20. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011 Dec;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
21. Samora JB, Lifchez SD, Blazar PE. Physician-rating web sites: ethical implications. J Hand Surg Am 2016 Jan;41(1):1-8. [doi: 10.1016/j.jhsa.2015.05.034] [Medline: 26304734]
22. Strech D. Physician rating sites from an ethical viewpoint: a roadmap. Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen 2010 Jan;104(8-9):674-681. [doi: 10.1016/j.zefq.2010.09.019] [Medline: 21129706]
23. Segal J, Sacopulos MJ, Rivera DJ. Legal remedies for online defamation of physicians. J Leg Med 2009 Sep;30(3):349-388. [doi: 10.1080/01947640903146121] [Medline: 19681012]
24. Ornstein C. The Washington Post. 2016. Doctors fire back at bad Yelp reviews — and reveal patients' information online URL: https://www.washingtonpost.com/news/to-your-health/wp/2016/05/27/docs-fire-back-at-bad-yelp-reviews-and-reveal-patients-information-online/ [accessed 2016-08-15] [WebCite Cache ID 6jmSIW59Q]
25. Azu MC, Lilley EJ, Kolli AH. Social media, surgeons, and the Internet: an era or an error? Am Surg 2012 May;78(5):555-558. [Medline: 22546128]
26. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014 Feb 20;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
27. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]

XSL•FO
RenderX

28. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
29. Adams S. Post-panoptic surveillance through healthcare rating sites. Info Commun Soc 2012;16(2):215-235.
30. Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Finucane ML, et al. Taking patients' narratives about clinicians from anecdote to science. N Engl J Med 2015 Aug 13;373(7):675-679. [doi: 10.1056/NEJMsb1502361] [Medline: 26267629]
31. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
32. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? Results from an online-based cross-sectional study. J Med Internet Res 2016 Sep 19;18(9):e254. [doi: 10.2196/jmir.5889]
33. Wold Health Organization (WHO). 2006. Quality of care: a process for making strategic choices in health systems URL: http://apps.who.int/iris/handle/10665/43470 [accessed 2017-03-17] [WebCite Cache ID 6p1nkiIqY]
34. Donabedian A. The seven pillars of quality. Arch Pathol Lab Med 1990 Nov;114(11):1115-1118. [Medline: 2241519]
35. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260(12):1743-1748. [Medline: 3045356]
36. Donabedian A. Evaluating the quality of medical care. Milbank Q 2005;83(4):691-729 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00397.x] [Medline: 16279964]
37. Maribito A. Looking for a good doctor (or realtor or mechanic): construing quality with credence services. 2008. URL: https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/ETD-TAMU-3019/MIRABITO-DISSERTATION.pdf?sequence=1&isAllowed=y
38. von der Gracht HA. Consensus measurement in Delphi studies. Technol Forecast Soc Change 2012 Oct;79(8):1525-1536. [doi: 10.1016/j.techfore.2012.04.013]
39. Campbell SM, Roland MO, Buetow SA. Defining quality of care. Soc Sci Med 2000 Dec;51(11):1611-1625. [Medline: 11072882]
40. FMH. Tarmed URL: http://www.fmh.ch/ambulante_tarife/tarmed-tarif/tarmed_grundlagen.html [accessed 2016-10-04] [WebCite Cache ID 6p1oEXCOG]
41. Equam. EQUAM Stiftung URL: https://www.equam.ch/de/equam-fuer-praxen/ [accessed 2017-03-30] [WebCite Cache ID 6pLPB6eqW]
42. QBM Stiftung. Stiftung für Qualitätsentwicklung in der ambulanten Medizin URL: http://www.qbm-stiftung.ch/home.html [WebCite Cache ID 6p0hSlHP3]
43. Hayes A. Assessing and quantifying reliability. In: Statistical Methods for Communication Science. Mahwah, NJ; London: Lawrence Erlbaum Associates Publishers; 2005:118-129.
44. Field A. Discovering Statistics Using SPSS, 3rd edition. London: SAGE Publications; 2009.
45. Gaetano J. Researchgate. Holm-Bonferroni sequential correction: an EXCEL calculator (1.1) [Microsoft Excel workbook] URL: https://www.researchgate.net/publication/236969037_Holm-Bonferroni_Sequential_Correction_An_EXCEL_Calculator [accessed 2017-03-17] [WebCite Cache ID 6p1olakTn]
46. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6(2):65-70.
47. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
48. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? An analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
49. Ejaz A, Spolverato G, Bridges JF, Amini N, Kim Y, Pawlik TM. Choosing a cancer surgeon: analyzing factors in patient decision making using a best-worst scaling methodology. Ann Surg Oncol 2014 Jun 4;21:3732-3738. [doi: 10.1245/s10434-014-3819-y]
50. Charaghi-Sohi S, Hole A, Mead N, McDonald R, Whalley D, Bower P, et al. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. Ann Fam Med 2008;6(2):107-115. [doi: 10.1370/afm.816]
51. Fung CH, Elliott MN, Hays RD, Kahn KL, Kanouse DE, McGlynn EA, et al. Patients' preferences for technical versus interpersonal quality when selecting a primary care physician. Health Serv Res 2005 Aug;40(4):957-977 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00395.x] [Medline: 16033487]
52. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Med Care Res Rev 2007 Apr;64(2):169-190. [doi: 10.1177/1077558707070640020301] [Medline: 17406019]
53. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
54. Hibbard J, Slovic P, Jewett J. Informing consumer decisions in health care: implications from decision-making research. Milbank Q 1997;75(3):395-414. [Medline: 9290635]

XSL•FO
RenderX

55. Schlesinger M, Kanouse DE, Martino SC, Shaller D, Rybowski L. Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. Med Care Res Rev 2014 Oct;71(5):38-64. [doi: 10.1177/1077558713496321] [Medline: 23999489]
56. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
57. 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]
58. Victoor A, Delnoij DM, Friele RD, Rademakers KJ. Determinants of patient choice of healthcare providers: a scoping review. BMC Health Serv Res 2012;12(1):272-288 [FREE Full text] [doi: 10.1186/1472-6963-12-272] [Medline: 22913549]
59. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? a systematic review. Med Care 2009;47(1):1-8. [doi: 10.1097/MLR.0b013e3181808bb5]
60. Schulz PJ, Nakamoto K. Patient behavior and the benefits of artificial intelligence: the perils of "dangerous" literacy and illusory patient empowerment. Patient Educ Couns 2013 Aug;92(2):223-228. [doi: 10.1016/j.pec.2013.05.002]

Abbreviations

PRW: physician rating website
eWOM: electronic word-of-mouth
QBM: Qualitäts-Basis Modul (Verdag)

Edited by G Eysenbach; submitted 24.10.16; peer-reviewed by S McLennan, S Martino; comments to author 22.12.16; revised version received 17.01.17; accepted 10.02.17; published 01.05.17.

Please cite as:
Rothenfluh F, Schulz PJ
Physician Rating Websites: What Aspects Are Important to Identify a Good Doctor, and Are Patients Capable of Assessing Them? A Mixed-Methods Approach Including Physicians' and Health Care Consumers' Perspectives
J Med Internet Res 2017;19(5):e127
URL: https://www.jmir.org/2017/5/e127/
doi:10.2196/jmir.6875
PMID:28461285

©Fabia Rothenfluh, Peter J Schulz. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mobile App for Treatment of Stress Urinary Incontinence: A Cost-Effectiveness Analysis

Malin Sjöström[1], MD, PhD; Lars Lindholm[1], PhD; Eva Samuelsson[1], MD, PhD

Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

Corresponding Author:
Malin Sjöström, MD, PhD
Department of Public Health and Clinical Medicine
Umeå University
Region Jämtland Härjedalen
Box 654
Umeå, SE-832 27
Sweden
Phone: 46 703269948
Fax: 46 63142221
Email: malin.sjostrom@regionjh.se

Abstract

Background: Mobile apps can increase access to care, facilitate self-management, and improve adherence to treatment. Stress urinary incontinence (SUI) affects 10-35% of women and, currently, an app with instructions for pelvic floor muscle training (PFMT) is available as first-line treatment. A previous randomized controlled study demonstrated that the app benefitted symptom severity and quality of life (QoL); in this study we investigate the cost-effectiveness of the app.

Objective: The objective of this study was to evaluate the health economy of the app for treating SUI.

Methods: This deterministic cost-utility analysis, with a 1-year societal perspective, compared the app treatment with no treatment. Health economic data were collected alongside a randomized controlled trial performed in Sweden from March 2013 to October 2014. This study included 123 community-dwelling women participants of 18 years and above, with stress urinary incontinence ≥1 time per week. Participants were self-assessed with validated questionnaires and 2-day leakage diaries, and then randomized to 3 months of treatment (app group, n=62) or no treatment (controls, n=61). The app focused on pelvic floor muscle training, prescribed 3 times daily. We continuously registered treatment delivery costs. Data were collected on each participant's training time, incontinence aids, and laundry at baseline and at a 3-month follow-up. We measured quality of life with the International Consultation on Incontinence Modular Questionnaire on Lower Urinary Tract Symptoms and Quality of Life, and calculated the quality-adjusted life years (QALYs) gained. Data from the 3-month follow-up were extrapolated to 1 year for the calculations. Our main outcome was the incremental cost-effectiveness ratios compared between app and control groups. One-way and multiway sensitivity analyses were performed.

Results: The mean age of participants was 44.7 years (SD 9.4). Annual costs were €547.0 for the app group and €482.4 for the control group. Annual gains in quality-adjusted life years for app and control groups were 0.0101 and 0.0016, respectively. Compared with controls, the extra cost per quality-adjusted life year for the app group ranged from −€2425.7 to €14,870.6, which indicated greater gains in quality-adjusted life years at similar or slightly higher cost.

Conclusions: The app for treating stress urinary incontinence is a new, cost-effective, first-line treatment with potential for increasing access to care in a sustainable way for this patient group.

(J Med Internet Res 2017;19(5):e154) doi:10.2196/jmir.7383

KEYWORDS

mobile application; pelvic floor; urinary incontinence, stress; self care; cost-benefit analysis

XSL•FO
RenderX

Introduction

One possible way to meet the future demands in the health care sector could be to empower patients by increasing self-management with mHealth [1]. Worldwide, there are approximately 5 billion mobile phone subscribers, and the smartphones are constant companions for many individuals [2]. The App Store and Google Play websites offer around 100,000 health apps, but few have been scientifically evaluated [3]. Mobile health apps could facilitate self-management and adherence to treatment; in addition, they could increase access to care for individuals with limited access or for those unwilling to seek ordinary health care [4].

Stress urinary incontinence (SUI), that is, urine leakage upon sneezing, coughing, or exertion [5], affects 10-35% of women [6,7]. This condition is suited to self-management. The diagnosis is based on patient-reported measures and does not require a physical examination [8]. The first-line treatment is pelvic floor muscle training (PFMT), which is safe, effective [8-10], and can be completed without health care personnel supervision [8]. Although SUI can decrease quality of life (QoL) [11], only around 20% of individuals seek care [12]. In some cases, the leakage is not considered as a major problem, but in other cases, the patient is too embarrassed to seek care [13]. Our research group has developed the mobile app, Tät, which serves as a first-line treatment for SUI, based on self-management. This app provides information and instructions for PFMT [14].

One concern in deciding the treatments to be delivered in the health care systems is the cost. One common way to evaluate cost is the cost-utility analysis, which compares the costs and effects of at least two treatment alternatives. This analysis allows comparison of diverse interventions [15]. Costs can be considered either from a health care perspective, which only includes costs borne by the health care system, or from a societal perspective, which includes other costs. Currently, the former perspective is recommended in the United Kingdom by the National Institute for Health and Clinical Exellence (NICE) [16], and the latter is recommended in the United States [17] and Sweden [18]. The utility of the treatment is defined as the added time gained with an improved QoL, calculated as quality-adjusted life years (QALYs).

In this study, we performed a cost-utility analysis of SUI treatment with the app, Tät, compared with no treatment.

Methods

Design

This deterministic cost-utility analysis had a 1-year societal perspective. It was performed according to the principles outlined by Drummond et al [15].

Population

We collected data for this analysis alongside a randomized controlled trial on SUI treatment with the Tät app. The trial was registered at clinicaltrials.gov (ID: NCT01848938), and the trial results were described in detail elsewhere [14].

Briefly, we recruited community-dwelling women, aged 18 years and above, with SUI of once or more than once a week, via our website. Interested women completed an online screening questionnaire. When they met the study criteria, we sent them a letter of information, a form to provide informed consent, and a 2-day leakage diary. After returning these, they completed a Web-based questionnaire that provided their background characteristics, medical history, symptom severity, and QoL. Exclusion criteria were ongoing pregnancy, maximum voiding volume <0.3 L, macroscopic hematuria, irregular menstrual bleeding, difficulty passing urine, previous incontinence surgery, previous or present malignancy in the lower abdomen, severe psychiatric disorder, or impaired mobility or sensibility in the legs or lower abdomen.

We consecutively randomized eligible women to either three months of treatment with the app (app group, n=62), or no treatment (control group, n=61). The app Tät contained information on SUI, provided a PFMT program, with 6 basic and 6 advanced levels, and it prescribed PFMT 3 times daily during treatment. At the end of treatment, the instructions were to continue PFMT 2 or 3 times per week for maintenance training [19]. The control group received no intervention. After the 3-month follow-up, we offered the participants in the control group the app, on an optional basis. We collected 3-month follow-up data with a Web-based questionnaire. There was no face-to-face contact with the participants at any time.

Symptom Severity

We measured symptom severity at baseline and at 3 months, with the validated [20] and recommended [8,10,21] questionnaire, the International Consultation on Incontinence Modular Questionnaire on Urinary Incontinence, Short Form (ICIQ-UI SF). It contained 3 items such as frequency, amount of leakage, and overall QoL impact. The total score ranged from 0 to 21, with higher scores indicating greater severity. The total scores were used to categorize the severity of the condition (1-5: slight; 6-12: moderate; 13-18: severe; and 19-21: very severe) [22]. After 3 months, the app group reported clinically relevant and significantly greater improvements in symptoms compared with the control group [14].

Costs

We evaluated costs from a 1-year societal perspective. Costs included the cost of mailing the 2-day leakage diary, and the estimated time spent by our study administrator in emailing each participant the link to the Web-based questionnaire. The cost for our study administrator's time was calculated based on her gross hourly wage. We did not include costs for the app development because these are one-time costs and are comparable with, for example, the costs for basic education of health care personnel; these are costs which are normally not included in health economic analyses. No other costs for the delivery of treatment were identified.

We collected baseline and follow-up data on the use of incontinence aids and any extra laundry due to leakage. In a previous study, we collected data on the different types of incontinence aids (large, medium, or small) used by women with SUI [23]. We then calculated a mean price per unit, based

XSL•FO
RenderX

on the prices for incontinence aids listed on the website of a large pharmacy brand (Apoteket). The price for laundry was derived from the literature [24].

When a societal perspective is applied, an estimate of the cost for the individual's time should be included in the health economic analysis [14]. At the 3-month follow-up, participants estimated how much time they had spent on PFMT during the last 4 weeks. We used this estimate to calculate the PFMT performed during the treatment period. To estimate PFMT for the entire year for the app group, we assumed that the participants would follow the prescription for maintaining PFMT over the remaining part of the year and we adjusted the time spent on PFMT accordingly. For the control group, we assumed the time spent on PFMT would remain constant throughout the year. To estimate the cost for each participant's time, we calculated the gross hourly wages for women with the same educational level in Sweden [25], a method which is commonly used [14].

For all other costs, we assumed that costs measured at the 3-month follow-up would remain constant throughout the year. We added up all the costs, and the sum represented the total societal cost. All costs are given in euro, and they were based on the 2013 year-end prices. At that time, the exchange rate for 1 EUR was 8.94 SEK (Swedish krona).

Quality of life, Utility Weights, and QALYs

To evaluate QoL, we used the validated [26-28] and recommended [8,10,24] condition-specific questionnaire, the International Consultation on Incontinence Modular Questionnaire on Lower Urinary Tract Symptoms and Quality of Life (ICIQ-LUTSqol). This questionnaire contained 19 items on aspects of everyday life that might be influenced by urinary leakage, such as travel, work, meetings with family and friends, exercise, sexual performance, mood, energy, and sleep. Items are scored 1-4 (1: not at all or never; 2: slightly or sometimes; 3: moderately or often; and 4: a lot or all the time). The overall score ranged from 19 to 76, with higher scores indicating more impact. The questionnaire, was derived from the Kings Health Questionnaire [26], which is widely used in health economic analyses; both questionnaires used the same method for calculating QALY [29].

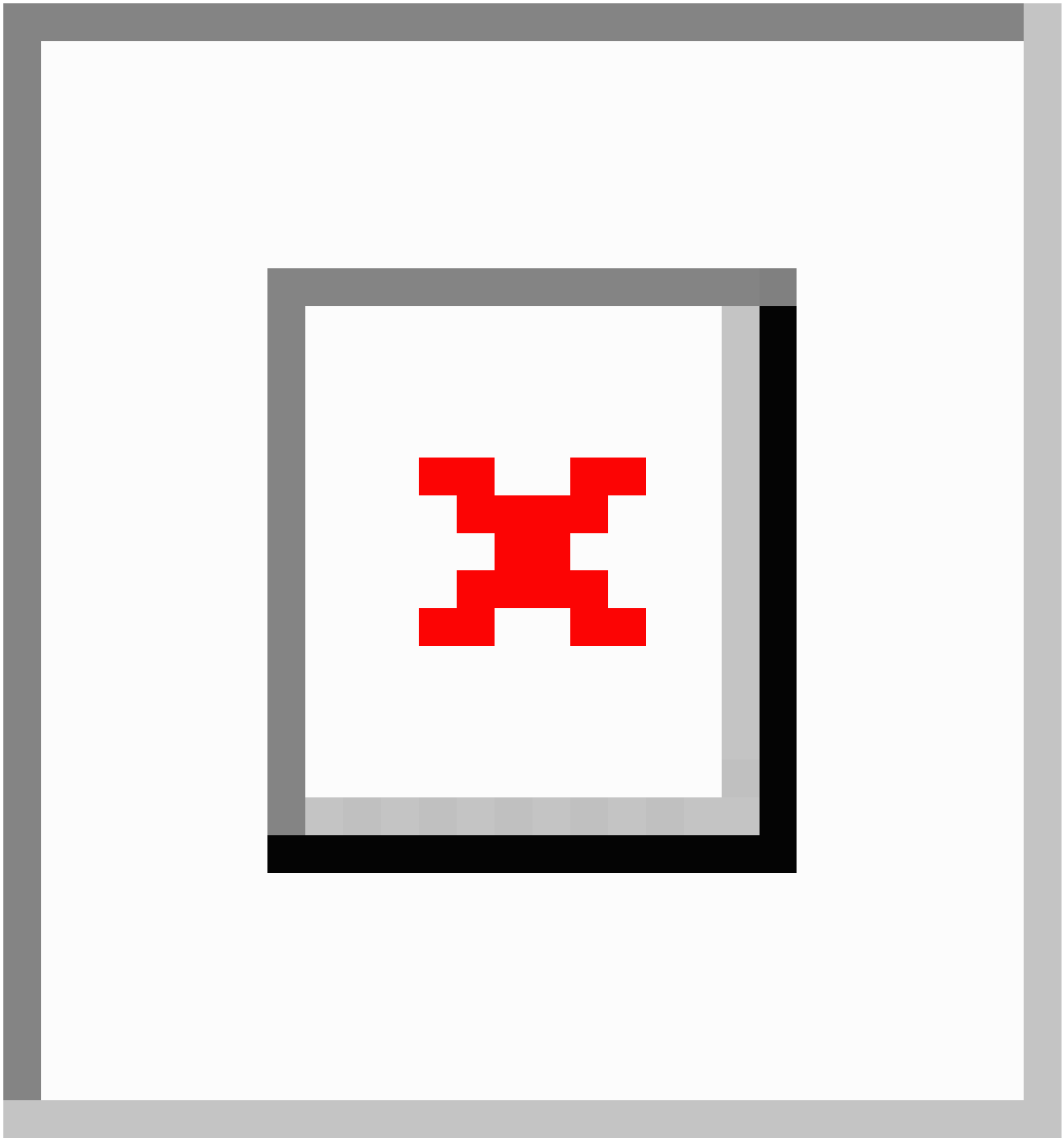
We based our QALY calculations on data from the ICIQ-LUTSqol, and we applied a preference-based index derived by Brazier et al [29], which incorporates 9 of the 19 items into a "health state" classification. We used the algorithm of this index to translate the health state classification into a utility weight, which ranged from 0 (worst imaginable health state) to 1 (best imaginable health state). We assumed that the utility weight calculated from 3-month follow-up data would remain stable for the remainder of the year in both groups.

Main Outcome

Our main outcome was the incremental cost-effectiveness ratio (ICER), defined as the difference in cost-effectiveness between the app group and the control group. We calculated the ICER as presented in Figure 1.

XSL•FO
RenderX

Figure 1. Equation for calculation of the incremental cost-effectiveness ratio (ICER) using costs and quality-adjusted life-years (QALYs) for the app and control group respectively.



Statistics

For group comparisons at baseline, we used the Student *t* test for continuous variables, the chi-square test for categorical variables, and the Mann-Whitney *U* test for ordinal variables. For analyses of treatment effects within each group, we used paired *t* test. For comparisons of treatment effects between groups, we used a linear mixed-model analysis, which incorporated all available data for the outcomes of symptom severity and QoL. The utility weights are expressed as the mean value with a 95% CI. Costs were assumed to change linearly, and QALYs were calculated based on an "area-under-the-curve." *P* values <.05 were considered statistically significant. We collected and analyzed data in SPSS for Mac, version 23.0 (IBM) and in Excel for Mac, version 14.6.3 (Microsoft Corporation).

Sensitivity Analysis

We considered the fact that PFMT could be performed while doing other things and that laundry caused by leakage could be washed with other garments. Therefore, we performed a one-way sensitivity analysis by varying input data on the time spent on PFMT, the time spent on laundry, and the cost of laundry, one at a time, to test the potential impact of these uncertainties. In addition, we performed a multiway analysis that incorporated all three variables.

XSL•FO
RenderX

Ethics

The Regional Ethical Review Board, Umeå University, approved of the study (number 2012-325-31M). All participants gave informed consent.

Results

Study Population

We performed this study in Sweden, from March 2013 to October 2014. We randomized 123 participants to receive the app (app group, n=62) or with no treatment (control group, n=61). Baseline characteristics, including age, education, symptom severity, and baseline QoL scores did not differ significantly between groups (Table 1).

At follow-up, we had lost one participant from each group. In addition, in the app group, we were missing outcome data on the ICIQ-LUTSqol for 3 participants.

Costs

The total assessment cost per participant was €6.4. The app group had higher total costs than the control group, mainly due to the extra time spent on PFMT. The total annual cost per participant in each group is presented in Table 2.

Table 1. Baseline characteristics of study participants.

Variable	App group (n=62)	Control group (n=61)
Age, mean years (SD[a])	44.8 (9.7)	44.7 (9.1)
University education ≥3 years, n (%)	52 (84)	46 (75)
BMI[b], mean kg/m^2 (SD)	24.0 (4.1)	24.5 (4.4)
Daily smoker, n (%)	2 (3)	3 (5)
Symptom severity, n (%)[c]		
Slight	3 (5)	0 (0)
Moderate	36 (58)	42 (69)
Severe	23 (37)	19 (31)
Overall score ICIQ-UI SF[c], mean (SD)	11.1 (3.0)	11.0 (2.6)
Overall score ICIQ-LUTSqol[d], mean(SD)	34.1 (6.1)	34.8 (6.1)

[a]SD: standard deviation.

[b]BMI: body mass index.

[c]Based on overall score from the International Consultation on Incontinence Modular Questionnaire on Urinary Incontinence, Short Form (ICIQ-UI SF).

[d]ICIQ-LUTSqol: International Consultation on Incontinence Modular Questionnaire on Lower Urinary Tract Symptoms, Quality of Life.

Table 2. Costs per participant included in a cost-effectiveness analysis with a 1-year societal perspective, for App group versus Control group in women with stress urinary incontinence.

Variable	Price per unit[a]	Amount used		Cost	
		App group	Control group	App group	Control group
Assessment	6.4	1	1	6.4	6.4
Treatment delivery	0	-	-	0	0
Participant costs					
Participant's time for PFMT[b,c], mean (h)	29.61	15.66	9.91	463.7	293.4
Participant's time for laundry[c], mean (h)	29.61	1.30	3.38	38.5	100.1
Incontinence aids[d], mean (n)	0.134	114.40	169.60	15.4	22.7
Extra laundry loads[e], mean (n)	2.21	10.40	27.04	23.0	59.8
Total cost				547.0	482.4

[a]Prices are in euro (€), based on the 2013 year-end prices. Exchange rate was 1 EUR=8.94 SEK.

[b]PFMT: pelvic floor muscle training.

[c]Based on 2013 mean income for Swedish women with a similar educational level.

[d]Based on mean consumption at baseline (prices acquired from Apoteket).

[e]Data from the literature [24].

XSL•FO
RenderX

Quality of Life, Utility Weights, and QALYs

In the app group, there was significant improvement in QoL at follow-up (mean ICIQ-LUTSqol reduction: 4.8, 95% CI 3.4-6.2). In contrast, the control group did not display a significant reduction in scores (mean ICIQ-LUTSqol reduction: 0.7, 95% CI −0.5 to 1.8). The difference between groups was highly significant (P<.001).

The utility weights and QALY changes for each group are presented in Figure 2. The QALYs gained corresponded to an extra 3.9 days in the best imaginable health state for the app group, and only 0.6 days for the control group.

Main Outcome and Sensitivity Analysis

In Table 3, we illustrate the ICERs for the base case and the sensitivity analysis. In all the analyses, except one (participant's time for PFMT halved), the costs in the app group were slightly higher than costs in the control group. However, in all cases, the app treatment was more effective compared with no treatment or control group.

Table 3. Incremental cost-effectiveness ratios (ICERs) for the app group versus the control group, including the base case and a sensitivity analysis.

Group		Total cost (€[a])	QALY[b]-gain	Δ Cost (€)	Δ QALY-gain	ICER[c]
Base case						
	Control group	482.4	0.00158			
	App group	547.0	0.01006			
	App group vs control group			64.6	0.00849	7615.5
Sensitivity analysis						
One-way: participant's time for PFMT[d] halved						
	Control group	335.7	0.00158			
	App group	315.1	0.01006			
	App group vs control group			−20.6	0.00849	−2425.7
One-way: cost for laundry halved						
	Control group	452.4	0.00158			
	App group	535.4	0.01006			
	App group vs control group			83.1	0.00849	9785.5
One-way: participant's time for laundry not included						
	Control group	382.3	0.00158			
	App group	508.5	0.01006			
	App group vs control group			126.2	0.00849	14870.6
Multiway: participant's time for PFMT and cost for laundry halved, participant's time for laundry not included						
	Control group	205.7	0.00158			
	App group	265.1	0.01006			
	App group vs control group			59.4	0.00849	6999.5

[a]€ refers to euro at 2013 year-end price.

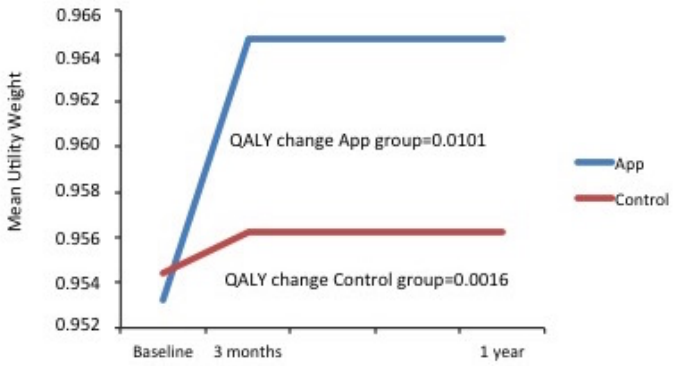
[b]QALY: quality-adjusted life years.

[c]ICER: Δ Cost/Δ QALY-gain.

[d]PFMT: pelvic floor muscle training.

XSL•FO
RenderX

Figure 2. Changes in utility weights reflect gains in quality-adjusted life years (QALYs). Assessments were recorded at baseline, and at 3-month follow-ups, and estimated at one-year, for individuals that received either the app treatment (top, blue line) or no treatment (bottom, red line) for stress urinary incontinence.



Discussion

Principal Findings

In this health economic evaluation, we demonstrated that SUI self-management with a mobile app that provided information and instructions for PFMT was a cost-effective first-line treatment alternative, compared with a control group that received no treatment. The results were consistent and stable in different scenarios with varying costs.

Strengths and Limitations With the Study

This study had several strengths. The calculations were based on known costs and on data collected directly from the participants. Our research group had previous experience with non–face-to-face SUI treatment; there were no disruptions or major technical problems during the study and the loss to follow-up was low. We applied existing guidelines and used validated and recommended outcomes. The diagnosis of SUI was well substantiated and the population was clinically relevant because the vast majority (120/123, 97.6%) of participants had moderate to severe symptoms and actively sought treatment.

This study also had some limitations. One was the relatively low number of participants (n=123), which increases the uncertainty of the data and might have affected the results. Another was that we did not have 1-year follow-up data. Instead, we assumed that costs and utility weights measured at the 3-month follow-up would remain constant over the year. This assumption was based on our previous study of long-term effects of Internet-based PFMT for SUI [30], where improvements achieved after 3 months of treatment were maintained after 1 and 2 years. We had no reason to believe that the outcome of the current app treatment would be different from that of the previous study. Moreover, although the no-treatment alternative was plausible, given the fact that only around 20% of affected women seek care [12], it would have been interesting to compare outcomes between the app and a care-as-usual alternative. However, care-as-usual varies substantially, because there is no gold standard for SUI treatment. Another limitation was that our population had a higher educational level (98/123, 79.7% had ≥3 years of university education) than the general population of Swedish women (≈30% of women aged 25-44 years have ≥3 years of university education) [31]. However, there are no indications that the educational level might affect the ability to perform PFMT [32].

Strengths and Weaknesses Compared With the Literature

The total cost per participant was higher in the app group (€547.0) than that in the control group (€482.4). Although savings on laundry and incontinence aids were larger in the app group, participants in this group spent more time on PFMT. However, our estimation of the cost for participant time might have been somewhat biased, due to the relatively high educational level of our participants, compared with that of the general population. Nevertheless, the results were consistent in all tested scenarios, and the costs were comparable with those reported in other studies on conservative SUI treatments. For example, in a previous study, we compared Internet-based programs and postal-treatment programs for PFMT, where we found that the total costs were €596.5 and €596.2, respectively [20]. Moreover, in a Dutch study on a care-as-usual SUI treatment, the total cost was €453 (including productivity losses, travel costs, patient out-of-pocket costs, and health care costs, but not time for PFMT) [33].

In this study, we most likely overestimated the QALY gains in the control group due to the fact that we considered it significant in the incremental analysis despite of controls not showing a significant improvement in QoL. The app-group gain in QALY (0.0101) might seem small, but it was comparable to QALY gains observed in other studies on SUI treatment. In a primary care setting, Albers-Heitner et al [33] reported incremental QALY gains of 0.01-0.02, when intense PFMT was performed under guidance of a specialist nurse and compared with a general practitioner (GP) care-as-usual alternative. Arlandis-Guzman et al [34] reported QALY gains of 0.01014, 0.00846, and 0.00957, with the antimuscarinic drugs fesoterodine, tolterodine, and solifenacin, respectively. However, second-line treatments with sling surgery could produce larger QALY gains (0.0504) [35]. Nevertheless, although the QALY gains with conservative treatments are low, in sheer numbers, the patient group that can potentially benefit from treatment is large; thus, the attainable total QALY gain is substantial.

XSL•FO
RenderX

We estimated that the extra cost per QALY for the app treatment was €7615.5, and the sensitivity analyses indicated a potential range of −€2425.7 to €14,870.6. In one of the scenarios tested (time for PFMT halved), the negative ICER value implied that, compared with doing nothing for this group of patients, the app treatment could increase the QoL for the individual at a reduced cost for the society. In the other scenarios, QALY gains were larger in the app group, but at greater cost, compared with the control group. The affordability of an additional cost per QALY depends on the willingness to pay for a more effective treatment, which might differ in different countries. In the United Kingdom, interventions with ICERs ≤€16,500-25,000 (£20,000–30,000) are typically recommended by NICE [16], and in US, those with ICERs of ≤€36,500 ($50,000) are usually recommended [36,37]. In Sweden, incremental costs of ≤€11,200 (100,000 SEK) are considered low, and incremental costs of ≥€60,000 (500,000 SEK) are considered high [38]. Data are scarce on the cost effectiveness of other health apps, due to the limited number of studies conducted.

We did not calculate an ICER from a health care perspective, because the assessment cost was the same (€6.3) in both the groups, and no costs could be identified for the delivery of treatment, that is, Δ Cost=0. When the app is implemented outside a study setting, the parameters will change. For example, additional costs for updates, bug fixes, and technical support must be taken into account. We estimate that an IT technician would require approximately 4 h per month for this maintenance, but on the other hand, as the number of users increases, the cost per person will diminish. Furthermore, the total cost for the health care system to deliver an app treatment is likely to be low compared with face-to-face treatment. For example, in Sweden, the estimated cost for a GP consultation was €173 [39]; in United Kingdom, the estimated cost for 3 months of PFMT under supervision of a trained nurse was €158 to €293 [40].

Future Research and Clinical Implications

The Tät app has been released free of charge in both Swedish and English. Although the cost per participant will decrease as the number of users increases, the effects might decline outside the study setting. We are currently continuing to follow the effects reported by users. In addition, the long-term effects of the app need to be established, and the app should also be evaluated as a possible complement to other treatments. Future perspectives include developing the app for treating other types of urinary incontinence.

SUI treatment with the Tät app will not suit all women, but it offers a cost-effective first-line treatment to many women. To our knowledge, this study was the first to evaluate the cost effectiveness of an app treatment for a common health condition. Modern health care systems face many challenges, and it is important for clinicians to deliver care in sustainable ways. The development of self-management apps could be a feasible way to deliver high-quality care in a cost effective, affordable manner to large patient groups. It could also be a way to provide treatment to women with limited access to care, for example, women in low or middle-income countries. While to adding value to the individual patient, these apps could reduce the need for support from primary care, and thus, those resources could be conserved for individuals with explicit needs.

Conclusion

Self-management of SUI with an app for PFMT is a cost-effective first-line treatment alternative.

Acknowledgments

We would like to thank the participating women for making this study possible. Many thanks also to the Department of ICT Services and System Development (ITS) at Umeå University for helping in developing the app. We would also like to express our gratitude to our colleagues, Ina Asklund and Emma Nyström, for assistance in the study and to our study administrator, Susanne Johansson. This study was funded by grants from the Swedish Council for Working Life and Social Research, the Region Jämtland Härjedalen, and Visare Norr, Northern County Councils, Sweden. The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

The application Tät was developed by Eva Samuelsson, Malin Sjöström, and Göran Umefjord in cooperation with the Department of ICT Services and System Development (ITS), Umeå University, Sweden. The application is a registered trademark by the Swedish Patent and Registration office for E Samuelsson at Umeå University. It is also CE-marked as a medical device Class 1, according to Swedish regulation LVFS 2003:11. It is available for free in English and Swedish at Appstore and Google Play.

Conflicts of Interest

None declared.

References

1. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
2. WHO. From innovation to implementation URL: http://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf?ua=1 [WebCite Cache ID 6nYmyUH5w]

XSL•FO
RenderX

3. 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]
4. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
5. Haylen BT, de RD, Freeman RM, Swift SE, Berghmans B, Lee J, et al. An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. Neurourol Urodyn 2010;29(1):4-20. [doi: 10.1002/nau.20798] [Medline: 19941278]
6. Hannestad YS, Rortveit G, Sandvik H, Hunskaar S, Norwegian EPINCONT study. Epidemiology of Incontinence in the County of Nord-Trøndelag. A community-based epidemiological survey of female urinary incontinence: the Norwegian EPINCONT study. Epidemiology of Incontinence in the County of Nord-Trøndelag. J Clin Epidemiol 2000 Nov;53(11):1150-1157. [Medline: 11106889]
7. Abrams P, Cardozo L, Khoury S, Wein A. Epidemiology of Urinary (UI) and Faecal (FI) Incontinence and Pelvic Organ Prolapse (POP). In: Incontinence 4th edition. Paris, France: Health Publications Ltd; 2009.
8. Shamliyan T, Wyman J, Kane R. AHRQ. Nonsurgical treatments for urinary incontinence in adult women: diagnosis and comparative effectiveness URL: http://www.effectivehealthcare.ahrq.gov/ehc/products/169/834/urinary-incontinence-treatment-report-130909.pdf [WebCite Cache ID 6nRS4bzsg]
9. Dumoulin C, Hunter KF, Moore K, Bradley CS, Burgio KL, Hagen S, et al. Conservative management for female urinary incontinence and pelvic organ prolapse review 2013: summary of the 5th International consultation on incontinence. Neurourol Urodyn 2016 Jan;35(1):15-20. [doi: 10.1002/nau.22677] [Medline: 25400065]
10. National Institute for HealthClinical Excellence. Urinary incontinence: the management of urinary incontinence in women URL: https://www.nice.org.uk/guidance/cg171/resources/urinary-incontinence-in-women-management-35109747194821 [accessed 2017-01-16] [WebCite Cache ID 6nYnEVk6h]
11. Coyne KS, Kvasz M, Ireland AM, Milsom I, Kopp ZS, Chapple CR. Urinary incontinence and its relationship to mental health and health-related quality of life in men and women in Sweden, the United Kingdom, and the United States. Eur Urol 2012 Jan;61(1):88-95. [doi: 10.1016/j.eururo.2011.07.049] [Medline: 21831517]
12. Hannestad YS, Rortveit G, Hunskaar S. Help-seeking and associated factors in female urinary incontinence. The Norwegian EPINCONT Study. Epidemiology of Incontinence in the County of Nord-Trøndelag. Scand J Prim Health Care 2002 Jun;20(2):102-107. [Medline: 12184708]
13. Kinchen KS, Burgio K, Diokno AC, Fultz NH, Bump R, Obenchain R. Factors associated with women's decisions to seek treatment for urinary incontinence. J Womens Health (Larchmt) 2003 Sep;12(7):687-698. [doi: 10.1089/154099903322404339] [Medline: 14583109]
14. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2016 Sep 09:a Epub ahead of print. [doi: 10.1002/nau.23116] [Medline: 27611958]
15. Drummond MS, Torrance G, O'Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes, 3rd edition. Oxford: Oxford University Press; 2005.
16. National Institute for Health and Clinical Exellence. 2013. Guide to the methods of technology appraisal URL: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 [accessed 2017-01-16] [WebCite Cache ID 6nYnPsghT]
17. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996 Oct 09;276(14):1172-1177. [Medline: 8827972]
18. The Dental and Pharmaceutical Benefits Agency. TLV. Health economics URL: http://www.tlv.se/In-English/medicines-new/health-economics/ [accessed 2017-01-11] [WebCite Cache ID 6nRSEic08]
19. Kraemer WJ, Adams K, Cafarelli E, Dudley GA, Dooly C, Feigenbaum MS, American College of Sports Medicine. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. Med Sci Sports Exerc 2002 Feb;34(2):364-380. [Medline: 11828249]
20. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. Neurourol Urodyn 2004;23(4):322-330. [doi: 10.1002/nau.20041] [Medline: 15227649]
21. Abrams P, Andersson KE, Birder L, Brubaker L, Cardozo L, Chapple C, Fourth International Consultation on Incontinence. Fourth International Consultation on Incontinence Recommendations of the International Scientific Committee: Evaluation and treatment of urinary incontinence, pelvic organ prolapse, and fecal incontinence. Neurourol Urodyn 2010;29(1):213-240. [doi: 10.1002/nau.20870] [Medline: 20025020]
22. Klovning A, Avery K, Sandvik H, Hunskaar S. Comparison of two questionnaires for assessing the severity of urinary incontinence: the ICIQ-UI SF versus the incontinence severity index. Neurourol Urodyn 2009;28(5):411-415. [doi: 10.1002/nau.20674] [Medline: 19214996]

XSL•FO
RenderX

23. Sjöström M, Umefjord G, Lindholm L, Samuelsson E. Cost-effectiveness of an Internet-based treatment program for stress urinary incontinence. Neurourol Urodyn 2015 Mar;34(3):244-250 [FREE Full text] [doi: 10.1002/nau.22540] [Medline: 24347521]
24. Subak L, Van Den Eeden S, Thom D, Creasman JM, Brown JS, Reproductive Risks for Incontinence Study at Kaiser Research Group. Urinary incontinence in women: direct costs of routine care. Am J Obstet Gynecol 2007 Dec;197(6):596.e1-596.e9. [doi: 10.1016/j.ajog.2007.04.029] [Medline: 17880904]
25. Statistics Sweden Statistical Database. SCB. Labour market 2013 URL: http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__AM__AM0112/TidsserieUtbniva/table/tableViewLayout1/?rxid=0408f212-2586-4d09-bc21-82819c677eb3 [WebCite Cache ID 6nYohsTuv]
26. Kelleher CJ, Cardozo LD, Khullar V, Salvatore S. A new questionnaire to assess the quality of life of urinary incontinent women. Br J Obstet Gynaecol 1997 Dec;104(12):1374-1379. [Medline: 9422015]
27. Coyne K, Kelleher C. Patient reported outcomes: the ICIQ and the state of the art. Neurourol Urodyn 2010 Apr;29(4):645-651. [doi: 10.1002/nau.20911] [Medline: 20432329]
28. Sjöström M, Stenlund H, Johansson S, Umefjord G, Samuelsson E. Stress urinary incontinence and quality of life: a reliability study of a condition-specific instrument in paper and web-based versions. Neurourol Urodyn 2012 Nov;31(8):1242-1246. [doi: 10.1002/nau.22240] [Medline: 22517196]
29. Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's health questionnaire. Med Decis Making 2008;28(1):113-126. [doi: 10.1177/0272989X07301820] [Medline: 17641139]
30. Sjöström M, Umefjord G, Stenlund H, Carlbring P, Andersson G, Samuelsson E. Internet-based treatment of stress urinary incontinence: 1- and 2-year results of a randomized controlled trial with a focus on pelvic floor muscle training. BJU Int 2015 Dec;116(6):955-964 [FREE Full text] [doi: 10.1111/bju.13091] [Medline: 25683075]
31. SCB. Educational attainment of the population 2014 URL: http://www.scb.se/Statistik/UF/UF0506/2014A01S/UF0506_2014A01_SM_UF84SM1601.pdf [WebCite Cache ID 6nYo8RXLz]
32. Henderson JW, Wang S, Egger MJ, Masters M, Nygaard I. Can women correctly contract their pelvic floor muscles without formal instruction? Female Pelvic Med Reconstr Surg 2013;19(1):8-12 [FREE Full text] [doi: 10.1097/SPV.0b013e31827ab9d0] [Medline: 23321652]
33. Albers-Heitner CP, Joore MA, Winkens RA, Lagro-Janssen AL, Severens JL, Berghmans LC. Cost-effectiveness of involving nurse specialists for adult patients with urinary incontinence in primary care compared to care-as-usual: an economic evaluation alongside a pragmatic randomized controlled trial. Neurourol Urodyn 2012 Apr;31(4):526-534. [doi: 10.1002/nau.21204] [Medline: 22275126]
34. Arlandis-Guzman S, Errando-Smet C, Trocio J, Arumi D, Rejas J. Cost-effectiveness analysis of antimuscarinics in the treatment of patients with overactive bladder in Spain: a decision-tree model. BMC Urol 2011 May 20;11:9 [FREE Full text] [doi: 10.1186/1471-2490-11-9] [Medline: 21599928]
35. Montesino-Semper MF, Jimenez-Calvo JM, Cabases JM, Sanchez-Iriso E, Hualde-Alfaro A, García-García D. Cost-effectiveness analysis of the surgical treatment of female urinary incontinence using slings and meshes. Eur J Obstet Gynecol Reprod Biol 2013 Nov;171(1):180-186. [doi: 10.1016/j.ejogrb.2013.08.035] [Medline: 24035322]
36. Eichler H, Kong SX, Gerth WC, Mavros P, Jönsson B. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? Value Health 2004;7(5):518-528 [FREE Full text] [doi: 10.1111/j.1524-4733.2004.75003.x] [Medline: 15367247]
37. Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Rev Pharmacoecon Outcomes Res 2008 Apr;8(2):165-178. [doi: 10.1586/14737167.8.2.165] [Medline: 20528406]
38. The National Board of Health and Welfare. Socialstyrelsen. Nationella riktlinjer för sjukdomsförebyggande metoder URL: https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/18484/2011-11-11.pdf [WebCite Cache ID 6nYncIFmF]
39. Nilsson FO, Linnér L, Samuelsson E, Milsom I. Cost-effectiveness analysis of newer anticholinergic drugs for urinary incontinence vs oxybutynin and no treatment using data on persistence from the Swedish prescribed drug registry. BJU Int 2012 Jul;110(2):240-246 [FREE Full text] [doi: 10.1111/j.1464-410X.2011.10729.x] [Medline: 22093912]
40. Imamura M, Abrams P, Bain C, Buckley B, Cardozo L, Cody J, et al. Systematic review and economic modelling of the effectiveness and cost-effectiveness of non-surgical treatments for women with stress urinary incontinence. Health Technol Assess 2010 Aug;14(40):1-188, iii [FREE Full text] [doi: 10.3310/hta14400] [Medline: 20738930]

Abbreviations

GP: general practitioner
ICER: incremental cost-effectiveness ratio
ICIQ-LUTSqol: International Consultation on Incontinence Modular Questionnaire on Lower Urinary Tract Symptoms and Quality of Life
ICIQ-UI SF: International Consultation on Incontinence Modular Questionnaire on Urinary Incontinence, Short Form

XSL•FO
RenderX

NICE: National Institute for Health and Clinical Exellence
PFMT: pelvic floor muscle training
QALY: quality-adjusted life year
QoL: quality of life
SUI: stress urinary incontinence

Edited by G Eysenbach; submitted 23.01.17; peer-reviewed by G Rortveit, A Wagg; comments to author 08.03.17; revised version received 19.03.17; accepted 20.03.17; published 08.05.17.

Please cite as:
Sjöström M, Lindholm L, Samuelsson E
Mobile App for Treatment of Stress Urinary Incontinence: A Cost-Effectiveness Analysis
J Med Internet Res 2017;19(5):e154
URL: http://www.jmir.org/2017/5/e154/
doi:10.2196/jmir.7383
PMID:28483745

©Malin Sjöström, Lars Lindholm, Eva Samuelsson. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Accelerating Digital Mental Health Research From Early Design and Creation to Successful Implementation and Sustainment

David C Mohr[1], PhD; Aaron R Lyon[2], PhD; Emily G Lattie[1], PhD; Madhu Reddy[3], PhD; Stephen M Schueller[1], PhD

[1]Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[3]Department of Communication Studies, Northwestern University, Evanston, IL, United States

Corresponding Author:
David C Mohr, PhD
Center for Behavioral Intervention Technologies
Department of Preventive Medicine
Northwestern University
750 N Lakeshore Drive, 10th Floor
Chicago, IL, 60611
United States
Phone: 1 3125031403
Fax: 1 312 503 1111
Email: d-mohr@northwestern.edu

Abstract

Mental health problems are common and pose a tremendous societal burden in terms of cost, morbidity, quality of life, and mortality. The great majority of people experience barriers that prevent access to treatment, aggravated by a lack of mental health specialists. Digital mental health is potentially useful in meeting the treatment needs of large numbers of people. A growing number of efficacy trials have shown strong outcomes for digital mental health treatments. Yet despite their positive findings, there are very few examples of successful implementations and many failures. Although the research-to-practice gap is not unique to digital mental health, the inclusion of technology poses unique challenges. We outline some of the reasons for this gap and propose a collection of methods that can result in sustainable digital mental health interventions. These methods draw from human-computer interaction and implementation science and are integrated into an Accelerated Creation-to-Sustainment (ACTS) model. The ACTS model uses an iterative process that includes 2 basic functions (design and evaluate) across 3 general phases (Create, Trial, and Sustain). The ultimate goal in using the ACTS model is to produce a functioning technology-enabled service (TES) that is sustainable in a real-world treatment setting. We emphasize the importance of the service component because evidence from both research and practice has suggested that human touch is a critical ingredient in the most efficacious and used digital mental health treatments. The Create phase results in at least a minimally viable TES and an implementation blueprint. The Trial phase requires evaluation of both effectiveness and implementation while allowing optimization and continuous quality improvement of the TES and implementation plan. Finally, the Sustainment phase involves the withdrawal of research or donor support, while leaving a functioning, continuously improving TES in place. The ACTS model is a step toward bringing implementation and sustainment into the design and evaluation of TESs, public health into clinical research, research into clinics, and treatment into the lives of our patients.

(J Med Internet Res 2017;19(5):e153) doi:10.2196/jmir.7725

KEYWORDS

eHealth; mHealth; methodology

Background

Mental health problems are common [1] and present a tremendous societal burden in terms of cost, morbidity, quality of life, and mortality [2,3]. About two-thirds of people with mental health problems want some form of psychological treatment [4-8]. However, most people experience barriers that prevent access to such treatments [9,10]. Furthermore, there are not enough mental health professionals to meet the needs of the population [11]. To deliver mental health care to all patients who need and desire it, the care system will require services that can be delivered cost effectively, remotely, and in settings

XSL•FO
RenderX

where people most frequently receive care such as primary care or community social services [11].

Digital mental health technologies such as Web-based and mobile apps are frequently cited as potential methods of extending effective care in a cost-effective manner [12]. Randomized controlled efficacy trials have consistently demonstrated that these technology-based tools, when coupled with support from a coach or clinician, can produce benefits similar to those seen for psychological treatments [13-18]. Yet for all the indications that digital mental health interventions can work, evidence is emerging that such interventions have not been effective in routine care settings. The largest implementation trial to date, conducted in England's National Health Service, compared two of the best-established coach-supported Web-based interventions (Beating the Blues and MoodGym) for depression with treatment as usual in primary care and found no significant benefits [19]. Patients did not engage with the Web-based treatment programs and even stopped answering the coaches' calls. This mirrors the experience of large health care organizations in the United States such as Kaiser Permanente, which have unsuccessfully tried many times to implement well-known digital mental health interventions such as Beating the Blues [20]. Similar to the experience in England, patients did not use them, coaches and providers were uncertain how to engage patients, and it was unclear how to integrate these treatments into the care system.

The aim of this paper is to describe the challenges that face the field of digital mental and behavioral health research related to this research-to-practice gap. We focus on the challenges of developing and sustainably implementing technology-enabled treatment services within care systems and not on standalone products such as those available through app stores. While digital behavioral health interventions such as those targeting diet and exercise also face substantial implementation challenges [21,22], the technology, commercial, and clinical contexts are quite different relative to mental health. Thus, while this paper may have relevance for the broader field of technology-enabled behavioral interventions, we begin here with a narrower focus on mental health. We propose a clinical research model that integrates several methodologies to rapidly move from initial design through to implementation and sustainment within care systems.

Challenges

The challenges can be grouped into the three Ds: duration of the research process, design, and denominator of recruitment [20].

Duration of Research Process

In medicine it can take up to 17 years to move 14% of original research into patient care [23]. Clinical science has developed frameworks for evaluation that aim to protect the interests of stakeholders, including patients, providers, and payers, by verifying the efficacy, effectiveness, and safety of interventions. In psychology, models are based on the US Food and Drug Administration's prescribed phases for the evaluation of pharmaceuticals. These five phases are (1) intervention generation and refinement, (2) efficacy in research clinics, (3) efficacy in community settings, (4) effectiveness, and (5) implementation [24,25]. Other models exist such as the deployment-focused model [26] that are less linear and more focused on effectiveness but have a similar number of steps.

These methods have made the process of bringing research into practice both inefficient and ineffective [27-30]. These problems are even more pronounced in digital health, where the pace of technological innovation is rapid and consumer expectations about the capabilities of technologies are rapidly evolving [31-33]. A great deal can change with technology in 17 years: iPhones were first launched in 2007 with Android following in 2008, and today smartphones are the dominant method of accessing the Internet for many Americans [34]. Research in digital health intervention must be translated rapidly into practice to avoid validation of interventions that rely on obsolete technologies [35].

Design

The design of digital interventions has been lacking in a number of ways. First, most digital mental health technologies, which have been evaluated through trials, have largely been designed top-down [36-38], with experts specifying interventions consistent with behavioral strategies derived from evidence-based treatments. The resulting interventions have been primarily Web-based and predominantly psychoeducational via text or video, with some simple interactive tools for common evidence-based practices such as tracking of relevant symptoms and scheduling and monitoring of targeted behaviors. While usability testing has gained currency in recent years [35], the design of these digital interventions has generally not included input from end users. Thus, the field has generally designed interventions to try to get people to do what experts believe is beneficial and has paid far less attention to what users want or how to fit tools into the fabric of users' lives.

Second, when design occurs, it focuses nearly entirely on the technology components. The evidence, however, indicates that digital interventions require some human support to obtain substantive and reasonably reliable benefits [16], and indeed, optimization of the design of human support services may have a greater impact on clinical outcomes than does the design of the technologies [39]. Thus, digital mental health interventions are essentially sociotechnical systems, which we call technology-enabled services (TESs) to emphasize that this patient-supporter interaction is at least as critical as patient-technology and supporter-to-technology.

Finally, implementation and sustainment are rarely considered during the design of TESs. Most TESs tend to be developed by academic or commercial teams outside of the settings where they would eventually be deployed [40]. Thus, designs do not include requirements, processes, and constraints of routine care settings such as designing and implementing referral processes; managing coaching or support in the context of a clinic, practice, or agency; communication needs among providers; the needs of administrators; or the information technology needs.

XSL•FO
RenderX

Denominator problem

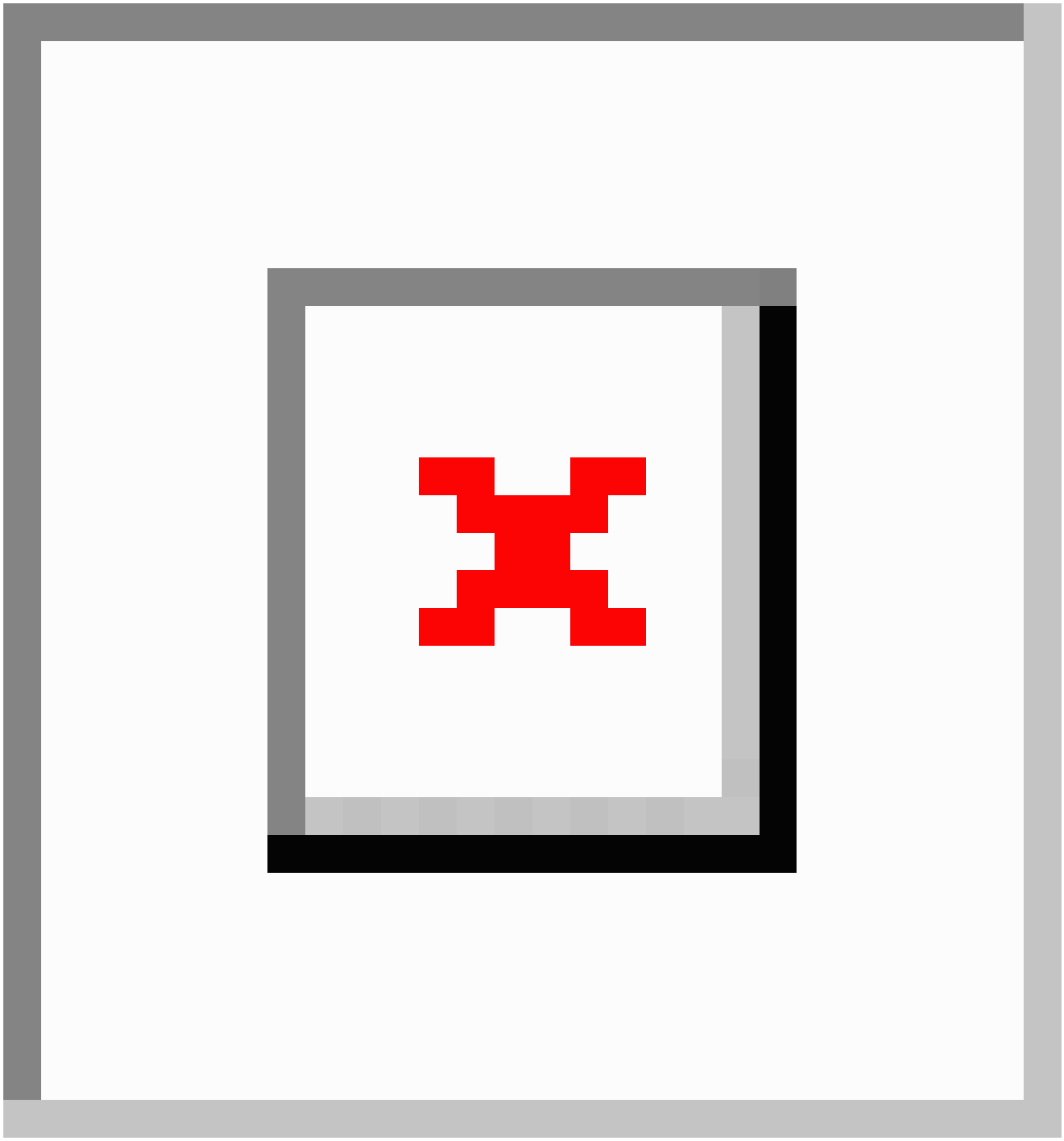
Most trials of digital interventions have recruited from very large pools of potential participants active on the Internet, raising questions of generalizability [20]. For example, one recent study evaluating a coached Web-based intervention had a relatively clear denominator, recruiting 406 participants from a pool of 8.7 million insurance plan members (0.00047% enrollment) over a year [41]. The ability to fill trials from an extraordinarily large pool of potential participants means that those recruited are likely unique and rare individuals who are motivated and willing to engage with developed TESs [20]. This introduces biases limiting the generalizability of both the research findings and the TESs themselves, thereby reducing the likelihood of successful implementation and sustainability.

Accelerated Creation-to-Sustainment Model

What is needed to overcome these challenges is an end-to-end approach that can move rapidly from the initial design stages of a TES through to implementation and sustainment while ensuring that factors critical for the entire process are evaluated and addressed. We propose the Accelerated Creation-to-Sustainment (ACTS) model that builds on existing methodologies, including human-computer interaction, implementation science, and trial methodology and aims to develop and sustainably implement a TES. As displayed in Figure 1, the ACTS model has three targets across three research phases. First, the Create phase aims to produce protocols for the service, an initial, functional version of the technology that supports that service, and an implementation blueprint. The Trial phase builds on an effectiveness-implementation trial [42] to evaluate the efficacy of the TES and implementation plan, optimize the TES and implementation plan, and development metrics for sustainment. The Sustainment phase withdraws research support gradually, leaving a sustainable TES in place. Figure 2 shows the iterative functions (evaluate and design) that occur for the three development targets (TES service and technology components and implementation and sustainment procedures) across each phase.

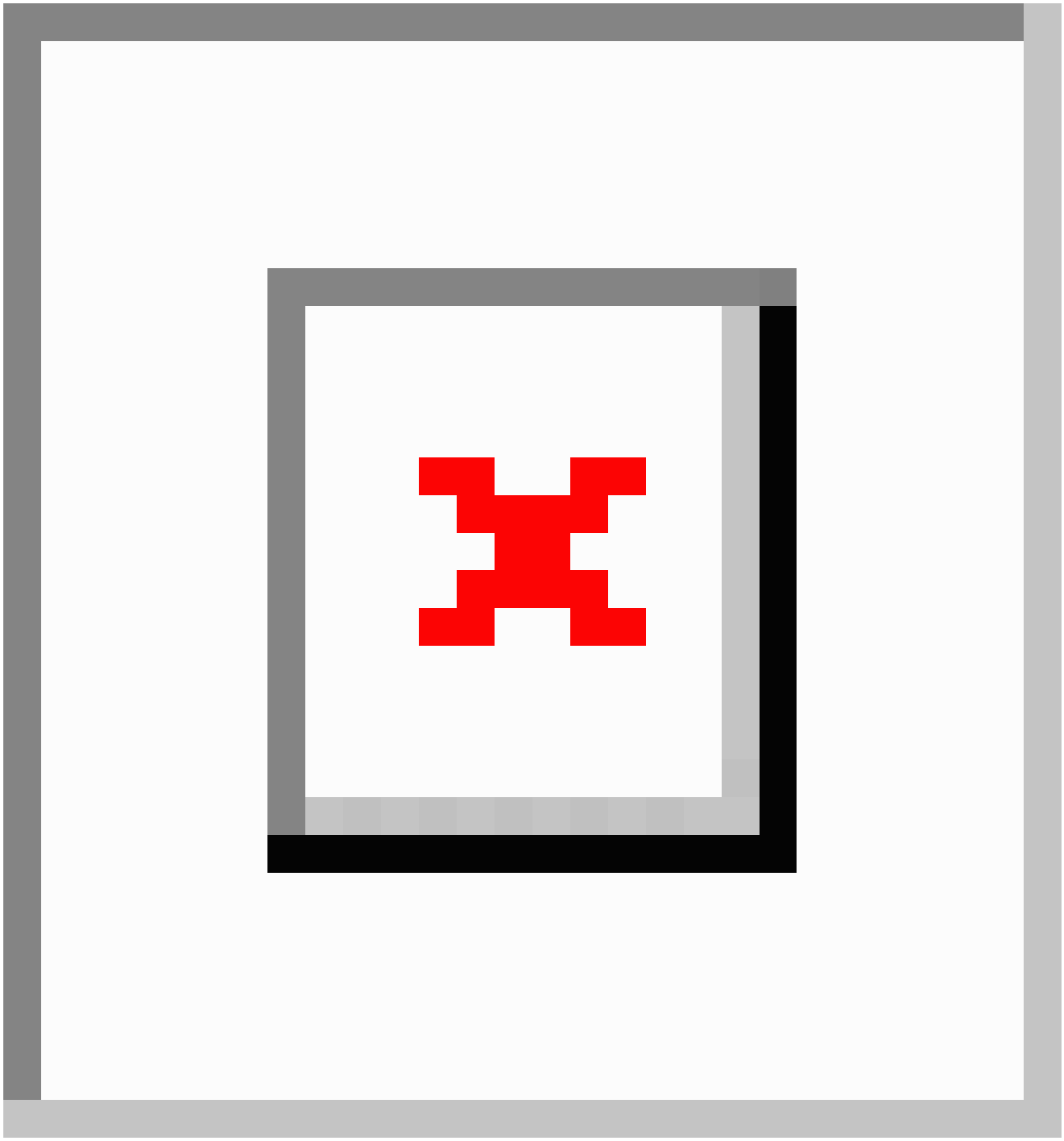
XSL•FO
RenderX

Figure 1. Aims for each development target in each phase.



XSL•FO
RenderX

Figure 2. Iterative evaluative and design functions at each phase.



Functions and Targets

The ACTS model is made up of two basic iterative functions, design and evaluate. These functions persist across each phase although the specific methods employed may vary. This stands in contrast to other methodological models for the development and evaluation of digital technologies for behavior change, which argue that design precedes assessment and sharing (eg, the Integrate, Design, Assess, and Share process [43]).

These functions are applied to three development targets relevant to digital mental health, each of which is critical to the ultimate goal of a sustainably implemented TES. The goal of creating a TES requires definition of the service and the technology. The service refers to the clinical goals, behavioral strategies, and expected roles of each of the actors in the intervention (including the patients, care managers, and any clinicians), much as a treatment protocol might for a standard behavioral intervention. The technology refers to the technologies that enable the service. While much of the research has been on the patient-facing side of technologies, as we shift toward viewing the technology as supporting the service, the design and evaluation of technologies will have to be extended to interfaces and systems that support all stakeholders, including patients, care managers, clinicians, and administrators. Implementation and sustainment targets produce processes that facilitate successful delivery of the TES within a treatment setting and its continued use over an extended period, even after any research support is terminated [44,45].

XSL•FO
RenderX

Virtually every mental health treatment technology has been designed in academic labs or by commercial developers, typically outside of clinical settings. As has been seen generally in health research, the chasm between research and implementation is wide and not easily bridged [29]. While it is difficult to overcome the research-to-practice gap for behavioral interventions generally, the technological components add additional cost and complexities to adapting the intervention to fit the context. To succeed in creating an effective, implementable, and sustainable TES, design and research must take place in the settings where they are expected to be deployed. Implementation and sustainment strategies must be designed into the TES from the very beginning [46].

Phases

The ACTS model has three general phases that rapidly carry the project from initial conceptualization to a sustainable TES. These are the Create, Trial, and Sustainment phases. The aims for each of these phases, displayed in Figure 1, reflect the product or output for that phase.

Create Phase

The first phase is focused on the development of the TES (both technology and services components) and implementation strategies. Although much has been written about the importance of "designing for dissemination" or "starting early for sustainment" [47], implementation and sustainment strategies are rarely developed alongside the services they are intended to support. A clear definition of the service is critical, because this frequently involves new roles and functions and potentially new professionals who can help support technologies and services [48]. This phase employs user-centered design [49,50] that emphasizes deep engagement with key stakeholders (eg, patients, providers, administrators) and their organizational and social contexts to produce a well-designed TES [51].

The aim of the Create phase is to develop a service protocol, a technology, and an implementation/sustainment plan that is ready for first deployment. Success at the end of the Create phase would include having a (1) clearly specified service delivery model that includes basic protocols for any staff involved in TES service provision, (2) a set of enabling technologies that are safe and have preliminary evidence that they meet criteria for technical and functional reliability (technical reliability means that the software and hardware perform consistently according to specifications; by functional reliability, we mean that users can use the technologies to perform the intended actions), and (3) a set of contextually appropriate implementation/sustainment strategies described in an implementation blueprint that are compelling, easy, and reliable to use. We emphasize that the TES and implementation plan do not need to be final products. Both to accelerate the process and avoid wasteful over-optimization early in the process, the goal of the Create phase is to produce a minimally viable TES and implementation plan that are expected to serviceable, which can then be optimized in the next phase.

To meet these aims, design processes are used that we refer to here as user-centered design but which have also been referred to as cooperative design, participatory design, and contextual design [52-54]. This user-centric approach to development recognizes that the success and adoption of technology depends on people's experiences and the ecosystem's support of that technology.

At the core of user-centered design is a focus on collaboration between the designers and stakeholders to provide stakeholders with low-effort ways to inform, contribute to, and interact with technologies during the formative processes. User-centered design's systematic approach to integrate information and iteratively improve products and services is useful to the scientific process of ensuring replicability and generalizability. User-centered design usually begins with acquiring information from key stakeholders to understand user requirements that can contribute to an initial design document [55]. This design document leads to low-fidelity prototypes, which are designs on paper, videos, or nonfunctional tools that contain key functions of interests. Such prototypes are useful because they can be presented to testers for evaluation and can be redesigned based on that evaluation. This allows assumptions to be tested early and makes use of the stakeholders' expertise in the domains in which they will use and support technologies in practice. This cycle continues with progressively more fully functioning, higher fidelity versions until the product is functional and free of major design flaws, at which point it is ready for initial implementation.

User-centered design typically begins with acquiring a basic understanding of the goals, challenges, and motivations of each of the stakeholders, particularly in relationship to the management of mental health conditions. In addition, an understanding of the affordances and constraints associated with organizational factors is important for developing the implementation blueprint.

Initial evaluation of user requirements, particularly when little is known, can begin with individual interviews and focus groups to understand the needs, wants, and limitations of key stakeholders [56,57]. However, gathering information about requirements often involves watching in addition to asking. Workflow observations can help illuminate important TES features (both technical and coaching) and how TES might be integrated into the larger care context. These observations can help identify the organizational barriers that users would face in using the TES that may not emerge from interviews or focus groups [58,59].

After a basic understanding is acquired, more interactive methods can deepen an understanding of the kinds of design elements that may prove useful. For example, codesign workshops bring together the researchers and stakeholders to help representative end-users begin designing their own solutions that address their needs [50,60,61]. Stakeholder participants can draw sketches of features or tools that may be of interest to them [62], describe potential services or interactions, and even interact with paper prototypes [63]. The role of the researcher is to help participants translate these solutions into effective design constructs [64]. At the core of the codesign concept is the idea that stakeholders themselves are best positioned to articulate these solutions.

XSL•FO
RenderX

In user-centered design, as user requirements become clear, the iterative process of design, evaluate, redesign begins. This iterative process refines the design ideas with collaboration between the designers and stakeholders that eventually will be developed into the technologies, services, and implementation strategies [50,64]. Task analysis is one method, in which stakeholders are presented low-fidelity prototypes, which are simple versions of a tool with no real functionality, and asked to complete tasks that are part of the service design. These actions are observed, recorded, and analyzed. Based on errors or problems, corrections can be made to the design of the prototype or to the service protocol, and testing begins again [65]. Heuristic evaluations require experts who judge the compliance of prototypes with recognized usability principles [66]. Cognitive walkthroughs ask stakeholders to work through a series of tasks, which can be useful for evaluating implementation protocols [67]. Once the major flaws in the service protocol, technology, and implementation plan have been eliminated, the TES can progress to the trial stage. We note that because further refinements can be made during the trial, it is not necessary to test to perfection. It is preferable to get into the field as quickly as possible.

Optimization, Effectiveness, and Implementation Hybrid Trial Phase

The second phase is focused on the evaluation of the TES and implementation strategies with regard to clinical goals. The objectives in this phase are to optimize the TES so that it meets its clinical objectives and is usable by all stakeholders, evaluate its effectiveness, and successfully implement the TES in the setting. These objectives are achieved by testing in real-world clinical contexts rather than research settings. Traditional methods of evaluation would require a sequential, phased series of trials to separately achieve each goal [24]. The ACTS model argues that Optimization, Effectiveness, and Implementation (OEI) studies can occur simultaneously in an OEI Hybrid trial. Optimization extends the Create phase from early testing to longitudinal, iterative evaluation and redesign in the context of deployment, continuing until all major problems have been identified and resolved. Effectiveness is evaluated to provide evidence that the intended outcomes are achieved, protecting stakeholders' interests. Implementation is evaluated to ensure that the TES can be seamlessly deployed in the intended setting; any implementation problems result in redesign of the implementation plan. Combining these goals in a single trial structure should dramatically accelerate the rate of translational gains of effective TESs, providing useful information to key decision makers in a far timelier manner.

The OEI Hybrid trial builds on Curran's work and is similar to the Type 2 trial in which effectiveness and implementation outcomes are tested simultaneously [42], with effectiveness components focusing on patient-level clinical outcomes and implementation focusing on contextual and organizational factors such as adoption, uptake, treatment fidelity, costs and cost effectiveness, and efficiency [45,68,69]. While this condensation of effectiveness and implementation trials can reduce the time from initial concept to delivery, the phased research steps before this (pilots and efficacy) will still likely result in obsolescence in the technological elements of the TES. For this reason, as many have argued, the phase model of treatment development and evaluation, which typically lasts many years, adds insufficient value above that provided by effectiveness trials [28]. The value provided by phase models is not worth the threat posed by technological obsolescence [31].

To accelerate the design-implementation pipeline for TESs we argue that optimization of the TES and the implementation plan using the iterative functions of evaluation and redesign must be integrated into the hybrid trial methodology. A traditional, linear phase model would require that a pilot or field trial be conducted prior to launching a TES in a trial to ensure the intervention and technology is working as intended [24,25]. However, optimization outside the clinical context is of limited benefit because many of the challenges of optimization are related to the deployment context. Thus, we argue that the additional 6 to 12 months spent piloting outside the clinical context is overoptimization that fails to collect necessary contextual information and is therefore superfluous. Optimization of the TES and implementation plan together is required because the TES must adjust to issues that arise from unanticipated contextual and organizational issues. Indeed, such optimization during a trial is so common that it is recognized in the Consolidated Standards of Reporting Trials (CONSORT) eHealth guideline, which requires reporting of changes made to the intervention during the trial [70]. Accordingly, we argue that once the TES and implementation plan meet basic usability requirements and there is reason to be believe the TES is safe and may be useful to patients, the OEI Hybrid Trial can begin.

Optimization can continue during the Hybrid Trial, relying on the dynamic, iterative processes of user-centered design. Optimization is facilitated by the articulation of clear specification for objectives, constraints, and design variables [71].

Objectives of optimization are the outcomes that are to be maximized by the TES and the implementation plan. Typically, the primary objective of a TES is a reduction in clinical markers (eg, symptoms of depression). Secondary objectives might include use variables such as communication patterns between the patient and provider. Technology objectives typically include markers of usability including satisfaction, usefulness, ease of use, and absence of errors [72]. Implementation objectives may include measures of penetration, adoption, and cost [45,73].

Constraints refer to conditions that must be met for the viability of the service, technology, and implementation. These constraints can be identified and defined during the Create phase as information regarding the needs and limitations of the stakeholders and organization is accumulated.

One class of constraints relates to methods required to ensure that optimization does not threaten the internal validity of the Hybrid Trial. Making changes to the TES or implementation strategy flies in the face of traditional methods drawn from trial methodology based on pharmaceutical trials in which the active agent is fixed. While it is questionable how locked down behavioral interventions have ever been in randomized controlled trials, technology-based interventions demand changes to fix bugs and prevent obsolescence. The Trials of

XSL•FO
RenderX

Intervention Principles method offers ways of integrating continuous quality improvement into the trial while imposing constraints on the scope of design changes to ensure that the TES and implementation plan remain conceptually consistent over the course of the study [32]. A clear framing and operationalization of the principles being evaluated serve as constraints that limit the types of changes that can be made through optimization. Documentation and reporting of changes creates transparency. For example, an intervention focused on treating depression in primary care using care coordinator–supported mobile tools could not suddenly change the outcome or shift to a fundamentally different intervention model (eg, psychotherapy or pharmacotherapy). If an aim is to examine a TES administered by care managers, shifting to mental health providers would constitute a new treatment and trial. On the other hand, some alterations to the patient-facing or care manager–facing tools or the frequency of contacts may be within scope.

Constraints may also be imposed by the organization. For example, the amount of time available from staff, allowable operations in a clinic, or technological requirements may all constitute constraints. New constraints may also be identified during the OEI Hybrid Trial as unintended consequences are unearthed. For example, a TES may end up serving as a conduit into traditional services or help retain patients in treatment, thereby aggravating care capacity problems [74], resulting in constraints to manage the emergent problem.

Design variables refer to those specifications of the service, technology, and implementation that are modifiable and that may affect the objectives (eg, definition of the service, design of technologies, or implementation blueprint). Optimization seeks to refine the design variables to maximize the objectives without exceeding the constraints.

Relative to the Create phase, where information gathering is intended to result in initial definition and development of the TES and implementation approaches, information gathering in the optimization part of the Hybrid Trial phase addresses adjusting the fit for the organizational contexts. This includes responding to unanticipated problems and opportunities encountered in implementation, removing unnecessary components, and improving processes and functionality. Information for optimization, particularly for objectives and constraints, may be collected through defined methods such as the acquisition of data on clinical outcomes, system use, and stakeholder satisfaction.

Allowing iterative processes of evaluation and design to occur during the trial phase supports ongoing learning to continuously adapt and improve the TES and its implementation. The TES optimization evaluation can employ use data and outcome data collected for the effectiveness and implementation trial as well as semistructured user feedback interviews. Numerous methodologies are available to support this learning and optimization including A/B testing [75], continuous evaluation of evolving behavioral intervention technologies [76], multiphase optimization strategy [77], and trials of intervention principles [32]. The idea of changing an intervention during the trial is generally accepted in digital mental health research, and indeed, the CONSORT eHealth guideline includes the requirement of reporting these changes in the publication of trials [70]. Methods of managing, constraining, and documenting the changes have been articulated that can preserve the validity of the trials necessary to generalize to new settings [32,78]. One would expect that optimization efforts would be more intensive at the beginning of the OEI Hybrid Trial and would diminish as problems are identified and corrected.

Sustainment Phase

The final phase tapers and removes research infrastructure and support for a TES that has met the aims of the Hybrid Trial phase [44]. Sustainment refers to the continued use of an intervention in a manner that brings benefits after this support is removed [45]. Research teams can play a role in this tapering to help facilitate sustainment processes (eg, ensuring the collection and accuracy of data and supporting analysis). However, eventually these roles should be assumed by clinical staff, allowing researchers to recede into the background and leaving the organization to drive a sustainable TES without assistance. The aim of this phase is not just to study sustainment but to leave the health care site with a TES that can function without external research support and continue to adapt to problems that might arise.

Even when initial implementation is successful, sustainment has rarely been examined and, when it has, a lack of sustainment is evident [79]. There are numerous potential threats to sustainment, particularly for interventions that rely on human services in the context of complex, multilayered systems including changes within the organization (eg, staff turnover, change in service organization, leadership, information technology infrastructure) or the larger context (eg, consumer technologies, patient attitudes and preferences, funding models) [80,81].

Given the challenges of sustainment, it is unlikely that a “set it and forget it” model of sustainment will be successful. Thus, the iterative cycle of evaluate and redesign, both to address problems that arise and adjust to changing needs, demands, and contexts, is likely required into the Sustainment phase. Emerging research suggests that sustainment requires active components to identify and correct potential problems and continue to improve and adapt the TES to meet changes in the organization, patient population, and larger context [80,81]. More recent conceptualizations of sustainment have incorporated features of learning health care systems that mirror our evaluation and redesign functions, using data inputs to continuously monitor and dynamically correct and adapt the intervention [35,46,82].

Data collection for dynamic sustainment models must be as effortless as possible so as not to add burden to care systems that are already functioning at or beyond capacity. Fortunately, TESs provide an increasingly large amount of unobtrusively collected data that can be used to monitor the health of the intervention along the entire pipeline. For example, penetration can be monitored by the flow of referrals through various stages of the TES. In a primary care setting, such metrics might include the proportion of appropriate patients referred to TES, proportion of those referred contacted by TES staff, proportion of those contacted who initiated TES treatment, and proportion of those

XSL•FO
RenderX

who initiated TES treatment who complete it. Effectiveness can be monitored through symptom measures that are commonly used in TESs for self-monitoring as well as any other electronic records that may be available. Fidelity of the provider to the service protocol is an important component of sustainment. Fidelity can be monitored by having structured protocols represented within provider-facing technology components such as a digital checklist for specific provider actions during patient interactions. Patient satisfaction and quality of use can be monitored with qualitative and quantative feedback from patients and technology use data.

These largely unobtrusive sustainment metrics could be developed for the specific TES during the OEI Hybrid Trial. Fidelity to service models (eg, frequency and timing of contacts) can be assessed in conjunction with the supervisory functions that are important to the success of sustainment, providing benchmarks against which passively collected data from TES use can be used to develop empirically derived fidelity markers [83]. Methods for assessing fidelity will likely become increasingly unobtrusive as natural language processing capabilities improve, enabling automated extraction of fidelity markers from phone calls and messaging [84].

Redesign efforts at the Sustainment phase should be minor. Once a TES is fully implemented and integrated, substantial redesign and redevelopment of technical components can be many times more complex and costly compared to identifying and optimizing during earlier phases [85]. Similarly, major changes to services and workflows are disruptive and challenging to put into practice once integration of TESs is complete.

While the cost and complexity of major redesign efforts during Sustainment is high, all systems eventually require substantial redesign to meet changing needs, preferences, and contexts. In such instances, reintroduction of optimization methods from the OEI Hybrid Trial may help to understand context-specific objectives and constraints to alter the design. This is represented by the recursive arrows in Figure 2.

Hypothetical Example

It is often helpful to have a concrete example of a model to help convey concepts that can be difficult to grasp through abstract descriptions and discussions. Unfortunately, given our presentation of the ACTS model is an initial proposal, no example currently exists. However, in Table 1 we present a brief hypothetical example of the development and implementation of a TES for a primary care setting. We provide examples of methods and evaluation at each stage, along with the resulting products and possible outcomes.

XSL•FO
RenderX

Table 1. Hypothetical example of the Accelerated Creation-to-Sustainment model for a technology-enabled service to treat depression in primary care.

Phase and stage	Methods and evaluation	Design products and outcomes
Create		
User-centered design: TES[a]	Individual interviews with primary care physicians and staff Representative patients, administration in psychiatry, primary care, and medical director's office Codesign workshops with patients, care managers, primary care physicians, mental health specialists, and researchers and developers Quality assurance conducted with staff examining technical and functional reliability Usability testing with patients and care managers	Sketches and paper prototypes Design document and specifications A functioning TES that can be delivered by care managers Patient-facing tools are mobile app-based; care manager tools are computer-based TES has received only limited longitudinal testing in target population but is minimally viable
User centered design: Implementation plan	Contextual evaluation Individual interviews with primary care physicians and staff, representative patients, and administrators Workflow observations of physicians and care managers Cognitive walkthroughs of typical clinical tasks	Implementation hassle map that outlines where hassles and breakdowns in care occurs Implementation plan defined Plan leverages existing processes in which patients assessed in primary care are referred to TES Processes are defined for TES support, supervision, quality, and safety monitoring Patient population and TES use cases defined
OEI[b] Hybrid Trial		
Trial design	Blocked randomized controlled trial	Clinical and services outcomes Improvements in symptoms of depression; use of TES by patients and care managers Patients referred to, initiating, and completing TES Efficiency of care managers at providing services
Optimization	Evaluation of outcomes, use data, and periodic user feedback interviews As problems in TES are identified, changes made and logged	Change log with changes described TES and implementation plan updated iteratively until they function well
Evaluation	Symptom evaluation by self-report from all patients at all sites meeting criteria for depression TES use data monitored Care management fidelity monitored using random ratings of recorded calls and communication logs Implementation monitored Medical and mental health costs monitored and compared across sites All metrics evaluated over time to explore changes in effectiveness	Changes in symptoms of depression; use of TES by patients and care managers
Implementation	Qualitative and quantitative evaluation of implementation TES use data monitored Care management fidelity monitored using random ratings of recorded calls and communication logs Implementation monitored Medical and mental health costs monitored and compared across sites All metrics evaluated over time to explore changes in implementation Interviews with system stakeholders (eg, care managers, primary care physicians and staff, administrators) to assess perception of benefit or barriers	Fidelity scores of services provided by case managers Service utilization and costs of services by patients Suggested enhancements for implementation (eg, training, deimplementation of ineffective elements)

XSL•FO
RenderX

Phase and stage	Methods and evaluation	Design products and outcomes
Sustainment		
	Unobtrusive markers	Ongoing benefits of TES system
	Ongoing monitoring of symptoms through within treatment evaluation, system usage by patients and care providers, markers of fidelity (eg, pattern of care manager outreach, outcomes across care managers), referral patterns	A functioning TES that is supported by clinical staff and feeds appropriate and actionable information back to staff, providers, and administrators

[a]TES: technology-enabled services.

[b]OEI: optimization, effectiveness, and implementation.

Managing Potential Failure Points and Problems

There are potential risks and failure points within each phase. The user-centered design process in the Create phase should result in a design and initial product. The primary risk during the design phase is that design challenges will take longer than expected to address. For example, aligning the objectives and preferences of patients, providers, and administrators may prove more challenging than initially expected. There are several potential failure points during the OEI Hybrid Trial. First, it is possible that design and implementation cannot be optimized sufficiently to create a workable solution. For example, it is possible that not enough people are recruited into the trial, which would reduce the amount of information gained and provide insufficient resources or rationale to continue optimization procedures. As we note above, this would be an indicator of a problem in design of the TES or implementation plan and could require returning to the Create phase (represented by the recursive arrows in Figure 2). It is also possible that even with good optimization and implementation, the intervention is simply not effective, which could result in a return to user-centered design in the Create phase or abandonment of the effort. Assuming successful implementation and evidence of effectiveness, the program can move forward to the Sustainment phase. There are numerous potential contributors to a failure of sustainment of a TES that has made it through the OEI Hybrid Trial, including lack of support and guidance of staff and care managers, lack of resources, lack of buy-in from senior management, or failure to detect and address problems that arise over time [83,86,87].

We recognize that the accelerated process evaluation, in particular moving a TES into a trial phase earlier in the optimization process, opens the possibility of increased risk to participants. Such risk is low; negative outcomes among patients receiving digital interventions is rare, is lower than in control conditions, and does not appear to occur at rates higher than standard treatments [88]. However, while such risk is low, it is important to evaluate and guard against. Monitoring for patient risk should occur in all phases, with particular attention paid to the possibility of iatrogenic effects during the Create and early OEI Hybrid Trial phases. Such monitoring should use both quantitative and qualitative methods to detect potential iatrogenic effects (eg, “Did you experience any problems or negative consequences in your treatment?”) and to obtain more detailed information when such effects are detected [89]. Introducing any new service into a care setting has the potential to introduce other negative effects at the provider and system perspective as well. For example, TESs could produce such a wealth of information that providers and systems are not able to process it into actionable guidance when appropriate (such as safety alerts in the case of suicidal patients).

Conclusion

There is an enormous research-to-practice gap in digital mental health, with strong and growing evidence from efficacy trials over more than 15 years yet virtually no successful and sustainable implementation. This failure is due to many factors. Our research models are exceedingly reductive, compartmentalizing aims into individual research programs (such as phase models) to answer isolated questions of design, intervention refinement, efficacy, effectiveness, implementation, and sustainment. This is not only inefficient but also ignores the intricacy of delivering TESs involving rapidly changing technological environments into the varied and complex circumstances of individual patients’ lives.

The ACTS model provides a framework for accelerating research and integrating design, evaluation, and sustainable implementation into a unified effort. Evaluation in the Create phase is intensive and qualitative, becoming more quantitative in the Trial phase, and finally leaning heavily on pragmatic methods such as unobtrusive, largely automated measurement in Sustainment. Design flexibility is maximal in the initial design phases and becomes increasingly hard to change and adjust as the TES becomes developed, deployed, and integrated into care settings. It is imperative that we build implementation and sustainment into the design process from the very inception, when there is maximum flexibility. This is especially true when technology is involved; it is far more cost effective to adjust and fix design problems early as opposed to once the technologies and services are in place [90].

While many of the components of this framework, including user-centered design, hybrid trials, integration of optimization and evaluation, and sustainment, have previously been articulated and applied in many contexts, they have not been put together in a single organized model. This paper is intended as a draft of a general blueprint for a new, expedited approach to research in TESs. We would expect and welcome disagreement and refinement. However, we can no longer afford to consider clinical research as divorced from public health. In a world of rapidly evolving technologies, we can no longer wait

XSL•FO
RenderX

more than a decade to move research into practice. The ACTS model is a step toward bringing implementation and sustainment into design and evaluation, research into clinics, public health into clinical research, and treatment into the lives of our patients.

Acknowledgments

Research grants from the US National Institute of Mental Health support the writing efforts of Dr Mohr (P20 MH090318, R01 MH095753, and R01 MH100482), Dr Schueller (K08 MH102336), and Dr Lyon (K08 MH095939).

Conflicts of Interest

Dr Mohr has received an honorarium from Optum Behavioral Health. The remaining authors declare they have no potential conflicts of interest.

References

1. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
2. Hu T. Perspectives: an international review of the national cost estimates of mental illness, 1990-2003. J Ment Health Policy Econ 2006 Mar;9(1):3-13. [Medline: 16733267]
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
4. Brody DS, Khaliq AA, Thompson TL. Patients' perspectives on the management of emotional distress in primary care settings. J Gen Intern Med 1997 Jul;12(7):403-406 [FREE Full text] [Medline: 9229277]
5. Dwight-Johnson M, Sherbourne CD, Liao D, Wells KB. Treatment preferences among depressed primary care patients. J Gen Intern Med 2000 Aug;15(8):527-534 [FREE Full text] [Medline: 10940143]
6. Bedi N, Chilvers C, Churchill R, Dewey M, Duggan C, Fielding K, et al. Assessing effectiveness of treatment of depression in primary care: partially randomised preference trial. Br J Psychiatry 2000 Oct;177:312-318 [FREE Full text] [Medline: 11116771]
7. Priest RG, Vize C, Roberts A, Roberts M, Tylee A. Lay people's attitudes to treatment of depression: results of opinion poll for Defeat Depression Campaign just before its launch. BMJ 1996 Oct 5;313(7061):858-859 [FREE Full text] [Medline: 8870574]
8. Churchill R, Khaira M, Gretton V, Chilvers C, Dewey M, Duggan C, Nottingham Counselling and Antidepressants in Primary Care (CAPC) Study Group. Treating depression in general practice: factors affecting patients' treatment preferences. Br J Gen Pract 2000 Nov;50(460):905-906 [FREE Full text] [Medline: 11141877]
9. 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]
10. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006 Dec;32(3):254-258. [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
11. Kazdin A, Blase S. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
12. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
13. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-154:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
14. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
15. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
16. 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]

XSL•FO
RenderX

17. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
18. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
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;351:h5627 [FREE Full text] [Medline: 26559241]
20. Mohr D, Weingardt K, Reddy M, Schueller S. Three problems with current digital mental health research—and three things we can do about it. Psychiatr Serv (in press) 2017:1 (forthcoming).
21. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-S12. [doi: 10.7326/M13-3005] [Medline: 25402403]
22. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. JAMA 2016 Sep 20;316(11):1161-1171. [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
23. Balas E, Boren S. Managing clinical knowledge for healthcare improvements. In: van Bemmel JH, McCray AT, editors. Yearbook of Medical Informatics. Stuttgart: Schattauer; 2000:2000-2070.
24. Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M. Reenvisioning clinical science: unifying the discipline to improve the public health. Clin Psychol Sci 2014 Jan 1;2(1):22-34 [FREE Full text] [doi: 10.1177/2167702613497932] [Medline: 25821658]
25. Rounsaville B, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from stage I. Clin Psychol Sci Pract 2001;8(2):133-142. [doi: 10.1093/clipsy.8.2.133]
26. Weisz J, Jensen A, McLeod B, Hibbs E, Jensen P. Milestones and methods in the development and dissemination of child and adolescent psychotherapies: a new deployment-focused model. In: Hibbs ED, editor. Psychosocial Treatments for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice, 2nd edition. Washington: American Psychological Association; 2005:9-39.
27. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
28. Glasgow RE, Davidson KW, Dobkin PL, Ockene J, Spring B. Practical behavioral trials to advance evidence-based behavioral medicine. Ann Behav Med 2006 Feb;31(1):5-13. [doi: 10.1207/s15324796abm3101_3] [Medline: 16472033]
29. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003 Aug;93(8):1261-1267. [Medline: 12893608]
30. Institute of Medicine. Envisioning a Transformed Clinical Trials Enterprise in the United States: Establishing an Agenda for 2020. Washington: National Academies Press; 2012.
31. Patrick K, Hekler E, Estrin D, Mohr D, Riper H, Crane D. Rapid rate of technological development and its implications for research on digital health interventions. Am J Prev Med (in press) 2017:1.
32. 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]
33. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
34. Smith A. Record shares of Americans now own smartphones, have home broadband. Washington: Pew Internet and American Life Project; 2017. URL: http://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/ [accessed 2017-04-28] [WebCite Cache ID 6q4CMWTWU]
35. Lyon AR, Wasse JK, Ludwig K, Zachry M, Bruns EJ, Unützer J, et al. The Contextualized Technology Adaptation Process (CTAP): optimizing health information technology to improve mental health systems. Adm Policy Ment Health 2016 May;43(3):394-409. [doi: 10.1007/s10488-015-0637-x] [Medline: 25677251]
36. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychol Med 2003 Feb;33(2):217-227. [Medline: 12622301]
37. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
38. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. J Med Internet Res 2010 Dec 19;12(5):e48 [FREE Full text] [doi: 10.2196/jmir.1370] [Medline: 21169164]
39. Alfonsson S, Olsson E, Linderman S, Winnerhed S, Hursti T. Is online treatment adherence affected by presentation and therapist support? A randomized controlled trial. Comput Hum Behav 2016 Jul;60:550-558. [doi: 10.1016/j.chb.2016.01.035]

XSL•FO
RenderX

40. Lyon A, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol Sci Pract 2016;23(2):180-200.
41. 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]
42. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
43. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
44. Rabin BA, Brownson RC, Haire-Joshu D, Kreuter MW, Weaver NL. A glossary for dissemination and implementation research in health. J Public Health Manag Pract 2008;14(2):117-123. [doi: 10.1097/01.PHH.0000311888.06252.bb] [Medline: 18287916]
45. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011 Jan;38(1):4-23 [FREE Full text] [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]
46. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013;8:117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]
47. Brownson RC, Jacobs JA, Tabak RG, Hoehner CM, Stamatakis KA. Designing for dissemination among public health researchers: findings from a national survey in the United States. Am J Public Health 2013 Sep;103(9):1693-1699. [doi: 10.2105/AJPH.2012.301165] [Medline: 23865659]
48. Ben-Zeev D, Drake R, Marsch L. Clinical technology specialists. BMJ 2015 Feb 19;350:1. [Medline: 25697164]
49. Schuler D, Namioka A, editors. Participatory Design: Principles and Practices. Hillsdale: Lawrence Erlbaum Associates; 1993.
50. Muller M. Participatory design: the third space in HCI. In: The Human-Computer Interaction Handbook. Hillsdale: Lawrence Erlbaum Associates; 2003:1051-1068.
51. Ratwani RM, Fairbanks RJ, Hettinger AZ, Benda NC. Electronic health record usability: analysis of the user-centered design processes of eleven electronic health record vendors. J Am Med Inform Assoc 2015 Nov;22(6):1179-1182. [doi: 10.1093/jamia/ocv050] [Medline: 26049532]
52. Norman D, Draper S, editors. User-Centered System Design: New Perspectives on Human-Computer Interaction, 1st edition. Hillsdale: Lawrence Erlbaum Associates; 1986.
53. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications; 2004:445-456.
54. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
55. Vredenburg K, Mao J, Smith P, Carey T. A survey of user-centered design practice. 2002 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2002; Minneapolis.
56. Lazar J, Feng J, Hochheiser H. Research methods in human-computer interaction. Hoboken: John Wiley & Sons; 2010:0470723378.
57. Krueger R, Casey M. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks: Sage Publications; 2014.
58. Reddy M, Shabot M, Bradner E. Evaluating collaborative features of critical care systems: a methodological study of information technology in surgical intensive care units. J Biomed Inform 2008 Jun;41(3):479-487 [FREE Full text] [doi: 10.1016/j.jbi.2008.01.004] [Medline: 18316247]
59. Reddy MC, McDonald DW, Pratt W, Shabot MM. Technology, work, and information flows: lessons from the implementation of a wireless alert pager system. J Biomed Inform 2005 Jun;38(3):229-238 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.010] [Medline: 15896696]
60. Bessant J, Maher L. Developing radical service innovations in healthcare? The role of design methods. Int J Innov Manage 2009;13(04):555-568.
61. Mackay W. The interactive thread: exploring methods for multi-disciplinary design. 2004 Presented at: Proceedings of the 5th conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques; 2004; Cambridge.
62. King S, Conley M, Latimer B, Ferrari D. Co-Design: A Process of Design Participation. New York: Van Nostrand Reinhold Company; 1989.
63. Sefelin R, Tscheligi M, Giller V. Paper prototyping—what is it good for? A comparison of paper- and computer-based low-fidelity prototyping. 2003 Presented at: CHI '03 extended abstracts on Human factors in computing systems; 2003; Ft. Lauderdale.
64. Spinuzzi C. The methodology of participatory design. Tech Commun 2005;52(2):163-174.
65. Diaper D, Stanton N. The Handbook of Task Analysis for Human-Computer Interaction. Mahwah: Lawrence Erlbausm Assoc; 2003.

XSL•FO
RenderX

66. Nielsen J. Finding usability problems through heuristic evaluation. 1992 Presented at: CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 1992; Monterey.
67. Militello LG, Hutton RJ. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. Ergonomics 1998 Nov;41(11):1618-1641. [doi: 10.1080/001401398186108] [Medline: 9819578]
68. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009 Jan;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
69. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
70. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
71. Kristensson P, Bi X, Howes A, Oulasvirta A, Murray-Smith R, Thimbleby H. Principles, techniques, and perspectives on optimization. 2015 Presented at: CHI ' 15 Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems; 2015; Seoul.
72. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012 Jun;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
73. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009 Jan;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
74. Kenter RM, van de Ven PM, Cuijpers P, Koole G, Niamat S, Gerrits RS, et al. Costs and effects of Internet cognitive behavioral treatment blended with face-to-face treatment: Results from a naturalistic study. Internet Interv 2015 Mar;2(1):77-83. [doi: 10.1016/j.invent.2015.01.001]
75. Hekler E, Klasnja P, Froelich JE, Buman M. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2013; Paris.
76. Mohr DC, Cheung K, Schueller SM, Hendricks BC, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
77. 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]
78. 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]
79. Wiltsey SS, Kimberly J, Cook N, Calloway A, Castro F, Charns M. The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. Implement Sci 2012 Mar 14;7:17 [FREE Full text] [doi: 10.1186/1748-5908-7-17] [Medline: 22417162]
80. Fixsen DL, Blase KA, Naoom SF, Wallace F. Core Implementation Components. Res Soc Work Prac 2009 May 27;19(5):531-540. [doi: 10.1177/1049731509335549]
81. Glisson C, Schoenwald SK. The ARC organizational and community intervention strategy for implementing evidence-based children's mental health treatments. Ment Health Serv Res 2005 Dec;7(4):243-259. [doi: 10.1007/s11020-005-7456-1] [Medline: 16320107]
82. Chambers DA, Feero WG, Khoury MJ. Convergence of implementation science, precision medicine, and the learning health care system: a new model for biomedical research. JAMA 2016 May 10;315(18):1941-1942. [doi: 10.1001/jama.2016.3867] [Medline: 27163980]
83. Novins DK, Green AE, Legha RK, Aarons GA. Dissemination and implementation of evidence-based practices for child and adolescent mental health: a systematic review. J Am Acad Child Adolesc Psychiatry 2013 Oct;52(10):1009-1025 [FREE Full text] [doi: 10.1016/j.jaac.2013.07.012] [Medline: 24074468]
84. Gallo C, Pantin H, Villamar J, Prado G, Tapia M, Ogihara M, et al. Blending qualitative and computational linguistics methods for fidelity assessment: experience with the Familias Unidas preventive intervention. Adm Policy Ment Health 2015 Sep;42(5):574-585 [FREE Full text] [doi: 10.1007/s10488-014-0538-4] [Medline: 24500022]
85. Boehm B, Basili V. Top 10 list [software development]. Computer 2001;34(1):135-137. [doi: 10.1109/2.962984]
86. Aarons GA, Green AE, Trott E, Willging CE, Torres EM, Ehrhart MG, et al. The roles of system and organizational leadership in system-wide evidence-based intervention sustainment: a mixed-method study. Adm Policy Ment Health 2016 Nov;43(6):991-1008. [doi: 10.1007/s10488-016-0751-4] [Medline: 27439504]
87. Ford JH, Alagoz E, Dinauer S, Johnson KA, Pe-Romashko K, Gustafson DH. Successful organizational strategies to sustain use of A-CHESS: a mobile intervention for individuals with alcohol use disorders. J Med Internet Res 2015 Aug 18;17(8):e201 [FREE Full text] [doi: 10.2196/jmir.3965] [Medline: 26286257]

XSL•FO
RenderX

88. Ebert DD, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol Med 2016 Oct;46(13):2679-2693. [doi: 10.1017/S0033291716001562] [Medline: 27649340]
89. Rozental A, Andersson G, Boettcher J, Ebert D, Cuijpers P, Knaevelsrud C. Consensus statement on defining and measuring negative effects of Internet interventions. Internet Interv 2014;1:12-19.
90. Stecklein J, Dabney J, Dick B, Haskins B, Lovell R, Moroney G. Error cost escalation through the project life cycle. 2004 Presented at: International Council on Systems Engineering (INCOSE): 14th Annual International Symposium; 2004; Toulouse.

Abbreviations

ACTS model: Accelerated Creation-to-Sustainment model
CONSORT: Consolidated Standards of Reporting Trials
OEI Hybrid Trial: Optimization, Effectiveness, Implementation Hybrid Trial
TES: technology-enabled service
UCD: user-centered design

Edited by G Eysenbach; submitted 20.03.17; peer-reviewed by D Chambers, D Ben-Zeev; comments to author 05.04.17; revised version received 15.04.17; accepted 16.04.17; published 10.05.17.

Please cite as:
Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM
Accelerating Digital Mental Health Research From Early Design and Creation to Successful Implementation and Sustainment
J Med Internet Res 2017;19(5):e153
URL: http://www.jmir.org/2017/5/e153/
doi:10.2196/jmir.7725
PMID:28490417

©David C. Mohr, Aaron R Lyon, Emily G Lattie, Madhu Reddy, Stephen M Schueller. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

More Than a Text Message: Dismantling Digital Triggers to Curate Behavior Change in Patient-Centered Health Interventions

Frederick Muench[1*], PhD; Amit Baumel[1*], PhD(Clin Psych)

Psychiatry, Northwell Health, Great Neck, NY, United States

[*]all authors contributed equally

Corresponding Author:
Frederick Muench, PhD
Psychiatry
Northwell Health
311
1010 Northern Blvd
Great Neck, NY, 11021
United States
Phone: 1 191 532 0623
Fax: 1 516 837 1688
Email: fmuench@northwell.edu

Abstract

Digital triggers such as text messages, emails, and push alerts are designed to focus an individual on a desired goal by prompting an internal or external reaction at the appropriate time. Triggers therefore have an essential role in engaging individuals with digital interventions delivered outside of traditional health care settings, where other events in daily lives and fluctuating motivation to engage in effortful behavior exist. There is an emerging body of literature examining the use of digital triggers for short-term action and longer-term behavior change. However, little attention has been given to understanding the components of digital triggers. Using tailoring as an overarching framework, we separated digital triggers into 5 primary components: (1) who (sender), (2) how (stimulus type, delivery medium, heterogeneity), (3) when (delivered), (4) how much (frequency, intensity), and (5) what (trigger's target, trigger's structure, trigger's narrative). We highlighted key considerations when tailoring each component and the pitfalls of ignoring common mistakes, such as alert fatigue and habituation. As evidenced throughout the paper, there is a broad literature base from which to draw when tailoring triggers to curate behavior change in health interventions. More research is needed, however, to examine differences in efficacy based on component tailoring, to best use triggers to facilitate behavior change over time, and to keep individuals engaged in physical and mental health behavior change efforts. Dismantling digital triggers into their component parts and reassembling them according to the gestalt of one's change goals is the first step in this development work.

(J Med Internet Res 2017;19(5):e147) doi:10.2196/jmir.7463

KEYWORDS

alerts; digital triggers; text messaging; haptic triggers; reminder systems; push alerts; mHealth; mobile health; engagement; marketing; behavior change; behavioral medicine

Introduction

Digital interventions have pervaded how we attempt to change individuals' behavior or encourage them to engage in a health routine and have shifted the focus from patients and providers to individuals, who can now engage in self-care around the clock outside of traditional health care settings (eg, [1-3]). However, individuals' engagement with digital interventions outside of traditional settings must compete with other events in their daily lives and fluctuating motivation to engage in effortful behavior [4]. As a result, few individuals engage proactively in mobile apps or websites across the behavior change spectrum for more than a couple of weeks without therapist contact [5].

Digital Triggers Overview

Alerts, prompts, and cues are all stimuli designed to prompt desired actions and reactions from users. We use the term "trigger" to refer to this broad category of digital stimuli. Triggers are especially important when developing interventions aimed at engaging people in desired therapeutic activities in a nonclinical setting characterized by an abundance of competing

XSL•FO
RenderX

internal and external events. Designed to prompt an internal or external reaction, these stimuli give salience to an internal or external goal by focusing an individual on the desired goal at the appropriate time. For example, placing an item near the door that we will pass on our way out or a reminder from a loved one to take an umbrella as it is about to rain is a useful environmental engineering trigger that helps us to better prepare for a desired outcome (eg, to not get wet) [6]. Digital triggers include calendars, push notifications (both standard and in-app), texts, images, haptics, and other types of digital alerts, most often delivered through a mobile phone, watch, or computer.

There is growing empirical evidence supporting the use of triggers to enhance the effectiveness of health interventions or as stand-alone health interventions [7]. Evidence suggests that digital triggers improve individuals' engagement in interventions with a specific target, such as appointment adherence, medication adherence, homework completion, and engagement in medical and psychological treatments [8-12], especially when compared with no treatment control conditions [8]. For example, Spohr and colleagues [13] found that individuals increased their physical activity immediately after receiving a push notification compared with individuals who solely completed a mobile self-report questionnaire. Despite the potential of triggers to foster engagement, however, reviews have suggested that the evidence has been mixed due to the heterogeneity in populations, delivery types (eg, email, phone), characteristics of the target itself, and different outcome measures used [14-17].

There is a robust body of literature on the use of alerts, primarily short message service (SMS), otherwise known as text messaging, and email as an intervention in itself. Meta-analyses and reviews have highlighted the effectiveness of digital triggers as interventions for smoking cessation, alcohol use reduction, and prenatal care [8,11,18]. Interestingly, a recent meta-analysis revealed that adding other components (eg, a website) does not significantly improve pure alert-based interventions for health behaviors [8]. However, there have been mixed results regarding the efficacy of pure digital trigger-based interventions for weight loss and physical activity. Results suggest that trigger-based interventions that are designed to increase education about a topic (eg, proper prenatal care) or those that are designed to reduce impulsive responses or increase psychological well-being may be most effective because they increase goal salience in one's natural environment at the right time [19].

Paper Aim

Despite the promising outcomes of alert-based interventions and the prevalence of these alerts and triggers in daily life, little attention has been paid to dismantling triggers into their component parts to deepen the understanding of how they function to improve physical and mental health outcomes. While not targeting health outcomes directly, the marketing community has embraced digital triggers as core components of engagement in products and services [20]; therefore, we will draw on findings from marketing research where applicable. Our goal is to present the core components of triggers so that those developing digital health interventions can understand how to tailor those components to best engage health intervention users in the context of their environment and daily lives.

Tailoring as the Overarching Framework

Tailoring is the method of personalizing an intervention with respect to characteristics such as content and timing to ensure an intervention's highest level of receptivity and engagement. Most reviews and meta-analyses of text message-based interventions have indicated that tailoring interventions to the individual, particularly when also targeting a specific condition, yields larger effect sizes than not tailoring [8,11,12,19,21]. Although the health intervention itself is typically tailored according to components of behavior change theories, with readiness or importance of change being the most frequently used [5,12,22], nearly every component of a digital trigger (ie, content, frequency, interactivity) can be tailored to increase engagement.

As mechanisms studies have revealed, tailoring can improve engagement by increasing self-referential and heuristic processing [23,24] and attention to a desired stimulus, and reducing effortful processing. Accordingly, the more that items are tailored, the more receptive an end user may become. For example, evidence from both the health and marketing literature has shown that tailoring images to end users' demographic characteristics increases engagement [25,26] along with other strategies that match user characteristics with a persuasive strategy. Additional components, such as personalization by name, time of a trigger, end-point goal, frequency, current state, or location of receiver, can be tailored [27]. Nowhere is tailoring more evident than in the targeted advertisements on social media pages. For instance, Facebook uses 98 different data points, ranging from simple demographics (eg, gender) to life events (eg, just married) and consumer preferences (eg, preferred types of restaurants), to enable advertisers to tailor their message to specific audiences [28]. Therefore, we consider tailoring as an overall framework when developing trigger-based interventions. Below we describe each trigger component and, where possible, its relationship to other components. Figure 1 presents the concept of baseline and adaptive trigger tailoring within the context of product and engagement planning; Table 1 provides an explanatory overview of the digital trigger components described in the body of this work; and Table 2 offers a summary of our examination.

XSL•FO
RenderX

Figure 1. Trigger tailoring within the context of product and engagement planning and ongoing adaptation.

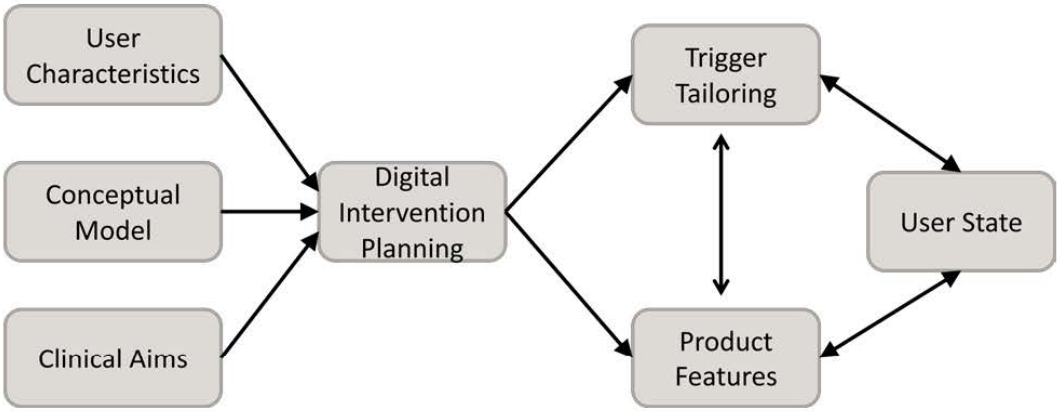


Table 1. Explanation of digital trigger components that can be tailored.

Component	Explanation	Examples
Who (sender)	The source of a digital trigger as it is recognized by the receiver	Automated, human, peer, sensor, organization, clinic, dog
How	The means by which the trigger is sent and presented	
Stimulus type	The type of trigger presentation	Text, sound, voice, image, video, moving image, light, vibration, pressure, electrical pulse
Delivery medium	The means by which the stimulus is delivered to the individual	Email, letter, push alert, text message, public alert system, digital banner (eg, social media)
When (just-in-time)	The best time to receive the trigger based on the recipient's ability to pay attention and the importance of the trigger at the moment	Fixed, customized, based on user's daily routine, context driven, data driven
How much	The frequency of triggers during the intervention phase; the context of frequency in relation to the trigger's impact	Frequency per day, week, etc
What	The actual content of a trigger; should increase the recipient's ability to *relate* to the intervention and to act in the desired way	
Trigger's target	The end-point goal the intervention developer is attempting to achieve in the user and the end-point goal of the user	Short-term (increase adherence, avoid forgetting), long-term (sustain engagement, avoid fatigue)
Trigger's structure	The arrangement of and relationships between the information parts within the medium type	Short or long message, statement, question, sound melody, vibration intensity
Developing a story through triggers	The creation of an individual journey that relates to intervention goals within the context of the user's life	

XSL•FO
RenderX

Table 2. Component summary.

Component	Highlights
Who	Messages from humans are more attended to than automated ones.
	People can attribute automated messaging to a human sender.
	The sender (or message writer) should be identified for increased credibility.
	The message source (eg, peer, loved one, clinician) should be modified based on the trigger's message (eg, reminder, motivational note, information).
How	For stimulus type, consider the end user's perceived burden based on the attention needed to process the information and the desired magnitude of reaction to the stimulus (eg, subtle changes in light vs electrical shock).
	Determine the delivery medium based on the strengths and weaknesses of each medium and their correspondence with the specific trigger's context.
	Relate to the receiver's age, communication preferences, limitations, and accessibility of the delivery method.
	Ensure trigger heterogeneity, as it will likely increase engagement.
When	Send triggers when people actually have a chance to comply with the request.
	Base triggering time on the individual's activities and not on fixed times.
	Examine the user's daily routine and preferences in order to send out triggers when the individual is most receptive.
	Enable users to easily customize trigger times.
How much	More is not necessarily better.
	The frequency of triggers should take into account the user's experience of importance and readiness for change.
What	Short-term goals refer to the immediate action item embedded within the trigger; they require or prompt immediate internal or external action.
	Long-term goals refer to the sustained engagement in an intervention to guide a user toward a long-term cognitive, emotional, or behavioral shift. In this case, the user does not necessarily need to trigger an action.
	Senders typically focus on goals that they want to achieve (eg, appointment adherence), rather than what the receiver wants to achieve with a trigger (eg, feeling better after a physical therapy session). Senders are advised to be aware of this distinction as they develop triggers.
	The more relevant triggers are to the end user without manipulation, the more effective they become.
	Including links, interactivity, or human support increases the likelihood that a trigger will be attended to in the short term.
	Triggers can be seen as the adaptive control mechanism for all components of an intervention; they focus on immediate action, but also on the larger story surrounding behavior change (ie, developing a story through triggers).

Dismantling Triggers

Who (Sender)

There is substantial literature discussing the impact of message source credibility on increasing engagement, receptivity, and actual behavior change [29]. The source of a digital trigger can be a provider, coach, peer, friend, individual, group, blog, online publication, mobile app, sensor, or anything that can be imagined by the receiver. Subtleties embedded in the message source can alter the impact and persuasiveness of an intervention, which highlights the need to carefully consider issues of framing. For example, one study revealed that visitors to a health information webpage were more likely to take action toward change when the information was on an official webpage rather than on a blog or personal page, regardless of the content [30].

Throughout the literature, there is evidence that (1) source credibility affects recipients' perception of delivered content [31]; (2) anonymous reviews are less credible than personally sourced reviews [32]; (3) messages from sources that are demographically closer to the receiver are more salient, and messages from people most important to the receiver are more likely to be viewed [33]; (4) messages are more persuasive if the sender is perceived as an expert before the message is sent [34,35]; and (5) digital interventions are enhanced when they are coupled with human support [36], possibly because of the combination of a human touch and the expectation of the individual's accountability [19].

It is worth noting that the receiver's attribution of the source might often be as powerful as the actual source of the message. Data show that people apply social rules such as politeness and reciprocity to computers [37], attribute humanistic characteristics to robots based on the latter's responsiveness [38], and develop a therapeutic alliance with self-help programs [39]. In a study conducted in a methadone clinic, participants responded to automated text messages as if it were a social interaction; they thanked the system for sending the message, despite having been informed that the messages were automated [40]. In this case, the participants' perception of the sender was more salient than the actual source. Perception can also be influenced by unknown source triggers where there is social reciprocity or perceived social acceptance, such as with friend request alerts

XSL•FO
RenderX

or other social media notifications. Taken together, it is clear that understanding and optimizing both the user's perception and the actual message source can have a substantial impact on trigger receptivity beyond the content. This is particularly true when the goal is to engage an end user in an actionable step that requires effort or burden.

How

The "how" relates to the means by which the trigger is presented and sent. Here, we divide the "how" into 2 different concepts: stimulus type, which refers to how the trigger is presented, and delivery medium, which refers to how the trigger is sent.

Stimulus Type

The stimulus type refers to the type of trigger presentation, for example, text, voice, sound, image, moving image, light, vibration, pressure, or electrical pulse. The type of stimulus used should take the specific situation into account; there is no "one type fits all" model. For example, literature on the picture-superiority effect has shown that individuals are more likely to remember images over words [41]. Furthermore, some emerging literature in the marketing field has highlighted that images increase click rates, engagement, and sharing over text alone [42] and over other types of links [43]. The primacy of image here may be attributed to the fact that images require less effort to process, can convey more information quickly, and may be remembered for longer periods of time [41]. At the same time, images may take more time to process than simple words [44], and individual differences may mediate the relationship between the trigger type and receptivity [45].

The stimulus type affects individuals' receptivity based on many factors that have not been sufficiently studied, yet should be taken into consideration: (1) the end user's perceived burden based on the time and level of attention required to process the information (eg, 160 characters vs app alert icon vs long email); (2) the complexity of the interaction (eg, simple alert via vibration vs conveying a complex message); (3) the desired magnitude of reaction to the stimulus (eg, subtle changes in light vs electrical shock); and (4) individual differences in aspects such as information processing.

Delivery Medium

The delivery medium is the means by which the stimulus is delivered to the individual, for example, text message, phone call, push notification, or digital banner in social media apps. While emails and push alerts are most often text-based communications, these delivery methods can include any stimulus type (eg, video, images, text). Mobile phones and wearable devices have expanded the opportunity to deliver an increasing number of stimulus types beyond what was imaginable only 10 years ago. In their review of wearable solutions to improve health, Zhao and colleagues [46] highlighted that current fall-detection systems send "audible alarms, vibrations, automatic voice calls, SMS, multimedia messaging service (MMS), emails, Twitter messaging, etc," highlighting the variability of the delivery medium. Not surprisingly, when compared with nothing, triggers delivered through almost any medium increase the individual's engagement in the intervention, when examined independently of moderating variables that may reduce overall engagement with the intervention (eg, trigger frequency, which may lead to alert fatigue).

Compared with the scientific literature, the marketing literature unrelated to health has made considerably more effort to understand response rates and engagement based on the delivery method of the stimulus. Overall, results have suggested that text messages are more likely to be viewed than emails, are clicked significantly faster than emails and push notifications, and have higher response rates overall [47]. Several small research studies in health have found that text messages are superior to phone calls in promoting appointment attendance, and that adding other mediums to text message-based interventions does not significantly improve outcomes [8,12]. In general, compared with other mediums, text messages appear proactively on the user's phone, are perceived as dual communication, require little programming or development experience by providers, and have the highest immediate view and response rates. At the same time, text messages require the user's phone number, are possibly the least secure form of health information delivery, can cost the end user money, and cannot convey complex information to the user.

Email can also be triggered at specific times but carries the additional benefits of being capable of transmitting complex information, cost efficient, easily shared with multiple individuals without the need for sophisticated technology, and typically stored for years by the end user. The major downside to email is that it is often not viewed by individuals upon receipt, making it unsuitable for use with actionable tasks (eg, notification of an appointment within the hour) [48].

Push notifications triggered directly from a native app fall somewhere in between text messages and email in terms of their positives and negatives. They are cheap and easy to develop, and are part of a larger platform, which allows for a rich user experience once clicked. However, push notifications have fairly low click rates, possibly because they are not perceived as part of the human communication spectrum [49] and are embedded in nearly every app on our mobile phone, possibly causing habituation.

While the data above imply the importance of choosing the right delivery method, comparative effectiveness trials have yielded insufficient evidence as to which delivery method is best at increasing engagement in interventions. Specifically, head-to-head trials directly comparing different methods have revealed little difference in health outcomes. A systematic review of 55 studies found that response rates are generally highest for text message, followed by email and then letters [12]. However, 1 of the studies revealed that participants preferred automated voice response reminders (72%) over text message reminders (28%) [50], highlighting the heterogeneity of outcomes when it comes to delivery mediums.

In light of this heterogeneity, we recommend determining the optimal trigger medium based on the medium's strengths and weaknesses in relation to the goals of the intervention. Should you add a digital banner or post on Instagram or Facebook, or send an email? The trigger's delivery medium should also vary based on other factors, such as the receiver's age (eg, email vs

XSL•FO
RenderX

text message), communication preferences (eg, Snapchat vs phone call), and limitations (eg, hearing vs visual impairment); the accessibility of the delivery method (eg, mobile phone vs email); and its development and programming capacity, among others.

Trigger Heterogeneity

Trigger heterogeneity refers to the practice of sending triggers through different mediums, at different times and places, and from different sources. This practice is based on the multichannel and source advertising literature, which has revealed that the most effective branding and advertising comes from a range of different sources. In effect, heterogeneity in the medium (eg, text, image, video), delivery method (eg, email, phone), context (eg, home, work), time (eg, same vs varied), and source (eg, provider, friend, automated) may allow for stimulus generalization and the reduction of the habituation that results from repetition. For example, social media alerts rarely indicate the actual sender of the alert but rather that "someone" has taken an action where the receiver was salient in their mind. Sometimes this is a request for a game, while other times it is an invitation to a party. Finally, the literature on learning processes has indicated that studying information in different contexts increases information retention and recall, highlighting the opportunity to increase information processing through heterogeneity [51].

When

The "when" refers to a certain point in time when the recipient is able to pay attention to a push notification and exert the desired amount of effort required to attain his or her goal. It derives from the concept "just-in-time," which focuses on the individual's receptivity within the context of ongoing tasks and daily routines [52]. People make quick decisions about how they attend to notifications as a combination of "Why am I receiving this?" and timing [53]. Although triggers can be disruptive, the recipient accepts them because they contain useful information (see "What" below). Nevertheless, adequate timing enables us to more effectively engage and support people when there is an opportunity for positive change [54,55]. Studies have repeatedly shown notifications to be effective in increasing people's immediate engagement across a broad range of health issues such as medication uptake (eg, [56-58]), attending scheduled appointments (eg, [59,60], and brief exercise [19]; these timely reminders help people to remember to engage in beneficially perceived actions during their daily routine [61]. Most trigger timing is based on self-selected alert times and general pattern awareness (eg, medication in the morning).

People's capacity to pay the desired amount of attention to a notification is influenced by the ongoing task they are engaged in when they receive the trigger. Mehrotra et al showed that perceived disruption increases as the complexity of an ongoing task increases [62]. In this study, mobile notifications were perceived as the most disruptive if they arrived when the user was in the middle of a task or just finishing it, and least disruptive if the user was idle or starting a new task. After work, when traveling, and at leisure were the tasks during which the users perceived notifications to be the most disruptive. However, knowing the target audience enabled the intervention developers to identify the right time windows for sending out triggers.

Similarly, MailChimp, an email marketing company that was sending billions of emails per month by 2014, analyzed the results of their time optimization system in terms of recipients' subscription following email distributions [63]. They came to the broad conclusion that the optimal time to send emails peaked at 10:00 AM in the recipients' own time zone between Monday and Friday (eg, work days). This finding is in line with another analysis presented by Localytics, a marketing company [64]. When categorizing the response rates for college students and bartenders, however, the optimal send time peak shifted to 1:00 PM; for neonatal nurses, who work in shifts, the results were not consistent. In effect, people tended to be more receptive to email notifications several hours after waking, but this time varied based on people's occupation. Baumel and Schueller [65] presented similar results showing that women with perinatal mood disorders preferred to receive services late at night, when clinicians were not available, due to the fact that they were awake and had pockets of availability at that time.

These empirical findings suggest that understanding the daily routines and availability of the target audience is key to planning the trigger time, because it enables the intervention developer to make sure that the trigger is received within the right context. Research also indicates that (1) relying on reminders supports repetition, but hinders habit development; (2) evidence-based cues (eg, sending data after lunch) increase automaticity; and (3) users prefer to customize and schedule alerts to support individual aspects of their daily routine [61,66] and their relevance. For example, Muench and colleagues [67] found that individuals wanted different behavior change techniques at different moments in the change process (eg, social support messages during relapse vs cognitive reframing during craving or prelapse states).

In the future, the rise of passive data collection systems via wearable, phone-based, and external sensors will provide significantly more opportunity to deliver just-in-time adaptive contextual digital triggers to end users when they need them most. While these systems will eventually allow for greater precision in triggering users at the highest levels of receptivity based on context, timing, and data from ongoing monitoring [68,69], they require significant learning. Thus, to benefit users in the short term, intervention developers should offer end users more flexibility in personally customizing their alerts. However, customization should not sacrifice best practices of when a message might be most effective. Consequently, guiding users to apply these alerts in certain time windows based on current knowledge is recommended.

How Much

"How much" refers to the frequency of triggers during the intervention phase. The question is whether sending out triggers more frequently increases users' compliance with the trigger's goal or results in trigger habituation, which eventually diminishes the reason for triggering. While "when" relates to the context of timing to maximize the trigger's impact, "how much" relates to the impact of the trigger's frequency on user behavior.

XSL•FO
RenderX

It is worth noting that the use of triggers in health interventions differs from the active utilization of a program. For example, there is extensive literature suggesting that, in Internet interventions, user utilization of the program results in better outcomes, implying a dose-response relationship [14,70-73]. However, this relationship can be mediated by the user's level of motivation and persistence [74]. Triggers are different, in that sending out more triggers to the user does not equate to the user making a greater effort to engage with the intervention. Therefore, it is important for the trigger designer to determine the number and frequency of triggers that will best encourage the user's efforts.

Determining this amount is a challenge, as data on the effect of a trigger's frequency on health outcomes are sparse. A recent review of messaging interventions for preventive health revealed that sending more versus fewer messages resulted in better outcomes; however, this review was based on few studies, making it hard to draw conclusions about the effects of trigger intensity [11]. In a study on different interventions to increase fruit and vegetable consumption, Heimendinger and colleagues did not find any effect of tailoring from 1 mailing, but found that tailoring was effective in increasing engagement when participants received 4 mailings [75]. Another study considered variation in frequency by examining whether the frequency of text messaging feedback (1 vs 3 weekly messages) could affect smoking cessation [76]. While the timing of the first weekly text message feedback (sent to both groups) was event based—sent out to individuals when they were most receptive (ie, after sending out an assessment)—the timing of the other 2 messages, sent only to the 3-weekly messages group, did not consider users' receptiveness; the messages were generally sent out on the 2 days following the initial message. No differences were identified between these 2 conditions. It is possible that the lack of "when" considerations for the subsequent 2 messages affected the outcomes, highlighting how the interdependency of components may affect a trigger's effectiveness.

While research has not directly examined the habituation to triggers in patient-centered health interventions, data from related fields suggest that an uncontrolled increase in the frequency of triggers results in a decrease in the triggers' effectiveness. Studies examining responsiveness to alerts among health providers have identified an alert fatigue: a decrease in the desired response to the alert due to its excessive appearance [77]. For example, in a literature review on physician response to drug safety alerts, van der Sijs et al found that these alerts were overridden by clinicians in 49% to 96% of cases because they receive so many irrelevant alerts [78]. When this occurs, the impact of all triggers becomes compromised. In a randomized study, Baseman et al [79] enrolled health care providers to receive public health messages and examined message content recall rates. The authors found that, for every increase of 1 local public health message per week, there was a statistically significant 41.2% decrease in the odds of the health care provider recalling the content of the study message.

Although direct empirical evidence on the impact of trigger frequency can be useful, looking more broadly into the research literature suggests that the frequency of triggers should take into account the user's experience of importance and readiness for change. For example, people who reported that it was more important to change their drinking behavior preferred to receive more messages than those with lower importance scores [80]. Since readiness for change affects the perceived importance of the health intervention (eg, [81-84]), it results in higher adherence to treatment, engagement with treatment, and response to suggestions made by clinicians across behavioral health issues (eg, [85-87]). Accordingly, the frequency of triggers should be modified based on the perceived importance of the intervention in the user's life along with other factors. Certainly, it is also possible to simply ask end users about their frequency preferences.

What

Relating the Recipient to the Intervention

The majority of research on trigger development has focused on the actual content of the message, whether it be based on a specific theoretical underpinning (eg, gain or loss framing based on the theory of reasoned action) or content-based tailoring based on multiple factors tailored to increase engagement and outcomes, as noted above [21,88]. As noted above, content tailoring is designed to increase the recipient's ability to *relate* to the trigger and to act in the desired way. As a result, researchers have often focused on the factors that will enable them to increase the recipient's *relatedness* to the message [89]. For example, Kocielnik and Hsieh demonstrated that using concepts that were cognitively close to the targeted behavior (eg, for exercising: strength training, aerobics, fitness) *or* cognitively close to the message's recipient (eg, benefits people care about or values they hold) increased recipients' perception of the informativeness and helpfulness of the triggers, and their perception of the triggers as motivators rather than just reminders. This relatedness also resulted in higher rates of completion of the desired activity [90]. While personalization has been shown to increase engagement, it is suggested that intervention developers guide and tunnel users to adhere to best practices or avoid potentially harmful triggers created through customization. For example, in a previous message preference study, 17.1% of participants chose "aggressive messages" such as "Do you seriously think that blaming others will help you change for the better?" over a neutral-toned message [91]. Because we are at the earliest stages of trigger development, using best practices from the general health behavior change literature appears to be a useful tool in personalization and customization efforts. Unfortunately, there has been significantly less research on variations in light-, haptic-, and other nontext- or nonimage-based triggers. Therefore, it is difficult to interpret how the content and structure of a haptic- or light-based trigger may affect recipients' engagement and goal achievement.

Despite message relevance and relatedness to the end user being identified as predictors of engagement and outcomes, many digital health interventions lean on behavior change theories (eg, social cognitive theory) as the core component of their content development. However, it is questionable whether theory-based content tailoring is more important to trigger-based interventions than increasing the recipient's relatedness to the trigger. For example, in a recent study, we compared different types of message framing and tailoring aimed at reducing

XSL•FO
RenderX

problem drinking. The results revealed little difference in drinking outcomes between theoretically distinct messaging groups (eg, gain vs loss framing vs static tailoring) [80]. Similar results were found in a recent meta-analysis, suggesting that theory may not be as important for building text message interventions as once thought [11].

In addition to the large body of literature on personalization of the trigger (eg, use of first name) and process-based tailoring (eg, motivation), we suggest 2 approaches to content framing less frequently articulated in the literature. These are the *target*, which refers to the short-term vs long-term reason or goal for which the trigger is being sent to the end user; and the *structure*, which includes the look, length, variation, or other within-trigger differentiators, and developing the behavior change narrative through a story.

Trigger's Target

The target of a trigger refers to the end-point goal the intervention designer is attempting to achieve in the user and the end-point goal of the user. Senders often focus on goals that they want to achieve (eg, appointment adherence), rather than what the receiver wants to achieve as a trigger (eg, feeling better after a physical therapy session). Senders are advised to be aware of this distinction as they develop triggers. The target can be further differentiated according to long-term versus short-term goals, such as adherence to an appointment versus ongoing appointment attendance, or a 1-night drinking plan versus ongoing disease management. Understanding both short-term and long-term goals allows the message content to be tailored beyond the overall condition or goal (eg, diabetes management) and placed within a specific framework of short- versus long-term goal attainment.

Short-Term Goals

Short-term goals refer to the immediate action item embedded within the trigger. They are often used in marketing campaigns and appointment reminders. For example, the goal to get someone to go to a sale is often achieved by highlighting (1) that a sale exists (ie, salience), (2) that the person needs what is for sale (ie, relevance), and (3) that the sale will end soon (ie, action urgency). In the health literature, short-term goals refer to immediate actionable items, such as taking a medication or not drinking for the next several hours. Short-term triggers are designed to help the user to approach or avoid a health behavior that is actionable in the moment through cognitive reframing or behavioral plans and guidance [92]. A short-term trigger usually involves some actionable reciprocity, interaction, or potential reward that will engage the user in the moment. Including links, interactivity, or human support has been found to increase the likelihood that a trigger will be attended to in the short term [93-95].

Long-Term Goals

Long-term goals refer to users' sustained engagement in an intervention to guide them toward a long-term cognitive, emotional, or behavioral shift. This is often the primary goal of chronic disease health campaigns. For example, general information on the long-term benefits of reducing alcohol intake on diabetes severity acts as a trigger to shift cognition in subtle ways. Fostering acceptance about an incurable illness is also a subtle long-term intervention target. While the intervention may result in short-term change, it does not require or prompt immediate internal or external action. Unlike with short-term targets, triggers directed toward long-term goals do not necessarily need to trigger an action on the part of the user but rather keep the end point salient over time. A recent review suggested that long-term interventions of about 6-12 months were more effective in promoting behavior change than those that were less than 6 months [11]. While little can be gleaned from these findings, they do suggest that trigger-based interventions can focus on long-term behavior change. For the most part, there are fewer examples of specific trigger types that are designed to foster long-term goal salience. While not specifically necessarily related to trigger-based interventions, an early study on virtual agents revealed that direct communication increased short-term engagement, whereas polite communication fostered long-term adherence [96,97].

Goal Relevance

The more that triggers are relevant to the end user without manipulation, that is, the end user agrees with the goal embedded within them, the more effective they become. Mehrotra and colleagues [62] highlighted that one of the top reasons triggers are opened is that they are relevant to the user, regardless of the senders' goals. Indeed, although they are not mutually exclusive, receivers' goals often have little overlap with the senders' goals. A good example of this distinction lies in medication adherence push notifications. People are not taking medications for the sake of taking medications; they are taking medications to avoid health consequences or to feel better. Therefore, if people can avoid health consequences and feel better without medication, medication triggers become irrelevant to them. This includes the burden associated with reminding oneself to take medications, which may trigger a recurrent feeling of being ill. In light of these potential tensions, understanding the framework through which users receive the trigger and using that to create relatedness is possibly one of the most important, yet underappreciated, components of intervention development.

Trigger's Structure

The structure of a trigger refers to the arrangement of and relationships between the information parts within the medium type. For example, with a text message, the structure can vary in many ways, such as long versus short messages, emoticons versus no emoticons, questions versus statements, and so on. The structure of a vibration stimulus can vary in terms of the intensity or frequency of the individual trigger. Images can vary in terms of brightness, contrast, and color. Does a sound alert employ a specific melody? Does the vibration follow a specific pattern for different message types (eg, mother vs friend) as in smartwatch alerts? These questions highlight the complex, multidimensional nature of trigger structures.

Unfortunately, little research has compared the effectiveness of different trigger structures. We recently compared users' preferences for different short message structures, revealing that people have a range of preferences with regard to structure. For example, we found that over 90% of the sample preferred

XSL•FO
RenderX

smiley emoticons, no "textese," and proper grammar, all of which have nothing to do with the actual content of the message [91]. We also identified some moderators of structure preference. For example, compared with participants with a college degree or higher, participants with less than a college degree were more likely to prefer short messages to long messages and messages that included smiley emoticons to messages that contained no emoticons.

The structure of a trigger often varies based on system rules that are preprogrammed and learned by an individual, or based on evolutionary rules. For example, in hospital settings, triggers for specific actions often take the form of varying patterns of sounds delivered through a public alert system or vibrations in a phone or pager. The importance of alerts can also be programmed into mobile phones in a way that only the user knows the meaning of the alert. In this way, often the individual learns the trigger's structure through slow pattern recognition. At the same time, a loud alarm alerts nearly all individuals to action without the need for overt learning and thus is often used in emergency situations. More research needs to be done to investigate how various alert structures prompt action.

Developing a Story Through Triggers

Triggers can be seen as actuators of an intervention or built as a component within a larger intervention framework. An actuator controls the movement or flow of system components. We think of triggers as the adaptive control mechanism for all components of an intervention that focus not only on immediate action, but also on the larger story surrounding behavior change in the context of a person's life and changing goals. Initial baseline tailoring efforts can help build a story around salient components of an individual's life (eg, social network), patterns of behavior (eg, usual heavy drinking times), and progress toward change in symptoms. In this way, triggers are framed to meet both long- and short-term needs and goals. A recent study on using messaging to reduce problem drinking revealed that the largest effect sizes for drinking reduction and subjective goal attainment were found among those whose messages were adapted weekly based on whether or not they met their goals for the week [80]. The adaptation of triggers in line with the users' experiences and goals can be applied in any context to curate tailored components immediately to meet the current demands of the individual and the entire trajectory of an intervention within the context of individuals' lives.

Limitations

Digital triggers offer unprecedented opportunities to increase goal salience without requiring significant effort on the part of the end user. However, as noted throughout the paper, implementation efforts can backfire without understanding of the potential pitfalls of each component. Alert fatigue is a major concern among emergency and medical groups and can occur in health-based interventions as well. Subgroups of individuals in studies have reported that messages can be burdensome, too frequent, and received at inopportune times, among other complaints. Habituation or ignoring triggers is another concern that can occur with long-term trigger-based interventions. Such user responses can reduce trigger salience. In addition to these concerns over intervention effectiveness, intervention developers need to take privacy and security concerns into account. For example, text messaging is possibly the least secure method of trigger delivery, as it is stored on the phone company's server, the messaging provider's server, and the end user's phone. Users should be aware of all privacy concerns prior to engaging in trigger-based adjuncts or interventions, like they would for any digital intervention.

In addition, as mentioned above in the Paper Aim paragraph, this review included papers from the marketing literature, where sometimes the goals of the intervention (eg, email subscription, special-sell) are very much different from the goals embedded within health interventions (lifetime changes needed to cope with a chronic illness). We accounted for this difference, however, by interpreting the findings from the marketing literature only within the particular context of the reviewed component (eg, when). In this way, the marketing literature could be perceived as an exploratory laboratory of specific human behaviors and reactions to triggers in the context of daily living (eg, when people are most available), which enabled us to generalize these data to the health domain.

Finally, due to the high variation in research and development methods of digital interventions that embed triggers (sometimes without clear information about component tailoring), and the broad disciplines from which data can be retrieved, we were not able to apply a systematic process to review and retrieve data (eg, systematic review). As research moves forward in this field, more evidence about the impact of trigger tailoring will emerge, and such a review would become more feasible.

Conclusions

Fostering ongoing goal salience through adaptive tailored triggers can enhance our models of behavioral health interventions. Triggers frame the interaction with the end user as the actuator that drives the story both in the short term and over time. Understanding the nature of digital triggers and how different trigger components facilitate action is of primary importance in realizing the trigger's potential. As evidenced throughout the paper, there is a broad literature base from which to draw when building trigger-based interventions. We recommend examining the existing literature in detail, with a special emphasis on the trigger developmental work that has primarily been done in the text messaging space [67,98]. The primary gaps in the literature appear in (1) long-term engagement appeals and (2) methods to build trigger interventions that can adapt to individuals' current state within the larger framework of their behavior change process. Dismantling digital triggers into their component parts and reassembling them according to the gestalt of one's change goals is the first step in this development work.

XSL•FO
RenderX

Acknowledgments

FM and AB equally contributed to this paper. This paper was partially supported by a grant from the National Institute on Alcohol Abuse and Alcoholism (R34AA021502). The funders had no role in the decision to publish or preparation of the manuscript.

Conflicts of Interest

FM has equity in a mobile health company for health behavior change and consults with several other companies using mobile technology to facilitate behavior change.

References

1. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
2. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
3. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015 Aug 19:1-12. [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
4. Baumeister RF, Vohs KD. Self-regulation, ego depletion, and motivation. Social Pers Psychol Compass 2007 Nov;1(1):115-128. [doi: 10.1111/j.1751-9004.2007.00001.x]
5. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
6. Papies EK. Health goal priming as a situated intervention tool: how to benefit from nonconscious motivational routes to health behaviour. Health Psychol Rev 2016 Dec;10(4):408-424 [FREE Full text] [doi: 10.1080/17437199.2016.1183506] [Medline: 27144729]
7. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
8. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
9. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
10. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]
11. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
12. De Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]
13. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
14. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
15. 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]
16. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
17. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
18. 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]
19. Versluis A, Verkuil B, Spinhoven P, van der Ploeg MM, Brosschot JF. Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. J Med Internet Res 2016;18(6):e152 [FREE Full text] [doi: 10.2196/jmir.5642] [Medline: 27349305]
20. Rettie R, Grandcolas U, Deakins B. Text message advertising: response rates and branding effects. J Targeting Meas Analysis Marketing 2005;13(4):304-312.

XSL•FO
RenderX

21. 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]
22. Lustria MLA, 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] [Medline: 23750972]
23. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
24. Chua HF, Liberzon I, Welsh RC, Strecher VJ. Neural correlates of message tailoring and self-relatedness in smoking cessation programming. Biol Psychiatry 2009 Jan 15;65(2):165-168 [FREE Full text] [doi: 10.1016/j.biopsych.2008.08.030] [Medline: 18926523]
25. Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? Am J Public Health 1994 Jan;84(1):43-49. [Medline: 8279610]
26. Kreuter MW, Sugg-Skinner C, Holt CL, Clark EM, Haire-Joshu D, Fu Q, et al. Cultural tailoring for mammography and fruit and vegetable intake among low-income African-American women in urban public health centers. Prev Med 2005 Jul;41(1):53-62. [doi: 10.1016/j.ypmed.2004.10.013] [Medline: 15916993]
27. Galazzo L. 5 push notifications strategies to increase app engagement. San Francisco, CA: Leanplum; 2016 May 27. URL: https://www.leanplum.com/blog/push-notifications-strategies-increase-engagement/ [accessed 2017-01-22] [WebCite Cache ID 6nh4rO4Fp]
28. Dewey C. 98 personal data points that Facebook uses to target ads to you. Washington, DC: The Washington Post; 2016 Aug 19. URL: https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/?utm_term=.6fbc6c8d0f0c [accessed 2017-01-22] [WebCite Cache ID 6nh4hMEvM]
29. Jones L, Sinclair R, Courneya K. The effects of source credibility and message framing on exercise intentions, behaviors, and attitudes: an integration of the elaboration likelihood model and prospect theory. J Appl Soc Psychol 2003;33(1):179-196. [doi: 10.1111/j.1559-1816.2003.tb02078.x]
30. Hu Y, Sundar SS. Effects of online health sources on credibility and behavioral intentions. Commun Res 2010;37(1):105-132.
31. 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]
32. Xie H, Miao L, Kuo P, Lee B. Consumers' responses to ambivalent online hotel reviews: the role of perceived source credibility and pre-decisional disposition. Int J Hospitality Manage 2011 Mar;30(1):178-183. [doi: 10.1016/j.ijhm.2010.04.008]
33. Wilson EJ, Sherrell DL. Source effects in communication and persuasion research: a meta-analysis of effect size. J Acad Marketing Sci 1993 Mar;21(2):101-112. [doi: 10.1007/BF02894421]
34. O'keefe DJ. The persuasive effects of delaying identification of high - and low - credibility communicators: A meta - analytic review. Commun Stud 1987 Jun;38(2):63-72. [doi: 10.1080/10510978709368231]
35. Ward C, McGinnies E. Persuasive effects of early and late mention of credible and noncredible sources. J Psychol 1974;86(1):17-23.
36. Schueller S, Tomasino K, Lattie E, Mohr D. Human support for behavioral intervention technologies for mental health: the efficiency model. 2016 Presented at: ACM CHI2016; May 7-12; San Jose, CA, USA.
37. Nass C, Moon Y. Machines and mindlessness: social responses to computers. J Soc Isssues 2000 Jan;56(1):81-103. [doi: 10.1111/0022-4537.00153]
38. Hoffman G, Birnbaum GE, Vanunu K, Sass O, Reis HT. Robot responsiveness to human disclosure affects social impression and appeal. 2014 Presented at: ACM/IEEE International Conference on Human-Robot Interaction; Mar 3-6, 2014; Bielefeld, Germany.
39. Clarke J, Proudfoot J, Whitton A, Birch M, Boyd M, Parker G, et al. Therapeutic alliance with a fully automated mobile phone and web-based intervention: secondary analysis of a randomized controlled trial. JMIR Ment Health 2016 Feb 25;3(1):e10 [FREE Full text] [doi: 10.2196/mental.4656] [Medline: 26917096]
40. Muench F, Adams M, McKay J, Morgenstern J, van Stolk-Cooke K. Integration of text messaging for adherence to behavioral health appointments in methadone treatment. 2011 Presented at: Annual Meeting of the Association for Behavioral Cognitive Therapies; Nov 10-13, 2011; Toronto, ON, Canada.
41. Childers TL, Houston MJ. Conditions for a picture-superiority effect on consumer memory. J Consumer Res 1984 Sep;11(2):643-654. [doi: 10.1086/209001]
42. Geerlings D. Twitter engagement study: photo vs. text tweets. Boston, MA: Brand Networks LLC; 2014 Jan 17. URL: https://bn.co/twitter-engagement-study-photo-vs-text-tweets/ [accessed 2017-01-22] [WebCite Cache ID 6nh51vr0i]
43. Redsicker P. Social photos generate more engagement: new research. Poway, CA: Social Media Examiner; 2014 May 13. URL: http://www.socialmediaexaminer.com/photos-generate-engagement-research/ [accessed 2017-01-22] [WebCite Cache ID 6nh5Br9vB]

XSL•FO
RenderX

44. Snodgrass JG, McCullough B. The role of visual similarity in picture categorization. J Exp Psychol Learn Mem Cogn 1986 Jan;12(1):147-154. [Medline: 2949047]
45. Fleischhauer M, Enge S, Brocke B, Ullrich J, Strobel A, Strobel A. Same or different? Clarifying the relationship of need for cognition to personality and intelligence. Pers Soc Psychol Bull 2010 Jan;36(1):82-96. [doi: 10.1177/0146167209351886] [Medline: 19901274]
46. Zhao F, Li M, Tsien J. The emerging wearable solutions in mHealth. In: Bonney W, editor. Mobile Health Technologies: Theories and Applications. Rijeka, Croatia: InTech; 2016.
47. Peccolo G. 45 texting statistics that prove businesses need to take SMS seriously.: OneReach Blog; 2015 Sep 10. URL: https://onereach.com/blog/45-texting-statistics-that-prove-businesses-need-to-start-taking-sms-seriously/ [accessed 2017-01-21] [WebCite Cache ID 6ngqyOKMG]
48. Tolentino J. SMS vs. push notification vs email: when should your app use what?. Amsterdam, Netherlands: The Next Web BV; 2015. URL: https://thenextweb.com/future-of-communications/2015/02/09/sms-vs-push-vs-email/ [accessed 2017-01-30] [WebCite Cache ID 6nuIcGOUF]
49. Essany M. SMS marketing wallops email with 98% open rate and only 1% spam. Traverse City, MI: Good News, Inc; 2014 Aug 6. URL: https://mobilemarketingwatch.com/sms-marketing-wallops-email-with-98-open-rate-and-only-1-spam-43866/ [accessed 2017-01-22] [WebCite Cache ID 6nh5Mc36y]
50. Greaney ML, Puleo E, Sprunck-Harrild K, Bennett GG, Cunningham MA, Gillman MW, et al. Electronic reminders for cancer prevention: factors associated with preference for automated voice reminders or text messages. Prev Med 2012 Aug;55(2):151-154 [FREE Full text] [doi: 10.1016/j.ypmed.2012.05.014] [Medline: 22659227]
51. Bouton ME, Nelson JB, Rosas JM. Stimulus generalization, context change, and forgetting. Psychol Bull 1999 Mar;125(2):171-186. [Medline: 10087934]
52. Nahum-Shani I, Smith S, Tewari A, Witkiewitz K, Collins L, Spring B, et al. Just in time adaptive interventions (jitais): an organizing framework for ongoing health behavior support. Technical Report Number 14-126. University Park, PA: The Methodology Center, The Pennsylvania State University; 2014. URL: https://methodology.psu.edu/media/techreports/14-126.pdf [accessed 2017-04-12] [WebCite Cache ID 6pg4NbNoG]
53. Canel C, Rosen D, Anderson E. Just-in-time is not just for manufacturing: a service perspective. Ind Manage Data Syst 2000;100(2):51-60.
54. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253. [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
55. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
56. Bennett JW, Glasziou PP. Computerised reminders and feedback in medication management: a systematic review of randomised controlled trials. Med J Aust 2003 Mar 03;178(5):217-222. [Medline: 12603185]
57. Patel S, Jacobus-Kantor L, Marshall L, Ritchie C, Kaplinski M, Khurana PS, et al. Mobilizing your medications: an automated medication reminder application for mobile phones and hypertension medication adherence in a high-risk urban population. J Diabetes Sci Technol 2013 May 01;7(3):630-639 [FREE Full text] [Medline: 23759395]
58. Keränen T, Liikkanen S. Medication reminder service for mobile phones: an open feasibility study in patients with Parkinson's disease. Telemed J E Health 2013 Nov;19(11):888-890. [doi: 10.1089/tmj.2013.0014] [Medline: 24050614]
59. Downer SR, Meara JG, Da Costa AC, Sethuraman K. SMS text messaging improves outpatient attendance. Aust Health Rev 2006 Aug;30(3):389-396. [Medline: 16879098]
60. Irigoyen MM, Findley S, Earle B, Stambaugh K, Vaughan R. Impact of appointment reminders on vaccination coverage at an urban clinic. Pediatrics 2000 Oct;106(4 Suppl):919-923. [Medline: 11044144]
61. Stawarz K, Cox A, Blandford A. Don't forget your pill!: designing effective medication reminder apps that support users daily routines. 2014 Presented at: 32nd Annual ACM Conference on Human Factors in Computing Systems; Apr 26-May 1, 2014; Toronto, ON, Canada.
62. Mehrotra A, Pejovic V, Vermeulen J, Hendley R, Musolesi M. My phoneme: understanding people's receptivity to mobile notifications. 2016 Presented at: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA, USA p. 7-12.
63. Foreman J. Insights from MailChimp's send time optimization system. Atlanta, GA: The Rocket Science Group, LLC; 2014 Jul 14. URL: http://blog.mailchimp.com/insights-from-mailchimps-send-time-optimization-system/ [accessed 2017-01-22] [WebCite Cache ID 6nh5UnVSU]
64. Todd J. Marketing Land. Redding, CT: Third Door Media, Inc; 2016 Apr 20. 5 must-do push notification best practices for mobile marketers URL: http://marketingland.com/5-must-push-notification-best-practices-mobile-marketers-172981 [accessed 2017-05-01] [WebCite Cache ID 6q88KEAoM]
65. Baumel A, Schueller SM. Adjusting an available online peer support platform in a program to supplement the treatment of perinatal depression and anxiety. JMIR Ment Health 2016 Mar 21;3(1):e11 [FREE Full text] [doi: 10.2196/mental.5335] [Medline: 27001373]

XSL•FO
RenderX

66. Stawarz K, Cox A, Blandford A. Beyond self-tracking and reminders: designing smartphone apps that support habit formation. 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, Republic of Korea.
67. Muench F, Weiss RA, Kuerbis A, Morgenstern J. Developing a theory driven text messaging intervention for addiction care with user driven content. Psychol Addict Behav 2013 Mar;27(1):315-321 [FREE Full text] [doi: 10.1037/a0029963] [Medline: 22963375]
68. Murray T, Hekler E, Spruijt-Metz D, Rivera D, Raij A. Formalization of computational human behavior models for contextual persuasive technology. 2016 Presented at: 11th International Conference on Persuasive Technology; April 5-7, 2016; Salzburg, Austria.
69. 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]
70. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
71. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of internet website use with weight change in a long-term weight loss maintenance program. J Med Internet Res 2010;12(3):e29 [FREE Full text] [doi: 10.2196/jmir.1504] [Medline: 20663751]
72. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
73. Zbikowski SM, Jack LM, McClure JB, Deprey M, Javitz HS, McAfee TA, et al. Utilization of services in a randomized trial testing phone- and web-based interventions for smoking cessation. Nicotine Tob Res 2011 May;13(5):319-327 [FREE Full text] [doi: 10.1093/ntr/ntq257] [Medline: 21330267]
74. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
75. Heimendinger J, O'Neill C, Marcus AC, Wolfe P, Julesburg K, Morra M, et al. Multiple tailored messages are effective in increasing fruit and vegetable consumption among callers to the Cancer Information Service. J Health Commun 2005;10 Suppl 1:65-82. [doi: 10.1080/10810730500263646] [Medline: 16377601]
76. Haug S, Meyer C, Schorr G, Bauer S, John U. Continuous individual support of smoking cessation using text messaging: a pilot experimental study. Nicotine Tob Res 2009 Aug;11(8):915-923. [doi: 10.1093/ntr/ntp084] [Medline: 19542517]
77. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. Health Aff (Millwood) 2011 Dec;30(12):2310-2317 [FREE Full text] [doi: 10.1377/hlthaff.2010.1111] [Medline: 22147858]
78. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
79. Baseman JG, Revere D, Painter I, Toyoji M, Thiede H, Duchin J. Public health communications and alert fatigue. BMC Health Serv Res 2013 Aug 05;13:295 [FREE Full text] [doi: 10.1186/1472-6963-13-295] [Medline: 23915324]
80. Muench F, van Stolk-Cooke K, Kuerbis A, Stadler G, Baumel A, Shao S, et al. A randomized controlled pilot trial of different mobile messaging interventions for problem drinking compared to weekly drink tracking. PLoS One 2017;12(2):e0167900 [FREE Full text] [doi: 10.1371/journal.pone.0167900] [Medline: 28146560]
81. Donovan D, Rosengren D. Motivation for Behavior Change and Treatment Among Substance Abusers. New York, NY: Guilford Press; 1999.
82. Laforge RG, Velicer WF, Richmond RL, Owen N. Stage distributions for five health behaviors in the United States and Australia. Prev Med 1999 Jan;28(1):61-74. [doi: 10.1006/pmed.1998.0384] [Medline: 9973589]
83. Marcus BH, Rakowski W, Rossi JS. Assessing motivational readiness and decision making for exercise. Health Psychol 1992;11(4):257-261. [Medline: 1396494]
84. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589]
85. Daley A, Duda J. Self-determination, stage of readiness to change for exercise, and frequency of physical activity in young people. Eur J Sport Sci 2006;6(4):231-243.
86. Kerns RD, Habib S. A critical review of the pain readiness to change model. J Pain 2004 Sep;5(7):357-367. [doi: 10.1016/j.jpain.2004.06.005] [Medline: 15501193]
87. Peterson KA, Hughes M. Readiness to change and clinical success in a diabetes educational program. J Am Board Fam Pract 2002;15(4):266-271 [FREE Full text] [Medline: 12150458]
88. Revere D, Dunbar PJ. Review of computer-generated outpatient health behavior interventions: clinical encounters “in absentia”. J Am Med Inform Assoc 2001;8(1):62-79 [FREE Full text] [Medline: 11141513]
89. Schmid KL, Rivers SE, Latimer AE, Salovey P. Targeting or tailoring? Maximizing resources to create effective health communications. Mark Health Serv 2008;28(1):32-37 [FREE Full text] [Medline: 18389854]

XSL•FO
RenderX

90. Kocielnik R, Hsieh G. Send me a different message: utilizing cognitive space to create engaging message triggers. 2017 Presented at: 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing; Feb 25-Mar 1, 2017; Portland, OR, USA.
91. Muench F, van Stolk-Cooke K, Morgenstern J, Kuerbis AN, Markle K. Understanding messaging preferences to inform development of mobile goal-directed behavioral interventions. J Med Internet Res 2014 Feb;16(2):e14 [FREE Full text] [doi: 10.2196/jmir.2945] [Medline: 24500775]
92. Kendzor DE, Shuval K, Gabriel KP, Businelle MS, Ma P, High RR, et al. Impact of a mobile phone intervention to reduce sedentary behavior in a community sample of adults: a quasi-experimental evaluation. J Med Internet Res 2016;18(1):e19 [FREE Full text] [doi: 10.2196/jmir.5137] [Medline: 26810027]
93. Wise K, Hamman B, Thorson K. Moderation, response rate, and message interactivity: features of online communities and their effects on intent to participate. J Comput Mediat Commun 2006 Oct;12(1):24-41. [doi: 10.1111/j.1083-6101.2006.00313.x]
94. Sundar S, Bellur S, Oh J, Jia H, Kim HS. Theoretical importance of contingency in human-computer interaction: effects of message interactivity on user engagement. Commun Res 2016;43(5):595-625.
95. Oh J, Sundar S. User engagement with interactive media: a communication perspective. In: O'Brien H, Cairns P, editors. Why Engagement Matters. Berlin, Germany: Springer; 2016:177-198.
96. Bickmore T, Schulman D, Yin L. Maintaining engagement in long-term interventions with relational agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
97. Bickmore T, Mauer D, Crespo F, Brown T. Persuasion, task interruption and health regimen adherence. In: Kort Y, IJsselsteijn W, editors. Persuasive Technology. Lecture Notes in Computer Science. Berlin, Germany: Springer; 2007:4744.
98. Bock BC, Rosen RK, Barnett NP, Thind H, Walaska K, Foster R, et al. Translating behavioral interventions onto mHealth platforms: developing text message interventions for smoking and alcohol. JMIR Mhealth Uhealth 2015;3(1):e22 [FREE Full text] [doi: 10.2196/mhealth.3779] [Medline: 25714907]

Abbreviations

MMS: multimedia messaging service
SMS: short message service

Edited by G Eysenbach; submitted 07.02.17; peer-reviewed by S Garcia, Y Bartlett; comments to author 28.02.17; revised version received 06.03.17; accepted 17.03.17; published 26.05.17.

Please cite as:
Muench F, Baumel A
More Than a Text Message: Dismantling Digital Triggers to Curate Behavior Change in Patient-Centered Health Interventions
J Med Internet Res 2017;19(5):e147
URL: http://www.jmir.org/2017/5/e147/
doi:10.2196/jmir.7463
PMID:28550001

©Frederick Muench, Amit Baumel. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 26.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Visits and Patient-Centered Care: Results of a Patient Survey and Observational Study

Kimberlyn Marie McGrail[1*], PhD; Megan Alyssa Ahuja[1*], MPH; Chad Andrew Leaver[2*], MSc, MBA

[1]Centre for Health Services and Policy Research, Vancouver, BC, Canada

[2]Canada Health Infoway, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Kimberlyn Marie McGrail, PhD
Centre for Health Services and Policy Research
201-2206 East Mall
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 778 998 3821
Fax: 1 604 822 5690
Email: kim.mcgrail@ubc.ca

Abstract

Background: Virtual visits are clinical interactions in health care that do not involve the patient and provider being in the same room at the same time. The use of virtual visits is growing rapidly in health care. Some health systems are integrating virtual visits into primary care as a complement to existing modes of care, in part reflecting a growing focus on patient-centered care. There is, however, limited empirical evidence about how patients view this new form of care and how it affects overall health system use.

Objective: Descriptive objectives were to assess users and providers of virtual visits, including the reasons patients give for use. The analytic objective was to assess empirically the influence of virtual visits on overall primary care use and costs, including whether virtual care is with a known or a new primary care physician.

Methods: The study took place in British Columbia, Canada, where virtual visits have been publicly funded since October 2012. A survey of patients who used virtual visits and an observational study of users and nonusers of virtual visits were conducted. Comparison groups included two groups: (1) all other BC residents, and (2) a group matched (3:1) to the cohort. The first virtual visit was used as the intervention and the main outcome measures were total primary care visits and costs.

Results: During 2013-2014, there were 7286 virtual visit encounters, involving 5441 patients and 144 physicians. Younger patients and physicians were more likely to use and provide virtual visits (P<.001), with no differences by sex. Older and sicker patients were more likely to see a known provider, whereas the lowest socioeconomic groups were the least likely (P<.001). The survey of 399 virtual visit patients indicated that virtual visits were liked by patients, with 372 (93.2%) of respondents saying their virtual visit was of high quality and 364 (91.2%) reporting their virtual visit was “very” or “somewhat” helpful to resolve their health issue. Segmented regression analysis and the corresponding regression parameter estimates suggested virtual visits appear to have the potential to decrease primary care costs by approximately Can $4 per quarter (Can –$3.79, P=.12), but that benefit is most associated with seeing a known provider (Can –$8.68, P<.001).

Conclusions: Virtual visits may be one means of making the health system more patient-centered, but careful attention needs to be paid to how these services are integrated into existing health care delivery systems.

(J Med Internet Res 2017;19(5):e177) doi:10.2196/jmir.7374

KEYWORDS

virtual visits; telehealth; primary care delivery; patient-centered care

XSL•FO
RenderX

Introduction

A greater orientation to information technology in primary care opens new possibilities for health care delivery, one of which is virtual care. "Virtual care" is a broad term meant to capture all clinical interactions in health care that do not involve the patient and provider being in the same room at the same time [1]. Consultations may be asynchronous, whereby patients answer structured clinical questions online and then receive care from a physician at a later time ("e-visits"), or synchronous, whereby patients interact with physicians in real time via telephone ("teleconsultations"), videoconference ("virtual visits") [2-5], or even by text [6].

Canada is a geographically large country, and physicians are disproportionately clustered within urban and semiurban settings, with known shortages of primary care providers in more rural and remote areas [7]. However, limited access to primary care is not a feature of rural areas only. There are well-documented accessibility issues in urban areas as well, whereby many patients do not have a regular primary care physician or cannot access their physician for in-person appointments within a timeframe that meets their current needs [8-10]. New modes of contact, including virtual visits, are one potential way to solve at least some of these access issues.

Some health systems are integrating virtual care into primary care practices as a complement to existing modes of care. The US health care provider Kaiser Permanente has altered their delivery of health care services to improve information continuity and to provide easy access to appropriate care for patients through initiatives such as electronic messaging with the care team, scheduled telephone visits, and a comprehensive patient portal [11]. This is, at least in part, a response to a growing focus on patient-centered care that aims to provide care that is both "accessible" and "acceptable" [12,13]. Virtual care can, for example, increase access to care for individuals who have difficulty presenting in-person for primary care services, such as those living in long-term care facilities and/or those with mobility issues [14].

Virtual care is also understood as a means of making care more convenient to patients and is gaining momentum in the United States and other advanced health systems alongside other innovations such as retail clinics and other forms of walk-in care [15-20]. There is, at present, limited empirical evidence about how patients view virtual care and more specifically virtual visits, how such care affects overall primary care use, and whether integration of virtual visits in existing relationships or more as a "walk-in" service matters. This research aims to fill that gap, focusing on British Columbia, Canada, a province with a population of 4.5 million people, which has had public funding of virtual visits since 2012.

This study has both descriptive and analytic objectives. The descriptive objectives are to assess the users and providers of virtual visits, as well as the reasons patients give for their use. The analytic objective is to assess empirically the influence of the use of virtual visits on overall primary care use and costs, paying specific attention to whether virtual care is provided by a known or a new primary care physician.

Methods

This was a mixed-methods study that included a patient survey and an analysis of administrative health care data. The study was undertaken in British Columbia, Canada, a province with universal health care coverage for a population of approximately 4.5 million. IRB-Services Canada provided research ethics approval for the patient survey, and the University of British Colombia Behavioural Research Ethics Board for the observational study.

Observational Study

Data Sources

We used data for fiscal years 2010/2011 to 2013/2014. The following files were linked:

Medical Services Plan (MSP) Payment Information File: this file includes all fee-for-service payment data for physicians and contains the information needed to identify virtual visit encounters between physicians and patients [21];

Consolidation file (MSP Registration and Premium Billing): this file includes patient characteristics for all BC residents who are eligible and receive publicly funded health care services, including patient demographic, location, and socioeconomic status (SES) information based on neighborhood income [22];

PharmaNet: this file includes all prescriptions filled in British Columbia, regardless of payment source. This file was used in a measure to classify physician practice style and to examine the frequency of prescribing as part of virtual visits [23]; and

MSP Practitioner Information File: this file includes demographic, specialty, location of training, and location of practice of all primary care providers in British Columbia [24].

We used the Johns Hopkins' adjusted clinical group (ACG) case-mix system to assess patients' morbidity burden. This system uses a risk-adjustment methodology to describe and predict expected use and cost of health care services and has been validated for use with BC administrative data [25]. More specifically, we used aggregated diagnosis groups (ADGs), a midlevel grouping based on diagnosis codes in health care encounters, to assess comorbidities. There are 32 ADGs in the overall system, eight of which are considered major; we used a count of major ADGs as our measure of health status [26].

Defining the Cohort and Comparison Group

The cohort was everyone who had one or more virtual visit fee codes billed on their behalf at any point in the study period. The population comparison group used for initial descriptive analyses were all other BC residents.

A second (and main) comparison group was matched (3:1) on 5-year age group, sex, health service delivery area (HSDA) of residence (there are 16 HSDAs in British Columbia), and number of major ADGs. This comparison group enabled analysis of the effects of virtual visits on overall patterns of health care services use. We anchored the match so that the comparison group had a primary care encounter at approximately the same time as the case's virtual visit. This was done because a virtual

XSL•FO
RenderX

visit will naturally create a spike in use because it represents an individual's decision to access the health care system. Matching based on that access creates a similar spike in the comparison group and thus increases the likelihood that any differences preceding or following that spike were attributable to the virtual visit.

Other Variables and Analysis

Our outcome of interest was costs associated with primary care visits. We used broader physician data to assess referrals to specialists and for laboratory and imaging services, but these aspects of care were not included in the cost outcome because, in the BC context, primary care physicians do not typically provide in-office laboratory or imaging services. Descriptive analyses assessed the age, sex, location, and morbidity distribution of individuals who had virtual visits, and the specialty, age, and sex of primary care physicians who provided them. We further separated users into those who saw a provider with whom they have had a traditional office visit versus seeing a new provider. Referrals and prescriptions within 30 days of a visit (except 90 days for imaging to allow for longer wait times) were captured using physician and pharmaceutical data.

Primary care physician practice style was classified into three groups—high responsibility, mixed practice, and low responsibility—based on a cluster analysis using five variables derived from fee-for-service payment data. This approach was developed previously [27] and was shown to create distinct groups, with high-responsibility physicians providing more comprehensive or full-service care and low-responsibility physicians providing care more consistent with a walk-in clinic-style practice.

Time series analysis was used to assess the effect of virtual visits on overall visits with primary care physicians. A time series approach helps identify changes in both the trend (slope) in service use and amount (level) of service use before and after an initial virtual visit encounter. No change in amount of service will be interpreted as virtual visits serving a substitute function for other health care services use.

More specifically, we will analyze time series data using segmented regression [28] in the form:

$$Y_{jkt}=\beta_0+\beta_1\times time_t+\beta_2\times group_k+\beta_3\times group_k\times time_t+\beta_4\times level_{jt}+\beta_5\times trend_{jt}+\beta_6\times level_{jt}\times group_k+\beta_7\times trend_{jt}\times group_k+\varepsilon_{jkt}$$

where *Y* is the mean number of primary care visits / costs per month, *j* is the intervention status, *k* is the group, and *t* represents time. In these models, β_0 is the intercept and β_1 is the existing trend in the matched comparison group, β_2 estimates the preintervention difference in level between the intervention group and its matched controls, and β_3 estimates the difference in trend, and β_4 and β_5 estimate the changes in level and trend for matched controls postintervention, respectively. The real parameters of interest are β_6 and β_7 because they estimate the difference in level (β_6) and trend (β_7) between the intervention group and matched controls following the intervention. Statistically significant values for these latter two parameters would indicate that the use of telemedicine had an effect on primary care services use. Analyses were completed using the autoreg procedure in SAS 9.0 and were assessed for autocorrelation.

Patient Survey

Recruitment and Analysis

Inclusion criteria for the patient survey were 18 years of age or older, had at least one virtual visit with a primary care physician in the past 6 months, currently living in British Columbia, and able to complete the survey in English. A total of 3025 patients were deemed eligible. The virtual visit technology provider issued an electronic invitation to participate in the survey and a single reminder to all eligible registrations. No monetary incentives were offered, but patients who completed the full survey were entered into a draw to win a prize (worth up to Can $500) identified by Harris-Decima, the survey company. Informed consent was obtained electronically before commencing the survey. Analyses focused on descriptive statistics, including univariate and bivariate analyses with tests of statistical significance. Analyses were conducted with SPSS version 21.

Results

Observational Study

In 2013/14, 144 of 5598 primary care physicians provided at least one virtual visit, or 2.57% of the total primary care physician population (Figure 1). Male and female providers were equally likely to provide these services (*P*=.89). There were physicians in every age group providing virtual visits, although the likelihood decreased monotonically with increasing age—from 5.3% (33/624) of physicians younger than 35 years to 0.8% (6/724) of those aged 65 years and older (*P*<.001). Physicians defined as low responsibility (ie, practices appear to be walk-in style) were the most likely to bill virtual visit fee items (3.24%, 51/1574), whereas those defined as high responsibility were least likely (1.40%, 14/1000, *P*=.04).

There were 5441 patients who had at least one virtual visit, which equates to 105.45 people per 100,000 of the provincial population. Table 1 shows demographics and health status of virtual visit users, both in numbers and as a percentage of the total population. Use of these services was highest in the 20 to 44 years age group (53.45%, 2908/5441) and lowest among those 85 years and older (0.77%, 42/5441). Use was highest (in percentage of population terms) in the Northern Health Authority (11.27%, 613/5441), a largely rural and remote part of the province, but the next highest rates of use were in the two most urbanized regions—Vancouver Coastal (24.74%, 1346/5441) and Fraser Health (45.07%, 2452/5441).

XSL•FO
RenderX

Figure 1. Percentage of primary care physicians who billed for virtual visits by sex, age group, and responsibility level in 2013/2014.

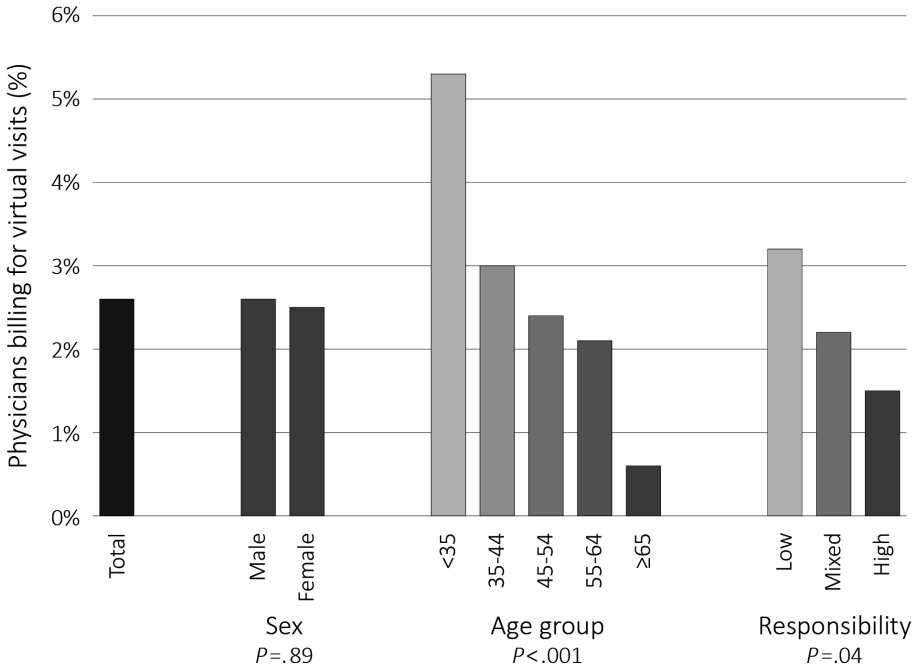
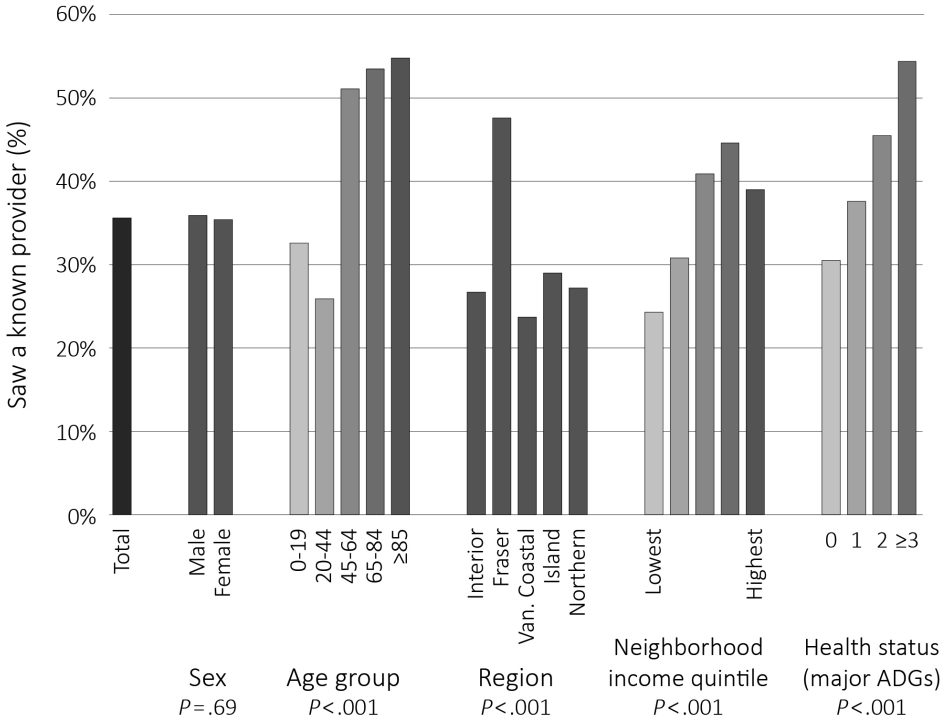


Figure 2. Percentage of patients who saw a known provider for first virtual visit, by sex, age group, region, neighborhood income, and health status in 2013/2014.



Note: P-values result from Chi-square comparing virtual visit users who see a known vs. new provider

XSL•FO
RenderX

Table 1. Demographics of virtual visit users (cohort), the BC total population (comparison group), and traditional primary care users (matched control).

Demographics	Cohort, n (%)	Comparison group, n (%)	Matched control, n (%)	Virtual visit users per 100,000 population, n
N	5441	5,154,164[a]	21,176	105.45
Sex				
Male	2246 (41.28)	2,567,271 (49.81)	6484 (41.21)	87.41
Female	3195 (58.72)	2,585,293 (50.16)	9251 (58.79)	123.43
Age group (years)				
0-19	668 (12.28)	1,041,727 (20.21)	1816 (11.54)	64.08
20-44	2908 (53.45)	1,759,334 (34.13)	8249 (52.42)	165.02
45-64	1356 (24.92)	1,454,896 (28.23)	4127 (26.23)	93.12
65-84	467 (8.58)	735,034 (14.26)	1402 (8.91)	63.49
≥85	42 (0.77)	163,173 (3.17)	141 (0.90)	25.73
Region				
Interior	558 (10.26)	821,389 (15.94)	1596 (10.16)	67.89
Fraser	2452 (45.07)	1,804,191 (35.00)	7150 (45.53)	135.72
Vancouver Coastal	1346 (24.74)	1,287,575 (24.98)	3850 (24.52)	104.43
Island	459 (8.44)	833,223 (16.17)	1343 (8.55)	55.06
Northern	613 (11.27)	318,779 (6.18)	1764 (11.23)	191.93
Income quintile				
Lowest	1065 (20.31)	1,028,089 (20.42)	3326 (21.42)	103.48
2	968 (18.46)	1,022,091 (20.30)	3312 (21.33)	94.62
3	1038 (19.79)	1,014,178 (20.14)	3183 (20.50)	102.24
4	1220 (23.26)	1,007,627 (20.01)	3020 (19.45)	120.93
Highest	953 (18.17)	963,518 (19.13)	2686 (17.30)	98.81
Health status				
0 major ADGs	3044 (55.95)	3,648,817 (70.79)	8699 (55.28)	83.35
1	1451 (26.67)	956,439 (18.56)	4297 (27.31)	151.48
2	569 (10.46)	330,074 (6.40)	1645 (10.45)	172.09
3+	377 (6.93)	218,834 (4.25)	1094 (6.95)	171.98

[a] The total population that lived in British Columbia at any time during 2010/2011 to 2013/2014.

When limited to the first virtual visit per person, just over one-third (35.58%, 1936/5441) of those visits were with providers already known to the patient and two-thirds (64.42%, 3505/5441) were with a new provider (Figure 2). Males and females were equally likely to have seen their physician before their virtual visit (P=.69). People older than 45 years, those who lived in the Fraser Health Authority (largest urban health authority), and those who had more complex health conditions (≥3 major ADGs) were the most likely to have seen their virtual visit provider previously in a nonvirtual setting (P<.001). There was no clear pattern by SES, but those in the lowest quintile were the least likely to have seen their virtual visit provider previously (24.32%, 259/1065; P<.001).

Virtual Visit Users Who Saw a Known Versus a New Provider

Table 2 provides the diagnoses for patients who had a virtual visit in the study period compared to the matched controls. A larger proportion of BC patients with a virtual visit sought care for a "mental disorder" compared to the general population (14.86%, 1262/8494 vs 8.50%, 2103/24,738; P<.001), and for "supplementary factors" (11.80%, 1002/8494 vs 2.50%, 618/24,738; P<.001), the latter of which was largely driven by requests for contraception and contraception advice.

Table 3 provides an overview of visit outcomes for virtual visit users segmented by those seeing a known provider (seen before) or new provider (not seen before) compared to the matched control group with "traditional" visits. The likelihood of receiving a prescription was higher for virtual visit users than in-person visits (49.40%, 3599/7286 vs 45.69%, 10,043/21,981;

XSL•FO
RenderX

P<.001). Referral to diagnostic imaging was lower for virtual visit users than in-person visits (1.88%, 137/7286 vs 6.59%, 1448/21,981; *P*<.001), but more likely if seeing a known provider (*P*<.001). Referral for laboratory testing was about the same for a virtual visit with a known provider (15.36%, 520/3385) and traditional visits (14.96%, 3289/21,981), but much lower (5.95%, 232/3901) for virtual visits with new providers (*P*<.001).

Figure 3 provides the interrupted times series results (regression parameter estimates) comparing patients seeing a known provider to patients seeing a new provider. Before the first virtual visit, the group who saw a known provider had both a higher cost (Can $16.41, *P*<.001) and a larger increasing trend in costs (Can $2.34, *P*<.001) compared to those who saw a new provider. After the intervention, the known provider group showed a decreasing trend (Can –$8.68, *P*<.001) compared to the new provider group, with the values approaching each other by the end of the follow-up period. Follow-up ended after six quarters because the numbers declined to the point that outcomes were not stable after this.

Figure 4 provides the interrupted time series results (regression parameter estimates) of the total virtual visit group compared to a matched comparison group of people who did not have a primary care virtual visit during the study period, but who do had at least one traditional general practitioner (GP) visit. In this case, both groups showed an increasing trend in primary care spending before the GP visit. After the intervention, the patients with a virtual visit had a lower trend than their matched controls (Can –$3.79, *P*=.01). The result was an apparent lower expenditure among the virtual visit group at the end of the follow-up period, but again this period was limited. The trends were the same when the outcome was primary care visits rather than costs (data not shown).

Patient Survey

A total of 399 of 3025 (13.19%) BC residents who had a virtual visit in the past year completed the online patient survey from April 17 to May 1, 2015. The survey took a mean 18 minutes to complete. The majority of respondents, were female (71.4%, 285/399), between 35 and 54 years (45.1%, 180/399), and married (67.4%, 269/399). Population comparisons were provided along with sample demographics, health service utilization, and detailed results are available in the full report [29].

A number of aspects of the patient-physician engagement on the virtual visit were viewed positively, with 93.2% (372/399) reporting that their most recent visit was of high quality, 95.0% (379/399) reporting confidence in the security and privacy of their personal information when using a virtual visit, and 79.0% (315/399) saying their most recent virtual visit was as thorough as an in-person visit.

In terms of visit outcome, 91.2% (364/399) of respondents reported that the virtual visit was “very” or “somewhat” helpful to resolve health issue for which they needed the appointment. Only 1.5% (6/399) of patients reported that they were advised to call 9-1-1 or visit an emergency department immediately. Nearly half (48.4%, 193/399) of patients indicated they would have gone to a walk-in clinic if the virtual visit had not been available, 20.3% (81/399) would have had an in-person visit with their doctor or regular place of care, and 10.8% (43/399) would have gone to the emergency department. A total of 12.5% (50/399) reported that they would not have sought care at that time.

Table 2. Diagnoses for virtual visits, 2011-2014.

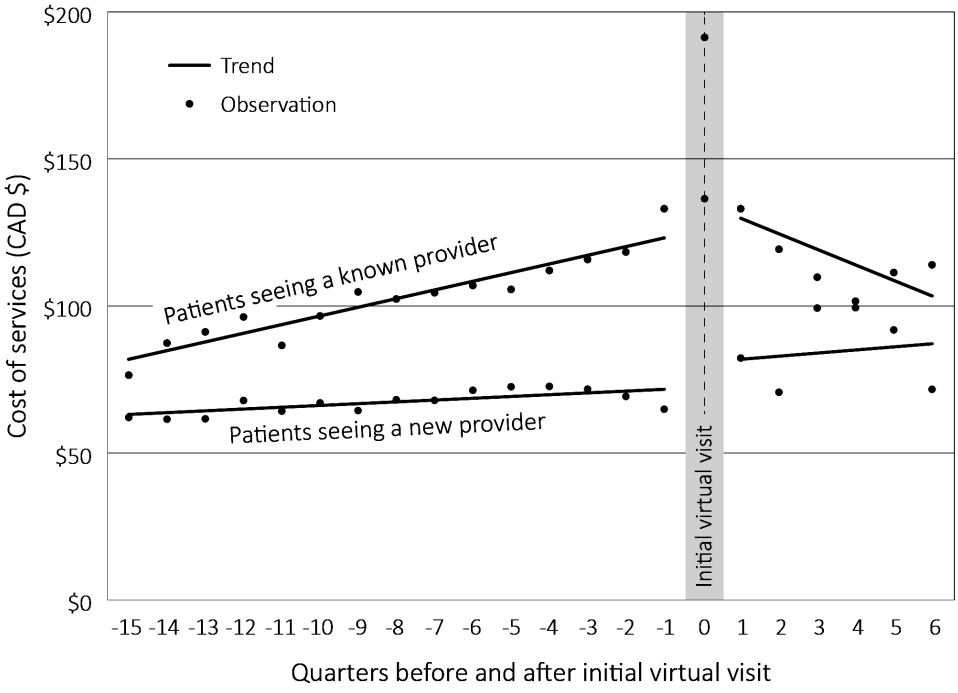
Diagnosis	Virtual visit users, n (%)	Traditional visit users (matched control), n (%)	*P*
Symptoms, signs, and ill-defined conditions	1299 (15.29)	4914 (19.86)	<.001
Mental disorders	1262 (14.86)	2103 (8.50)	<.001
Supplementary factors influencing health status and contact with health services	1002 (11.80)	618 (2.50)	<.001
Diseases of the respiratory system	852 (10.03)	3337 (13.49)	<.001
Diseases of the musculoskeletal system and connective tissue	540 (6.36)	1932 (7.81)	<.001
Diseases of nervous system and sense organs	525 (6.18)	1600 (6.47)	<.001
Endocrine, nutritional and metabolic diseases and immunity disorders	507 (5.97)	1387 (5.61)	<.001
Infections and parasitic diseases	467 (5.50)	1317 (5.32)	<.001
Diseases of the circulatory system	381 (4.49)	1418 (5.73)	<.001
Diseases of the genitourinary system	376 (4.43)	1662 (6.72)	<.001

XSL•FO
RenderX

Table 3. Number of visits resulting in a referral or other follow-up care for virtual and matched control visits, 2013/2014.

Referral/follow-up	Virtual visits (n=7286)				Matched control visits (n=21,981)	
	n (%)	Known provider, %	New provider, %	*P*	n (%)	*P*
General practice	113 (1.55)	2.22	0.97	<.001	314 (1.43)	.45
Medical specialists	125 (1.72)	1.60	1.82	.46	356 (1.62)	.58
Surgical specialists	78 (1.07)	1.18	0.97	.39	480 (2.18)	<.001
Imaging	137 (1.88)	2.90	1.00	<.001	1448 (6.59)	<.001
Laboratory test	752 (10.32)	15.36	5.95	<.001	3289 (14.96)	<.001
Prescription	3599 (49.40)	49.28	49.50	.85	10,043 (45.69)	<.001

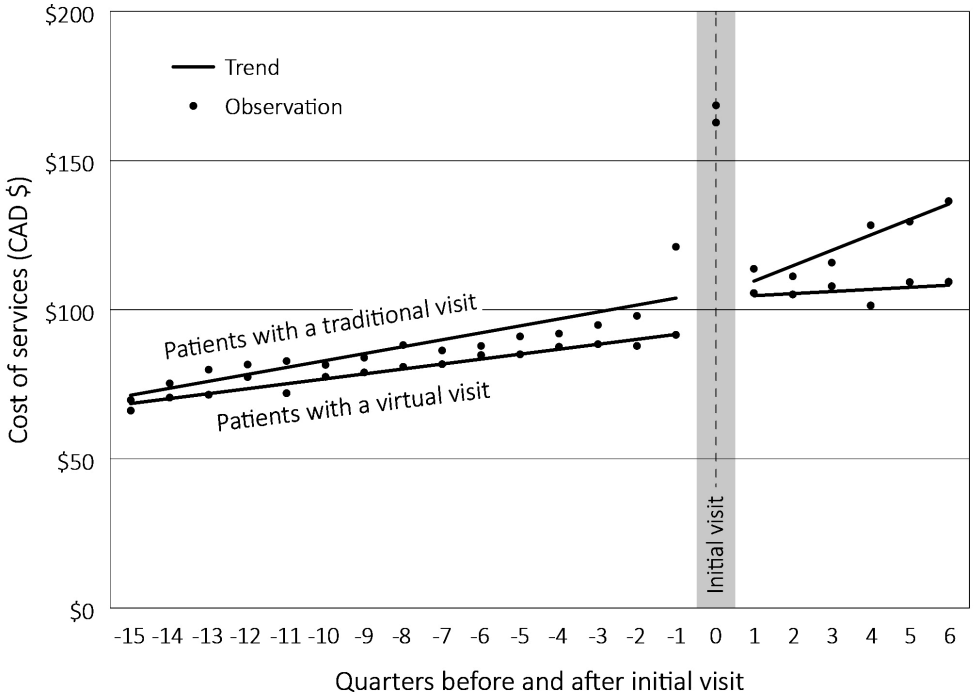
Figure 3. Time series analyses comparing virtual visit patients seeing a known provider to patients seeing a new provider, 2011-2014.



	Inter-cept	Trend in controls	Diff. in level prior	Diff. in trend prior	Chg. in level for controls	Chg. in trend for controls	Diff. in level after	Diff. in trend after
Known vs. new provider	62.51 *P*<.001	0.61 *P*=.01	16.41 *P*<.001	2.34 *P*<.001	8.54 *P*=.22	0.45 *P*=.79	0.45 *P*=.96	-8.68 *P*<.001

XSL•FO
RenderX

Figure 4. Time series analyses comparing patients with a virtual visit to matched controls with a traditional visit, 2011-2014.



	Inter-cept	Trend in controls	Diff. in level prior	Diff. in trend prior	Chg. in level for controls	Chg. in trend for controls	Diff. in level after	Diff. in trend after
Virtual visits vs. trad. visits	68.99 *P*<.001	2.32 *P*<.001	-2.11 *P*=.51	-0.67 *P*=.06	-1.81 *P*=.69	2.84 *P*=.01	12.41 *P*=.06	-3.79 *P*=.01

Discussion

Patients from all demographic and medical backgrounds are using virtual visits. Similar to other studies, younger patients are more likely to have a virtual visit compared to older individuals [3,4], perhaps emphasizing a digital divide between younger patients more comfortable with technology than older patients [30] or patients most likely to not have a regular care provider [31]. Patients with one or more major ADGs are more highly represented, which makes sense if these services are being used for monitoring existing conditions and providing services such as prescription refills that help patients avoid taking time and money to attend an in-person visit.

A small number of physicians are providing virtual visits. Male and female physicians are equally likely to be providing virtual visits, but there is a clear age gradient, with younger physicians far more likely to be using this new technology for patient care. In addition, we see that physicians who are identified as operating low-responsibility practices are significantly more likely to be providing virtual visits than those running high-responsibility practices. This suggests that at least some proportion (and likely a high proportion) of virtual visits may be essentially a virtual walk-in clinic [27].

Virtual visits were being provided in communities in British Columbia regardless of whether they were in urban and rural settings, indicating that virtual visits may not necessarily be filling all geographic gaps in primary care delivery [32]. Our analyses show that approximately one-third of people using virtual visits are seeing physicians with whom they previously interacted in a traditional office visit setting, whereas the remainder are seeing new providers. There were no sex differences in these percentages, but older people and those with more health problems were more likely to see a known provider. These appear to be positive trends from a patient care standpoint, consistent with monitoring of chronic conditions [33] and provision of patient-centered care [34]. Not seeing a known provider may reflect a desire for convenience over continuity [35], but may also indicate (at least in some cases) a specific preference for a new provider, such as for questions or care (eg, contraception) that patients wish to keep from their regular providers. Although there is no socioeconomic gradient in the use of virtual visits overall, individuals from lower income neighborhoods are less likely to see a known provider in their virtual visit.

The time series analyses comparing virtual versus traditional visits suggests that virtual visits may be beneficial in moderating total primary care costs over the longer term. At the same time, seeing a known versus new provider is better from an overall cost/use perspective. Putting these together, the conclusion is that virtual visits may have a beneficial effect—they are well-liked by patients (as seen in survey results) and they appear

XSL•FO
RenderX

to control costs—but that benefit is most associated with seeing a known provider. Some caution is needed with these interpretations given the small N and limited follow-up period, but they are at least suggestive. The implication is that it matters (potentially quite a lot) *how* virtual visits are embedded in health care delivery systems.

This is consistent with previous research showing that virtual visits can complement existing patterns of care and, in fact, reduce overall primary and urgent care visits for patients [18,36]. Similarly, an evaluation by the US National Institute of Justice found that virtual care more generally (including synchronous and asynchronous care) has specific value for defined patient populations, reducing external visits to specialists and costly off-site transfers for care of prisoners [37] and providing a cost-effective solution for patients with limited access to mental health services [38].

Limitations

Patient-initiated virtual visits in primary care are a relatively new model of access to primary care. Their use is growing rapidly, including in British Columbia, but is still a very small portion of total primary care. Although interrupted time series is a strong quasi-experimental analytic approach for evaluating population-level health interventions [39,40], it does not deal with selection bias and that may influence what we see in these analyses. Patients are not randomized to receive virtual visits, they choose to pursue them, and those who do and do not choose to use these visits may be different in unmeasured ways that are also related to primary care costs. Our comparison of users who see known versus new providers is further limited by the fact that these two groups are different in demographic and health status characteristics. Not all physicians choose to provide virtual visits, so selection at the physician level has some influence on which patients receive health care in this new format. These analyses consider only the costs of primary care. Future analyses should consider broader health care system effects of virtual visits and also the timing of virtual visits in longitudinal episodes of care.

Conclusions

British Columbia is unique in Canada in offering publicly funded virtual visits in primary care. The province-wide implementation—both available for interactions among patients and physicians who know each other and for walk-in-type visits—makes this research valuable to other jurisdictions in Canada and beyond. The number of virtual visits is continuing to increase and it appears that there is no simple conclusion about their effect on existing patterns of health care services use. In some cases, the care is complementary, providing a new way for patients and providers with existing relationships to interact. In other cases, the care may be easing access, perhaps providing patients with needed care at a convenient point in time, but not necessarily displacing subsequent service use. A patient-centered system is one that is organized to respond to patient need without these unintended side effects. Virtual visits may be one means by which the system can be reoriented to be more patient-centered.

In the context of primary care transformation, there are discussions of new models of care and the role of technology in care, with the understanding that technology may be one way to improve care delivery [35]. Our analyses suggest that it is important to consider how such technologies are integrated into the system, whether as an adjunct to existing relationships or simply another way to see a provider on-demand. As technology continues to develop, it is important to ensure health care systems harness it to increase opportunities for patient-centered care.

Acknowledgments

This independent study, including publication costs for this article, was funded by Canada Health Infoway Inc., a non-profit corporation funded by the Government of Canada. The opinions, results, and conclusions reported here are those of the authors. No endorsement by the University of British Columbia or Canada Health Infoway is intended or should be inferred. The funders had no role in study design, data collection and analysis, or decision to publish the manuscript. Population Data BC provided the data for the study with unique identifiers used for practitioners and patients to prevent identification of specific individuals. All inferences, opinions, and conclusions drawn in this report are those of the authors, and do not reflect the opinions or policies of the Data Steward(s). We would like to thank Dawn Mooney for development of tables and figures for the manuscript.

Conflicts of Interest

None declared.

References

1. Jamieson T, Wallace R, Armstrong K, Agarwal P, Griffin B, Wong I, et al. Virtual Care: A Framework for a Patient-Centric System. Toronto, ON: Women's College Hospital Institute for Health Systems Solutions and Virtual Care; 2015. URL: http://www.womenscollegehospital.ca/assets/pdf/wihv/WIHV_VirtualHealthSymposium.pdf [accessed 2017-01-16] [WebCite Cache ID 6natfkSzo]
2. Al-Mahdi I, Gray K, Lederman R. Online medical consultation: a review of literature and practice. In: Proceedings of the 8th Australasian Workshop on Health Informatics and Knowledge Management. 2015 Presented at: 8th Australasian Workshop on Health Informatics and Knowledge Management; Jan 27-30, 2015; Sydney, Australia p. 97-100 URL: http://crpit.com/confpapers/CRPITV164Al-Mahdi.pdf

XSL•FO
RenderX

3. Jung C, Padman R. Virtualized healthcare delivery: understanding users and their usage patterns of online medical consultations. Int J Med Inform 2014 Dec;83(12):901-914. [doi: 10.1016/j.ijmedinf.2014.08.004] [Medline: 25193501]
4. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. Characteristics of patients who seek care via eVisits instead of office visits. Telemed J E Health 2013 Jul;19(7):515-519 [FREE Full text] [doi: 10.1089/tmj.2012.0221] [Medline: 23682589]
5. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May;4(3):666-684 [FREE Full text] [Medline: 20513335]
6. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442]
7. Pong RW, Pitbaldo JR. Geographic Distribution of Physicians in Canada: Beyond How Many and Where. Ottawa, ON: Canadian Institute of Health Information; 2005. URL: https://secure.cihi.ca/free_products/Geographic_Distribution_of_Physicians_FINAL_e.pdf [accessed 2017-01-15] [WebCite Cache ID 6natjrcb5]
8. Sibley LM, Weiner JP. An evaluation of access to health care services along the rural-urban continuum in Canada. BMC Health Serv Res 2011 Jan 31;11(1):20 [FREE Full text] [doi: 10.1186/1472-6963-11-20] [Medline: 21281470]
9. The College of Family Physicians of Canada. 2010 National Physician Survey. 2010. URL: http://nationalphysiciansurvey.ca/wp-content/uploads/2012/05/NPS2010-National-Binder.pdf [accessed 2017-01-14] [WebCite Cache ID 6natm0xTa]
10. Osborn R, Squires D, Doty MM, Sarnak DO, Schneider EC. In new survey of eleven countries, US adults still struggle with access to and affordability of health care. Health Aff (Millwood) 2016 Dec 01;35(12):2327-2336. [doi: 10.1377/hlthaff.2016.1088] [Medline: 27856648]
11. McCarthy D, Mueller K. The Commonwealth Fund. 2009 Jun. Kaiser Permanente: Bridging the quality divide with integrated practice, group accountability, and health information technology URL: http://www.commonwealthfund.org/~/media/files/publications/case-study/2009/jun/1278_mccarthy_kaiser_case_study_624_update.pdf [accessed 2017-01-16] [WebCite Cache ID 6natoGRTA]
12. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. Med Care 1981 Feb;19(2):127-140. [Medline: 7206846]
13. Stewart M. Towards a global definition of patient centred care. BMJ 2001 Feb 24;322(7284):444-445 [FREE Full text] [Medline: 11222407]
14. Shah MN, Gillespie SM, Wood N, Wasserman EB, Nelson DL, Dozier A, et al. High-intensity telemedicine-enhanced acute care for older adults: an innovative healthcare delivery model. J Am Geriatr Soc 2013 Nov;61(11):2000-2007. [doi: 10.1111/jgs.12523] [Medline: 24164485]
15. Deloitte. eVisits: The 21st Century Housecall. 2014. URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/technology-media-telecommunications/deloitte-au-tmt-evisits-011014.pdf [accessed 2017-01-15] [WebCite Cache ID 6nau4z5gb]
16. Terry K. Physicians and telehealth: is it time to embrace virtual visits? Med Econ 2015 Jul 10;92(13):48-52. [Medline: 26510241]
17. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
18. Gordon AS, Adamson WC, DeVries AR. Virtual visits for acute, nonurgent care: a claims analysis of episode-level utilization. J Med Internet Res 2017 Feb 17;19(2):e35 [FREE Full text] [doi: 10.2196/jmir.6783] [Medline: 28213342]
19. Mehrotra A, Lave JR. Visits to retail clinics grew fourfold from 2007 to 2009, although their share of overall outpatient visits remains low. Health Aff (Millwood) 2012 Sep;31(9):2123-2129 [FREE Full text] [doi: 10.1377/hlthaff.2011.1128] [Medline: 22895454]
20. Ashwood JS, Gaynor M, Setodji CM, Reid RO, Weber E, Mehrotra A. Retail clinic visits for low-acuity conditions increase utilization and spending. Health Aff (Millwood) 2016 Mar;35(3):449-455. [doi: 10.1377/hlthaff.2015.0995] [Medline: 26953299]
21. British Columbia Ministry of Health. Medical Services Plan (MSP) Payment Information File. Vancouver, BC: Population Data BC; 2015. URL: https://www.popdata.bc.ca/data/internal/health/msp [accessed 2017-05-15] [WebCite Cache ID 6qU2dtS5G]
22. British Columbia Ministry of Health. Consolidation File (MSP Registration & Premium Billing). Vancouver, BC: Population Data BC; 2015. URL: https://www.popdata.bc.ca/data/internal/population/consolidationfile [WebCite Cache ID 6nf0XTyLQ]
23. British Columbia Ministry of Health. PharmaNet. Vancouver, BC: Population Data BC; 2015. URL: https://www.popdata.bc.ca/data/external/PharmaNet [accessed 2017-01-19] [WebCite Cache ID 6qU2wAz3d]
24. British Columbia Ministry of Health. Medical Services Plan (MSP) Practitioner Information File. Vancouver, BC: Population Data BC; 2015. URL: https://www.popdata.bc.ca/data [accessed 2017-01-20] [WebCite Cache ID 6nf0XTyLQ]
25. Reid RJ, MacWilliam L, Verhulst L, Roos N, Atkinson M. Performance of the ACG case-mix system in two Canadian provinces. Med Care 2001 Jan;39(1):86-99. [Medline: 11176546]
26. Puyat JH, Kazanjian A, Goldner EM, Wong H. How often do individuals with major depression receive minimally adequate treatment? A population-based, data linkage study. Can J Psych 2016 Mar;61(7):394-404 [FREE Full text]

XSL•FO
RenderX

27. McGrail K, Lavergne R, Lewis SJ, Peterson SL, Barer M, Garrison SR. Classifying physician practice style: a new approach using administrative data in British Columbia. Med Care 2015 Mar;53(3):276-282. [Medline: 25634088]
28. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002 Aug;27(4):299-309. [Medline: 12174032]
29. Canada Health Infoway. Virtual Visits in British Columbia: 2015 Patient Survey and Physician Interview Study. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3105-virtual-visits-in-british-columbia-2015-patient-survey-and-physician-interview-study/view-document [accessed 2017-01-17] [WebCite Cache ID 6nau7rcHs]
30. Kontos E, Blake KD, Chou WY, 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]
31. Statistics Canada. 2015. Access to a regular medical doctor, 2014 URL: http://www.statcan.gc.ca/pub/82-625-x/2015001/article/14177-eng.htm [accessed 2017-01-17] [WebCite Cache ID 6nauA0Qle]
32. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and quality of care in direct-to-consumer telemedicine. Telemed J E Health 2016 Apr;22(4):282-287. [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]
33. Hernandez C, Mallow J, Narsavage GL. Delivering telemedicine interventions in chronic respiratory disease. Breathe (Sheff) 2014 Sep;10(3):198-212 [FREE Full text] [doi: 10.1183/20734735.008314] [Medline: 26843894]
34. Dixon RF. Enhancing primary care through online communication. Health Aff (Millwood) 2010 Jul;29(7):1364-1369 [FREE Full text] [doi: 10.1377/hlthaff.2010.0110] [Medline: 20606189]
35. Daniel H, Sulmasy LS, Health and Public Policy Committee of the American College of Physicians. Policy recommendations to guide the use of telemedicine in primary care settings: an American College of Physicians position paper. Ann Intern Med 2015 Nov 17;163(10):787-789. [doi: 10.7326/M15-0498] [Medline: 26344925]
36. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009 Apr;28(2):323-333 [FREE Full text] [doi: 10.1377/hlthaff.28.2.323] [Medline: 19275987]
37. Abt Associates Inc. Telemedicine Can Reduce Correctional Health Care Costs: An Evaluation of a Prison Telemedicine Network. Washington, DC: US Department of Justice Office of Justice Programs; 1999. URL: https://www.ncjrs.gov/pdffiles1/175040.pdf [accessed 2017-05-02] [WebCite Cache ID 6qBeonZuz]
38. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
39. Shadish W, Cook T, Campbell D. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA: Houghton Mifflin; 2002.
40. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2017;46(1):348-355 [FREE Full text] [doi: 10.1093/ije/dyw098] [Medline: 27283160]

Abbreviations

EMR: electronic medical record
GP: general practitioner
HSDA: health service delivery area
ACG: adjusted clinical group
ADG: aggregated diagnosis group
SES: socioeconomic status

Edited by G Eysenbach; submitted 20.01.17; peer-reviewed by T Jamieson, N Lea; comments to author 23.02.17; revised version received 06.04.17; accepted 14.04.17; published 26.05.17.

Please cite as:
McGrail KM, Ahuja MA, Leaver CA
Virtual Visits and Patient-Centered Care: Results of a Patient Survey and Observational Study
J Med Internet Res 2017;19(5):e177
URL: http://www.jmir.org/2017/5/e177/
doi:10.2196/jmir.7374
PMID:28550006

©Kimberlyn Marie McGrail, Megan Alyssa Ahuja, Chad Andrew Leaver. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 26.05.2017. This is an open-access article distributed under the terms of the Creative Commons

XSL•FO
RenderX

Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using a Medical Intranet of Things System to Prevent Bed Falls in an Acute Care Hospital: A Pilot Study

Henri U Balaguera[1,2], MD; Diana Wise[3]; Chun Yin Ng[3], MBA, MPH; Han-Wen Tso[4], PhD; Wan-Lin Chiang[4], PhD; Aimee M Hutchinson[1], RN, MSN; Tracy Galvin[1], MSN; Lee Hilborne[5], MPH, MD; Cathy Hoffman[1], MHA; Chi-Cheng Huang[1,2], MD; C Jason Wang[6], MD, PhD

[1]Lahey Hospital and Medical Center, Burlington, MA, United States

[2]Tufts University School of Medicine, Boston, MA, United States

[3]MedicusTek USA Corporation, Newport Beach, CA, United States

[4]MedicusTek, Inc, Taipei, Taiwan

[5]David Geffen School of Medicine at UCLA, Department of Pathology and Laboratory Medicine, Los Angeles, CA, United States

[6]Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Henri U Balaguera, MD
Lahey Hospital and Medical Center
41 Mall Road
Burlington, MA, 01805
United States
Phone: 1 781 744 3939
Fax: 1 781 744 1597
Email: balaguera@gmail.com

Abstract

Background: Hospitalized patients in the United States experience falls at a rate of 2.6 to 17.1 per 1000 patient-days, with the majority occurring when a patient is moving to, from, and around the bed. Each fall with injury costs an average of US $14,000.

Objective: The aim was to conduct a technology evaluation, including feasibility, usability, and user experience, of a medical sensor-based Intranet of things (IoT) system in facilitating nursing response to bed exits in an acute care hospital.

Methods: Patients 18 years and older with a Morse fall score of 45 or greater were recruited from a 35-bed medical-surgical ward in a 317-bed Massachusetts teaching hospital. Eligible patients were recruited between August 4, 2015 and July 31, 2016. Participants received a sensor pad placed between the top of their mattress and bed sheet. The sensor pad was positioned to monitor movement from patients' shoulders to their thighs. The SensableCare System was evaluated for monitoring patient movement and delivering timely alerts to nursing staff via mobile devices when there appeared to be a bed-exit attempt. Sensor pad data were collected automatically from the system. The primary outcomes included number of falls, time to turn off bed-exit alerts, and the number of attempted bed-exit events. Data on patient falls were collected by clinical research assistants and confirmed with the unit nurse manager. Explanatory variables included room locations (zones 1-3), day of the week, nursing shift, and Morse Fall Scale (ie, positive fall history, positive secondary diagnosis, positive ambulatory aid, weak impaired gait/transfer, positive IV/saline lock, mentally forgets limitations). We also assessed user experience via nurse focus groups. Qualitative data regarding staff interactions with the system were collected during two focus groups with 25 total nurses, each lasting approximately 1.5 hours.

Results: A total of 91 patients used the system for 234.0 patient-days and experienced no bed falls during the study period. On average, patients were assisted/returned to bed 46 seconds after the alert system was triggered. Response times were longer during the overnight nursing shift versus day shift (P=.005), but were independent of the patient's location on the unit. Focus groups revealed that nurses found the system integrated well into the clinical nursing workflow and the alerts were helpful in patient monitoring.

Conclusions: A medical IoT system can be integrated into the existing nursing workflow and may reduce patient bed fall risk in acute care hospitals, a high priority but an elusive patient safety challenge. By using an alerting system that sends notifications directly to nurses' mobile devices, nurses can equally respond to unassisted bed-exit attempts wherever patients are located on

XSL•FO
RenderX

the ward. Further study, including a fully powered randomized controlled trial, is needed to assess effectiveness across hospital settings.

(J Med Internet Res 2017;19(5):e150) doi:10.2196/jmir.7131

KEYWORDS

accidental falls; acute care; nursing; patient safety; patient-centered care; sensor devices and platforms; health care technology; mobile apps; patient monitoring; health technology assessment

Introduction

Each year in the United States, between 700,000 and 1,000,000 hospitalized patients fall [1]. Studies report fall rates between 2.6 and 17.1 per 1000 patient-days with 30% to 50% resulting in injury [2-6]. Several observational studies report that 60% to 80% of in-hospital falls occur when a patient is moving to, from, and around the bed, and 80% of falls are unassisted [7-9]. Cost analysis studies estimate inpatient falls add approximately US $3500 to US $14,000 per patient stay, depending on whether there is serious injury [5,10]. The US Centers for Medicare and Medicaid Services (CMS) lists falls as one of 14 categories of preventable hospital-acquired conditions. Since 2008, CMS no longer pays for the extra health care costs associated with hospital-acquired falls, shifting the cost burden to hospitals [2,11,12].

Fall prevention strategies using innovative medical technology can potentially improve health care delivery and patient safety [13]. A range of medical-alerting devices have already begun using wireless sensor technologies to target pressure injuries and vital sign monitoring; however, only a few have published their results [14-15].

Although alerting technology can help nurses, inaccurate or false alarms can have the opposite effect on both caregivers and patients. Studies show 80% to 99% of monitor alarms are false or clinically insignificant [16-18]. With low positive predictive value (PPV) for monitoring alerts, nurses may experience alarm fatigue. Repeated false alarms leads to desensitization and true alarms requiring intervention could be ignored [19]. An alert monitoring system with a high PPV, routed to the appropriate caregiver or clinician, is needed to provide alerts that will lead to timely action. Furthermore, increasing alert PPV could minimize sleep disruption and positively affect recovery for patients [20].

We report the use of a medical Intranet of things (IoT) system to assess how quickly nurses respond to bed exits in an acute care hospital setting where an automated bed sensor pad system was used to analyze real-time patient movement data and provide timely alerts to nursing staff, with the potential to ultimately reduce bed falls (see Figure 1 for system setup in the hospital setting).

Figure 1. Components of the SensableCare System.



XSL•FO
RenderX

Methods

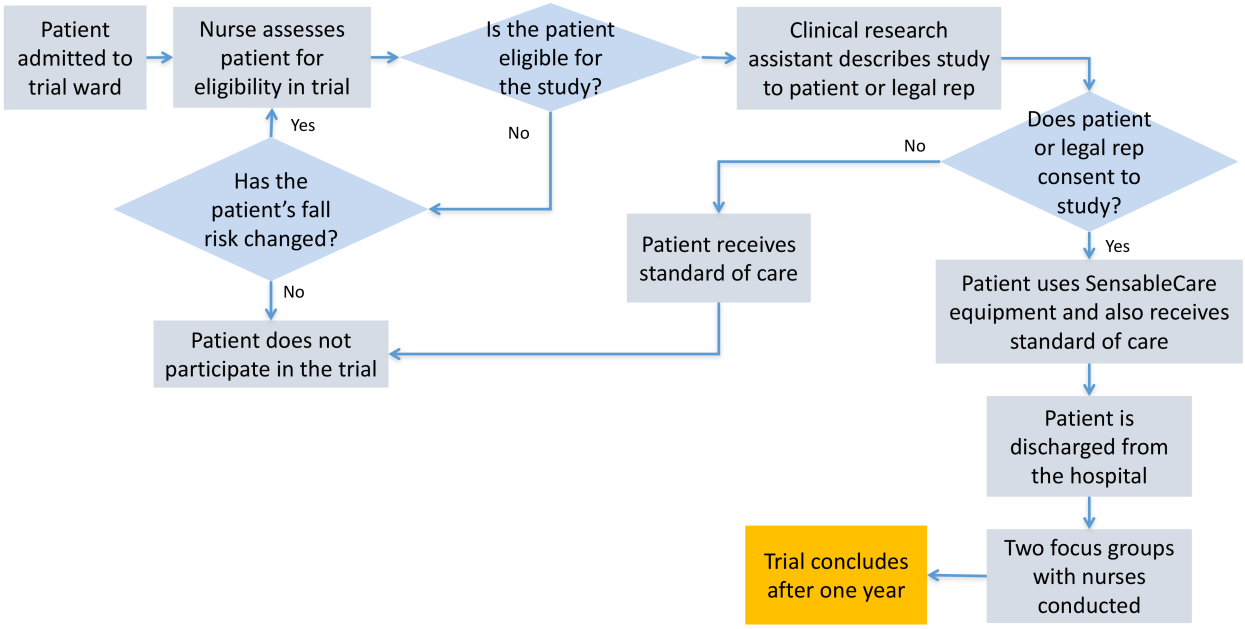
Setting

This study was conducted at a 317-bed suburban Massachusetts teaching hospital for 12 months from August 4, 2015 to July 31, 2016. Patients were recruited from a 35-bed medical-surgical ward. Patients 18 years and older and deemed a high fall risk (Morse Fall Scale score ≥45) were eligible for inclusion in the study [21]. Vulnerable populations (eg, prisoners, patients undergoing stem cell transplant) were excluded. Patients were placed throughout the ward; high fall risk patients (eg, those with encephalopathy) were generally placed in rooms in zone 1 (within 40 feet of the nursing station). Zones 2 and 3 rooms were 41 to 85 feet and 86 to 110 feet away from the nursing station, respectively.

Patient Consent

Clinical research assistants talked with eligible patients about the study when they were first admitted to the ward. Patients interested in participating signed a written consent form. For patients who could not readily consent on their own, family members and legal representatives could consent on behalf of the patient (see Figure 2). Basic demographic information was collected from eligible patients along with their Morse Fall Scale to confirm eligibility. This study was approved by the Lahey Hospital and Medical Center Institutional Review Board (IRB).

Figure 2. Recruitment methodology.



Medical Intranet of Things System Setup

The study tested the SensableCare System for reducing bed falls. A pressure-sensing pad was placed between the mattress surface and the bed sheet (Figure 3). The sensor pad was positioned so it monitored movement from the patients' shoulder to thighs. The pad is approximately 46 inches by 30 inches and encased in a waterproof nylon cover that is coated to provide bacterial resistance. It covered 49.3% of the bed and reached 2.5 inches from the edge of the bed. The pad sensor array detects when pressure is applied. When the system recognizes an increased probability of a bed exit, the system's software algorithm alerts the nursing staff via an app on a mobile device (ie, mobile phone) that nurses carry and to a dashboard at the nursing station. When the patient attempts to leave the bed, an audible message comes from a control box next to the patient's bed reminding the patient not to leave and that a nurse will arrive shortly. Concurrently, nurses receive the alerts at a dashboard and through their mobile phone via audible, vibratory, and visual alerts to prompt them for rapid response. All the system components are connected to the Intranet, allowing caregivers to receive actionable alerts whenever their mobile device is online. Each device on the network is identified by its IP address or MAC address.

The SensableCare System allows nurses to visualize how the patient is positioned in bed at the time of the alert, from which room number the alert is being generated, and informs other nurses when a patient is being assisted. Using a predictive algorithm, the system can also identify patients' activities in bed (eg, whether a patient begins to stir in bed, sits up in bed), before the patient attempts to leave the bed, or is already out of bed.

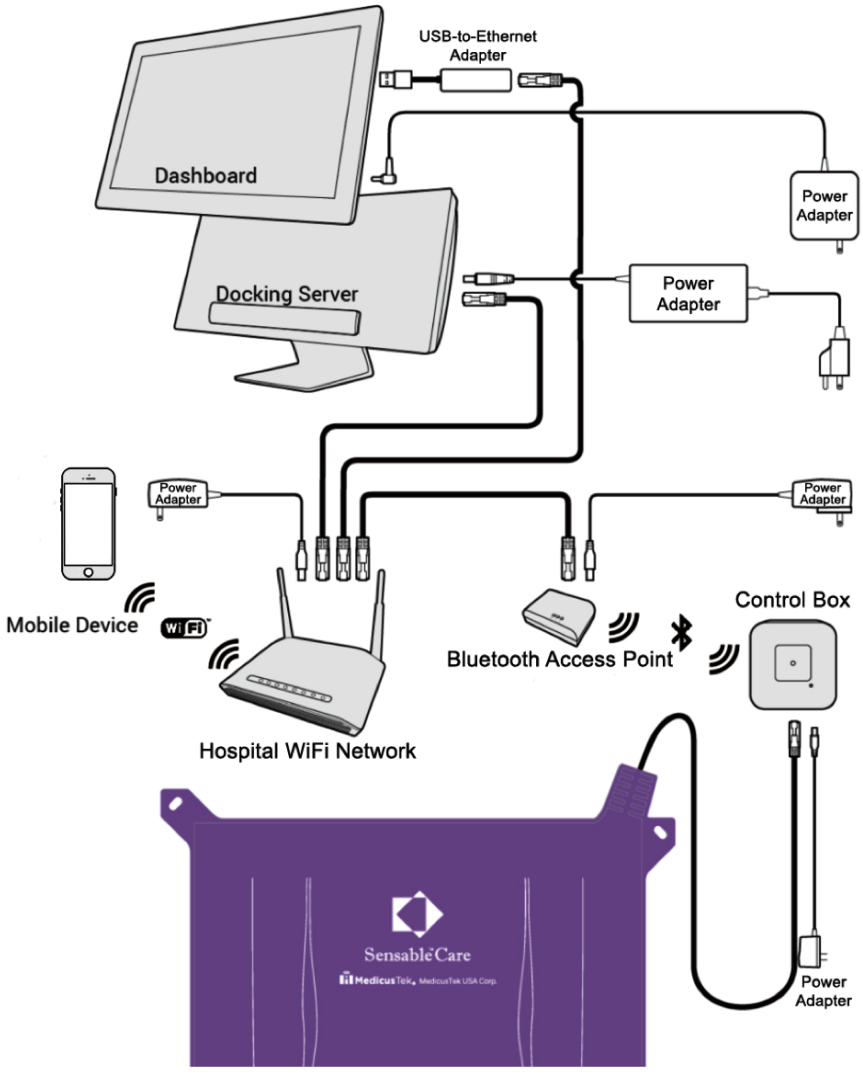
Nurses are able to customize each individual patient's alert settings. Nurses receive bed-leaving alerts when the system detects the patient attempting to leave the bed or is out of the bed. Nurses can also choose to receive alerts earlier, when a patient is stirring after being still for longer than 20 minutes and/or when a patient is sitting up in bed. Both the nurse and nurse aide responsible for an enrolled patient would receive

XSL•FO
RenderX

alerts directly through their mobile phones via a real-time push notification. The charge nurse at the nursing station would also receive alerts through the SensableCare app on the dashboard. Nurses respond to the alert by tapping the control box in the patient's room or returning the patient back to bed, in order to turn off the alerts at the dashboard and on other nurses' mobile phones (ie, Apple iPhone for this trial).

As a condition for the trial, the hospital's existing bed alarm (Stryker Secure II hospital bed with a bed-exit system) was maintained and used concurrently with the SensableCare System for patients who consented to the study. Those who did not consent to the study only used the existing Stryker bed alarm. The Stryker bed signals when a patient is exiting the bed via an audible alert. A high-pitched sound signifies the patient is leaving or is out of bed. A hallway light outside the patient room flashes in conjunction with the audible alert. The nurses would listen for the sound and go to the room with the flashing hallway light to assist the patient. Patients not participating in the study continued to only use the hospital's existing bed fall alarms.

Figure 3. Intranet of things architecture of the SensableCare System. The sensor pad sends data through a cable to the control box located at the patient's bedside. The control box wirelessly transmits this data to the Bluetooth routers located throughout the ward. This information then travels through the hospital network to the dashboard and docking server where the data is analyzed. When an alert is sent to the nurse via an app on their mobile phone, it is wirelessly transmitted through the hospital's Wi-Fi network.



Bed Fall Alerts and Patient Positioning Data

Sensor pad data were collected automatically from the system. These data included patient positioning data, the times the alert was sent to nursing staff, time to turn off bed-exit alerts, when a patient returned to bed, and the duration a patient spent out of bed. Data on patient falls were collected by the clinical research assistants and confirmed with the nurse manager of the unit.

In a previous engineering validation study of 47 patients in a hospital using a low-resolution camera to match movement captured with the sensor pad data, the PPV was 78.5% (unpublished data). System specificity is difficult to calculate in this system because it refers to the true negative rate in which a patient did not leave the bed and the system did not generate an alert. In this study, we calculated the PPV as it was implemented at Lahey Hospital and Medical Center. Because this was a 1-year pilot study approved to understand the integration of medical technology into hospital workflow,

XSL•FO
RenderX

sample size calculation was not performed prior to the study. However, the fall rates using the SensableCare System can be compared to the study unit's historical fall rate with the existing bed alarm during a 12-month period before the trial. The historical unit fall rate during the 12-month period prior to the study was 2.4 falls per 1000 patient-days.

Quantitative Data Analysis

Independent and Dependent Variables

The primary dependent variable of this study was nurse response time to alerts. Time-to-response data were calculated for all events when a nurse responded to an alert within 5 minutes; alarms turned off after 5 minutes were excluded because they were observed to be situations in which patients were assisted without the nurse first disabling the alarm.

To assess possible factors that may predict nurses' response times to alerts, this study included the following independent variables: demographic characteristics of patients (sex, age, weight, and fall risk factors) and characteristics of alerts (alert room, alert day, and alert time). Alert room was categorized as follows: zone 1, which is closest to the nursing station (within 40 feet from nursing station); zone 2, rooms further from the nursing station (41-85 feet away); and zone 3, rooms furthest from nursing station (86-110 feet away). Alert day was classified as "weekday" or "weekend." Alert time was divided into three groups: day shift, evening shift, and overnight shift.

We conducted descriptive analyses to provide a sociodemographic profile of patients and the correlations of alert characteristics and nurses' response times to alerts. After descriptive analysis, linear regression with generalized estimating equations (GEE) was used to examine the relationship among various factors and the alert response time. All *P* values were two-sided. Analyses were done using SAS version 9.4 (SAS Institute Inc, Cary, NC, USA).

Qualitative Study With Focus Groups

Qualitative data regarding staff interactions with the system were collected during two focus group sessions with a total of 25 nurses, each lasting approximately 1.5 hours. The focus groups were held on consecutive days to allow nurses separate opportunities to attend. Two research staff conducted the focus group with one serving as moderator and the other as note taker. Nurse managers invited all hospital staff working in the ward to voluntarily participate without monetary compensation. Research staff used a prepared list of questions approved by the IRB to facilitate focus group discussion. Detailed transcriptions were analyzed for themes related to barriers and facilitators for integrating the system into nurses' clinical workflow.

Results

Falls and Response to Alerts

During the study period, 91 patients used the system for 234.0 patient-days and experienced no bed falls. Both male and female patients with a range of ages, weights, and Morse fall scores participated in the trial (Table 1).

XSL•FO
RenderX

Table 1. Patient profiles (N=91).

Characteristics	Participants
Sex, n (%)	
Male	44 (48)
Female	47 (52)
Age (years), n (%)	
<65	20 (22)
65-74	22 (24)
≥75	49 (54)
Weight (lb), n (%)	
<150	24 (26)
150-199	37 (41)
≥200	26 (29)
Missing[a]	4 (4)
Morse Fall Score, n (%)	
45-60	41 (45)
65-80	20 (22)
85-125	30 (33)
Zone,[b] n (%)	
Zone 1	29 (32)
Zone 2	29 (32)
Zone 3	33 (36)
Patient-days on trial	
Total, n	234.0
Per patient, mean (SD)	2.6 (2.1)
Morse Fall Score items (Morse Fall Score points), n (%)	
Falls history	
No (0)	39 (43)
Yes (25)	52 (57)
Secondary diagnosis	
No (0)	2 (2)
Yes (15)	89 (98)
Ambulatory aid	
Bed rest/nurse assist (0)	43 (47)
Crutches/cane/walker (15)	48 (53)
Furniture (30)	0 (0)
IV/saline lock	
No (0)	2 (2)
Yes (20)	89 (98)
Gait/Transferring	
Normal/bedrest/immobile (0)	13 (14)
Weak (10)	74 (81)
Impaired (20)	4 (4)

XSL•FO
RenderX

Characteristics	Participants
Mental status	
Oriented to own ability (0)	74 (81)
Forgets limitations (15)	17 (19)

[a] Weights were not recorded in the electronic health record.

[b] Zone 1: closest to nursing station (within 40 feet from nursing station); zone 2: rooms further from nurse station (41-85 feet from nursing station); zone 3: rooms furthest from nursing station (86-110 feet from nursing station).

Nursing staff responded to alerts on average within a mean 45.9 (SD 64.7) seconds (Table 2). The SensableCare System bed alarm's PPV was 62.1% (260 positive bed-leaving attempts/419 total bed alerts). The alert reset times among the three shifts were mean 42.3 (SD 62.4) seconds for day shift; mean 43.7 (SD 61.2) seconds for night shift, and mean 57.1 (SD 74.0) seconds for overnight shift; the overnight shift required more time to respond to alerts (*P*=.004; see Table 2).

Table 2. Alert characteristics and response.

Alert characteristics	Events, n	Alerts reset (within 5 min),[a] n	Time to turn off alert (sec), mean (SD)	*P* [b]
Total events (bed leaving/out of bed)	1645	1416	45.9 (64.7)	
Alert room[c]				.002
Zone 1	710	628	41.3 (58.3)	
Zone 2	413	363	42.8 (61.1)	
Zone 3	522	425	55.2 (75.0)	
Alert day				.17
Weekday	1200	1040	47.2 (66.3)	
Weekend	445	376	42.1 (59.9)	
Alert time[d] (by shift)				.004
Day shift	703	581	42.3 (62.4)	
Night shift	612	543	43.7 (61.2)	
Overnight shift	330	292	57.1 (74.0)	

[a] If alert was reset or patient return to bed was >5 minutes, they were considered unaddressed alerts and not included in this table.

[b] Using *t* test (two groups) or ANOVA (three groups).

[c] Zone1: within 40 feet of nursing station; zone2: within 40-85 feet of nursing station; zone 3: within 85-110 feet of nursing station.

[d] Day shift: 07:00-14:59; evening shift: 15:00-22:59; overnight shift: 23:00-06:59 #

In multivariate linear regression using GEE (Table 3), response times were longer during the overnight nursing shift (beta=14.22, *P*=.005) compared to the day shift. Patient characteristics, alert room, and alert day were not significantly associated with alert response time.

XSL•FO
RenderX

Table 3. Factors predicting response time: multivariate linear regression using GEE.

Response time factors	Beta (SE)	*P*
Patient characteristics		
Sex (ref: female)		
Male	–5.07 (6.44)	.43
Age, year (ref: <65)		
65-74	7.58 (10.10)	.45
≥75	6.63 (8.34)	.43
Weight, lb (ref: 150-199)		
<150	–11.88 (6.45)	.07
≥200	13.01 (9.52)	.17
Falls history (ref: no)		
Yes	8.93 (5.07)	.08
Ambulatory aid (ref: bed rest/nurse assist)		
Crutches/cane/walker	–4.69 (5.39)	.38
Gait/Transferring (ref: normal/bedrest/immobile)		
Weak/impaired	–5.05 (8.79)	.57
Mental status (ref: oriented to own ability)		
Forgets limitations	–1.77 (5.94)	.77
Alert characteristics		
Alert room (ref: zone 2+3)		
Zone 1	–0.62 (4.76)	.90
Alert day (ref: weekend)		
Weekday	0.99 (3.84)	.80
Alert time (ref: day shift)		
Evening shift	0.65 (4.00)	.87
Overnight shift	14.22 (5.08)	.005

User Experience

Focus group participants found the SensableCare System easy to use and valued it as an effective technology for reducing bed falls: “It is helpful for really weak and unsteady patients,” reported one participant. Participants were satisfied overall with the user interface, system design, and mobile phone notifications: “...instead of browsing, you can see [in] which room [the alarm] is going off...notifications on the phone instead of browsing through the hallway...I find it positive, takes you less time.” Specific alerts on patient statuses (ie, stirring, sitting up, leaving, out of bed) were informative in understanding patient behaviors and comfort levels: “I like to know that [the patient] has moved in bed, my patient attempted to get out of bed, etc” and “Stirring shows patient is a little uncomfortable in the bed so check on them anyway.” There were self-reported instances in which nurses did not respond to an alert (ie, nurses were not carrying their mobile phones, when they were caring for another patient, or when they were busy and otherwise unable to check the phone): “Sometimes I can’t drop what I am doing to check the phone.” Nurses generally found the current SensableCare System useful and integrated well into their nursing workflow. Participants expressed that they could envision the system in multiple hospital settings.

When there were false alerts, nurses expressed that the system sometimes would sound when the patient was going back to bed. This was quickly remedied by tapping the control box when the nurse was at the patient bedside. Patients who rolled over toward the edge of the bed could also trigger a leaving alert. However, in those instances, nurses expressed that it was a minor inconvenience to have the ability to prevent bed falls for their patients.

Discussion

To our knowledge, this is the first study using a predictive algorithm with data collected from sensors to proactively reduce the fall risk in an acute care setting. By providing nurses with alerts to unassisted bed exits, the study looked at whether nurses responded quickly using this system. The ability for such a bed alert system to help prevent unassisted bed-leaving events might depend on at least three key technology factors: increased time for nursing/caregiver response, enabling nurses to respond from

XSL•FO
RenderX

wherever they are, and eliciting a specific response to prevent bed falls.

On average, nurses reset the alarm approximately 46 seconds after an alert was triggered. Because the system's software algorithm enables an alert to nursing staff's mobile device directly when a patient is initiating a departure from bed, they have more time to respond, hence decreasing the likelihood of bed falls. Furthermore, nurses can choose to receive earlier notifications (eg, patients stirring, sitting up), creating a graduated sequence of potential bed-exit notifications. Giving nurses additional time to respond to a potential bed exit allows them to respond in an appropriate and timely manner. Because nurses must balance multiple patient care responsibilities, it may take some time to drop what they are currently doing to address an immediate patient safety risk. Furthermore, once nurses arrive at the bedside, they are likely to assist the patient first before turning off the sounding alarm. This sequence of events may increase the observed response time recorded by our system.

Currently, high fall risk patients are placed in zones closest to the nursing station for better monitoring and responsiveness to bed-leaving events. Our study found there is no significant correlation between a patient's location on the ward and nursing response time. This may suggest that by using an alerting system that sends notifications directly to nurses' mobile devices, nurses can equally respond to unassisted bed exits wherever patients are located on the ward. Nurses may no longer have to move patients to a room closer to the nursing station.

The PPV with the SensableCare System in this study is 62.1%. One multicenter study found that alarms from a typical physiologic monitor have PPVs of 27% [22]; the same study saw only 5.9% of the monitor alarms led to a nurse's response [22]. Alarm fatigue may cause health care professionals to lower the alarm volume, adjust alarm settings to a point that is unsafe for the patient, or even ignore or deactivate the alarm. One study found that desensitization to alarms and missing alarms have been attributed to patient deaths [23]. SensableCare System's high PPV may have reduced concerns for false alarms.

A medical IoT system may provide information that is helpful in improving hospital operations. Comparing unit alert response times to fall rates may give administrators a metric for potentially reducing bed falls. Nurses can use the number of alerts generated during a particular shift to think about staffing levels and whether they are adequate to address potential bed fall events in a timely manner. Moreover, patient experience will improve with less noise on the ward when alerts are sent to specific nurses.

In summary, we have demonstrated that the SensableCare System is feasible and can be integrated to acute care hospitals. Nurses responded to the system bed alerts quickly and saw no falls for patients who participated during the pilot study. Further work to understand the total cost of ownership to operate a medical IoT can help hospital administrators calculate the cost benefits of using such a system. These costs can be determined if a hospital uses the SensableCare System to optimize nurse staffing or uses a dedicated staff member to carry a mobile device to quickly respond and help patients in a ward.

There are some limitations to this pilot study, including the small number of patients, recruitment of only high fall risk patients, and data from a single hospital unit. Moreover, the standard of care for preventing bed falls, which was the hospital's existing bed alarm, was turned on while the study was taking place on the same beds. This could underestimate the effect of the system if used alone. Also, because nurses were aware that fall risk and the SensableCare System were being studied, they may have been more vigilant (ie, Hawthorne effect) when using the system.

In conclusion, the preliminary evidence suggests that a technological solution using IoT may mitigate the long-standing patient safety fall risk in acute care hospitals while providing hospitals with baseline information for quality improvement, including response time from alarm to assist. Further study is necessary to fully assess the effectiveness of such systems in hospital settings.

Acknowledgments

Rosemary Tin, MSM, Lahey Hospital and Medical Center, provided hospital data for comparison. Mark Anderson, PhD, MedicusTek USA, contributed to the analysis of sensor data, and reviewed the manuscript. Paul Wise, MD, MPH, of Stanford University, Stanford, CA, critically reviewed the manuscript.

MedicusTek USA Corporation was the sponsor and source of funding for this study. The sponsor reviewed and approved the manuscript for publication.

Conflicts of Interest

MedicusTek, the maker of the medical sensor-based Intranet of things, was the sponsor for this study at Lahey Hospital and Medical Center. Several of the authors are affiliated with MedicusTek: Diana Wise, Chun Yin Ng, Han-Wen Tso, and Wan-Lin Chiang are employed by MedicusTek. Drs C Jason Wang, Lee Hilborne, and Chi-Cheng Huang are scientific advisors at MedicusTek.

References

1. Butcher L. The no-fall zone. Hosp Health Netw 2013 Jun;87(6):26-30. [Medline: 23885481]

XSL•FO
RenderX

2. Bouldin EL, Andresen EM, Dunton NE, Simon M, Waters TM, Liu M, et al. Falls among adult patients hospitalized in the United States: prevalence and trends. J Patient Saf 2013 Mar;9(1):13-17 [FREE Full text] [doi: 10.1097/PTS.0b013e3182699b64] [Medline: 23143749]
3. Shorr RI, Mion LC, Chandler AM, Rosenblatt LC, Lynch D, Kessler LA. Improving the capture of fall events in hospitals: combining a service for evaluating inpatient falls with an incident report system. J Am Geriatr Soc 2008 Apr;56(4):701-704 [FREE Full text] [doi: 10.1111/j.1532-5415.2007.01605.x] [Medline: 18205761]
4. Oliver D, Healey F, Haines TP. Preventing falls and fall-related injuries in hospitals. Clin Geriatr Med 2010 Nov;26(4):645-692. [doi: 10.1016/j.cger.2010.06.005] [Medline: 20934615]
5. Joint Commission. Preventing falls and fall-related injuries in health care facilities. Sentinel Event Alert 2015 Sep 28(55):1-5. [Medline: 26422837]
6. Fields J, Alturkistani T, Kumar N, Kanuri A, Salem DN, Munn S, et al. Prevalence and cost of imaging in inpatient falls: the rising cost of falling. Clinicoecon Outcomes Res 2015;7:281-286 [FREE Full text] [doi: 10.2147/CEOR.S80104] [Medline: 26082653]
7. Hitcho EB, Krauss MJ, Birge S, Claiborne DW, Fischer I, Johnson S, et al. Characteristics and circumstances of falls in a hospital setting: a prospective analysis. J Gen Intern Med 2004 Jul;19(7):732-739 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30387.x] [Medline: 15209586]
8. Oliver D. Bed falls and bedrails--what should we do? Age Ageing 2002 Sep;31(5):415-418. [Medline: 12242214]
9. Hanger HC, Ball MC, Wood LA. An analysis of falls in the hospital: can we do without bedrails? J Am Geriatr Soc 1999 May;47(5):529-531. [Medline: 10323644]
10. Quigley PA, White SV. Hospital-based fall program measurement and improvement in high reliability organizations. Online J Issues Nurs 2013 May 31;18(2):5 [FREE Full text] [Medline: 23758423]
11. Spoelstra SL, Given BA, Given CW. Fall prevention in hospitals: an integrative review. Clin Nurs Res 2012 Feb;21(1):92-112. [doi: 10.1177/1054773811418106] [Medline: 21862700]
12. Waters TM, Daniels MJ, Bazzoli GJ, Perencevich E, Dunton N, Staggs VS, et al. Effect of Medicare's nonpayment for Hospital-Acquired Conditions: lessons for future policy. JAMA Intern Med 2015 Mar;175(3):347-354. [doi: 10.1001/jamainternmed.2014.5486] [Medline: 25559166]
13. Gottlieb S, Makower J. A role for entrepreneurs: an observation on lowering healthcare costs via technology innovation. Am J Prev Med 2013 Jan;44(1 Suppl 1):S43-S47 [FREE Full text] [doi: 10.1016/j.amepre.2012.09.006] [Medline: 23195166]
14. Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. Am J Med 2014 Mar;127(3):226-232. [doi: 10.1016/j.amjmed.2013.12.004] [Medline: 24342543]
15. Siddiqui A, Behrendt R, Lafluer M, Craft S. A continuous bedside pressure mapping system for prevention of pressure ulcer development in the medical ICU: a retrospective analysis. Wounds 2013 Dec;25(12):333-339. [Medline: 25867745]
16. Atzema C, Schull MJ, Borgundvaag B, Slaughter GR, Lee CK. ALARMED: adverse events in low-risk patients with chest pain receiving continuous electrocardiographic monitoring in the emergency department. A pilot study. Am J Emerg Med 2006 Jan;24(1):62-67. [doi: 10.1016/j.ajem.2005.05.015] [Medline: 16338512]
17. Siebig S, Kuhls S, Imhoff M, Gather U, Schölmerich J, Wrede CE. Intensive care unit alarms--how many do we need? Crit Care Med 2010 Feb;38(2):451-456. [doi: 10.1097/CCM.0b013e3181cb0888] [Medline: 20016379]
18. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med 1997 Apr;25(4):614-619. [Medline: 9142025]
19. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
20. Matthews EE. Sleep disturbances and fatigue in critically ill patients. AACN Adv Crit Care 2011;22(3):204-224 [FREE Full text] [doi: 10.1097/NCI.0b013e31822052cb] [Medline: 21808157]
21. Ganz D, Huang C, Saliba D, Shier V, Berlowitz D, Lukas C, et al. Preventing falls in hospitals: a toolkit for improving quality of care (No 13-0015-EF). Rockville, MD: Agency for Healthcare Research and Quality; 2013. URL: http://www.ahrq.gov/sites/default/files/publications/files/fallpxtoolkit.pdf [accessed 2016-11-18] [WebCite Cache ID 6m7NBzlet]
22. Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. Intensive Care Med 1999 Dec;25(12):1360-1366. [Medline: 10660842]
23. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care 2013;24(4):378-386; quiz 387. [doi: 10.1097/NCI.0b013e3182a903f9] [Medline: 24153215]

Abbreviations

GEE: generalized estimating equation
IoT: Intranet of things
IRB: Institutional Review Board
PPV: positive predictive value

XSL•FO
RenderX

Edited by A Keepanasseril; submitted 07.02.17; peer-reviewed by T Wray, K Saleem; comments to author 06.03.17; revised version received 02.04.17; accepted 04.04.17; published 04.05.17.

Please cite as:
Balaguera HU, Wise D, Ng CY, Tso HW, Chiang WL, Hutchinson AM, Galvin T, Hilborne L, Hoffman C, Huang CC, Wang CJ
Using a Medical Intranet of Things System to Prevent Bed Falls in an Acute Care Hospital: A Pilot Study
J Med Internet Res 2017;19(5):e150
URL: http://www.jmir.org/2017/5/e150/
doi:10.2196/jmir.7131
PMID:28473306

©Henri U Balaguera, Diana Wise, Chun Yin Ng, Han-Wen Tso, Wan-Lin Chiang, Aimee M Hutchinson, Tracy Galvin, Lee Hilborne, Cathy Hoffman, Chi-Cheng Huang, C Jason Wang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 04.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Smart Toy to Enhance the Decision-Making Process at Children's Psychomotor Delay Screenings: A Pilot Study

María Angeles Gutiérrez García[1*], PhD; María Luisa Martín Ruiz[2*], PhD; Diego Rivera[3*], MEng; Laura Vadillo[2*], MEng; Miguel Angel Valero Duboy[2*], PhD

[1]Universidad Autónoma de Madrid, Departamento de Didáctica y Teoría de la Educación, Facultad de Formación de Profesorado y Educación, Madrid, Spain

[2]Grupo de Investigación Tecnologías para la sociedad de la información y el conocimiento (T>SIC), Departamento de Ingeniería Telemática y Electrónica, Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

[3]Universidad de Alcalá de Henares, Departamento de Automática, Escuela Politécnica Superior, Madrid, Spain

[*]all authors contributed equally

Corresponding Author:
María Luisa Martín Ruiz, PhD
Grupo de Investigación Tecnologías para la sociedad de la información y el conocimiento (T>SIC)
Departamento de Ingeniería Telemática y Electrónica
Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación, Universidad Politécnica de Madrid
Carretera de Valencia, Km 7
Madrid, 28031
Spain
Phone: 34 91 3363770
Fax: 34 91 3363770
Email: marialuisa.martinr@upm.es

Abstract

Background: EDUCERE ("Ubiquitous Detection Ecosystem to Care and Early Stimulation for Children with Developmental Disorders") is an ecosystem for ubiquitous detection, care, and early stimulation of children with developmental disorders. The objectives of this Spanish government-funded research and development project are to investigate, develop, and evaluate innovative solutions to detect changes in psychomotor development through the natural interaction of children with toys and everyday objects, and perform stimulation and early attention activities in real environments such as home and school. Thirty multidisciplinary professionals and three nursery schools worked in the EDUCERE project between 2014 and 2017 and they obtained satisfactory results. Related to EDUCERE, we found studies based on providing networks of connected smart objects and the interaction between toys and social networks.

Objective: This research includes the design, implementation, and validation of an EDUCERE smart toy aimed to automatically detect delays in psychomotor development. The results from initial tests led to enhancing the effectiveness of the original design and deployment. The smart toy, based on stackable cubes, has a data collector module and a smart system for detection of developmental delays, called the EDUCERE developmental delay screening system (DDSS).

Methods: The pilot study involved 65 toddlers aged between 23 and 37 months (mean=29.02, SD 3.81) who built a tower with five stackable cubes, designed by following the EDUCERE smart toy model. As toddlers made the tower, sensors in the cubes sent data to a collector module through a wireless connection. All trials were video-recorded for further analysis by child development experts. After watching the videos, experts scored the performance of the trials to compare and fine-tune the interpretation of the data automatically gathered by the toy-embedded sensors.

Results: Judges were highly reliable in an interrater agreement analysis (intraclass correlation 0.961, 95% CI 0.937-0.967), suggesting that the process was successful to separate different levels of performance. A factor analysis of collected data showed that three factors, trembling, speed, and accuracy, accounted for 76.79% of the total variance, but only two of them were predictors of performance in a regression analysis: accuracy (P=.001) and speed (P=.002). The other factor, trembling (P=.79), did not have a significant effect on this dependent variable.

Conclusions: The EDUCERE DDSS is ready to use the regression equation obtained for the dependent variable "performance" as an algorithm for the automatic detection of psychomotor developmental delays. The results of the factor analysis are valuable to simplify the design of the smart toy by taking into account only the significant variables in the collector module. The fine-tuning

XSL•FO
RenderX

of the toy process module will be carried out by following the specifications resulting from the analysis of the data to improve the efficiency and effectiveness of the product.

(J Med Internet Res 2017;19(5):e171) doi:10.2196/jmir.7533

KEYWORDS

research instruments; questionnaires and tools; Information retrieval; Internet of things; clinical information and decision making; Web-based and mobile health interventions; developmental delays; smart toys

Introduction

In early childhood, children learn by playing in natural settings such as the playground, home, or kindergarten. By playing, children develop various cognitive, perceptual, motor, linguistic, and communicative skills. When a child plays alone, he or she interacts with objects, usually toys, and performs various movements, such as picking or throwing objects, placing them in a row, or stacking them on top of one another. Not all these movements are simple; for example, building a tower with stackable cubes can be difficult for children between ages 2 and 3 years, especially if the cubes are small and the tower is high. Standardized developmental tests include items that ask children to make a tower by stacking cubes [1]. At 24 months of age, children must be able to stack at least five cubes. If a child does not succeed, he or she will be below the standard 24-month-old score, which may indicate some kind of delay in psychomotor development, the seriousness of which depends on the number of cubes the child can stack. Of course, a child development test is composed of several items that measure not only psychomotor development, but also language development, for example, through the understanding of instructions.

The research performed in the EDUCERE (Ecosistema de Detección Ubicua, atenCión y Estimulación tempRana para niños con trastornos del dEsarrollo; "Ubiquitous Detection Ecosystem to Care and Early Stimulation for Children with Developmental Disorders") project focuses on psychomotor development and analyzes how toddlers build a tower of cubes by electronically recording all significant movements performed by the children. The tower of five stacking cubes is an example of the psychomotor behavior of toddlers aged between 2 and 3 years. The aim is to determine if children are capable of building the tower and to analyze how they do it in order to detect minimal delays in development, which may lead to preventive monitoring of some children or, when appropriate, to the implementation of an early attention program.

Ubiquitous computing and ambient intelligence could support innovative application domains, such as the detection of motor impairments within the home environment [2]. Hence, the embedding of different kinds of sensors into everyday toys will allow the collection of systematic information processes and actions in order to make an early detection of potential problems that may affect development in the field of mobility. Furthermore, the detection of a potential motor problem paves the way to the utilization of this technology for early attention to children through educational activities that can mitigate possible additional effects in the future.

Further to this contextualization, the goal of the EDUCERE project [3], including the cube-based smart toys design presented in Rivera et al [4], is to build a developmental delay screening system (DDSS) at home or school that can record children's behavior and skills in order to detect early psychomotor developmental problems and promote stimulation activities. A multidisciplinary group of 30 researchers specialized in disciplines of childhood development (educators, psychologists, physiotherapists) and computer science engineers defined the smart toys model, the set of detectable measures, the embeddable sensors kits, and useful numeric information. The first toy selected by the working team was a set of cubes and the defined activity was to build a tower of cubes [4]. The aim of the DDSS is to provide authorized professionals with sufficient reliable information about the activity performed by a child. Smart toys help professionals by giving information about the following parameters [4]: (1) motion pattern while the child is moving a cube, including time of activity, acceleration, speed, and shaking data, and (2) tower status, including knowledge about how the children made the cube tower, how long it took, and how accurate was the alignment of cubes in the tower.

Early detection of developmental problems is a critical matter to assure the wellness of children [5]. Nowadays, most experts use different activities to evaluate the evolution of child development and motor skills, and many of these activities involve the manipulation of toys and other objects [6]. In fact, there are developmental scales, such as Merrill-Palmer [1] or Bayley [7] scales, which use specific toys and activities done by the child and are employed to identify possible delays or difficulties.

There are many childhood disorders, such as autism spectrum disorders, that can be detected by using information from the child's movements in certain activities [8-10]. Therefore, using sensors to obtain this kind of information is a logical step toward a more accurate detection process [11]. For instance, Marschollek et al [12] showed a classification of sensors to be used for these tasks, although oriented to wearable devices. In addition, unobtrusive wearable devices such as wristbands help to detect and measure movements [13]. Moreover, Taffoni et al [14] described how wearable sensors can measure children's movement when stacking a pile of cubes.

These approaches are partially intrusive because children have to use special wearable devices that are not part of their everyday routines. The goal of the EDUCERE project is to embed the measurement tools in ordinary objects, extending the Internet of things (IoT) paradigm to an "Internet of toys" experience by creating smart toys based on everyday objects equipped with low-cost sensors. This allows the acquisition of accurate information without interfering in children's daily activities.

XSL•FO
RenderX

This approach to the IoT using toys as smart objects is not exclusive for the purposes presented in this paper. For instance, Wang et al [15] explored the relationship of the IoT and toys (Internet of toys) in terms of interaction between toys and social networks. In addition, the Disney Research laboratories worked in the Internet of toys through the European CALIPSO project [16], which the main goal was to provide networks of connected smart objects, but their main efforts were on the design of low-consumption and low-latency communication protocols between the objects [17].

Materials

Smart Toys Development

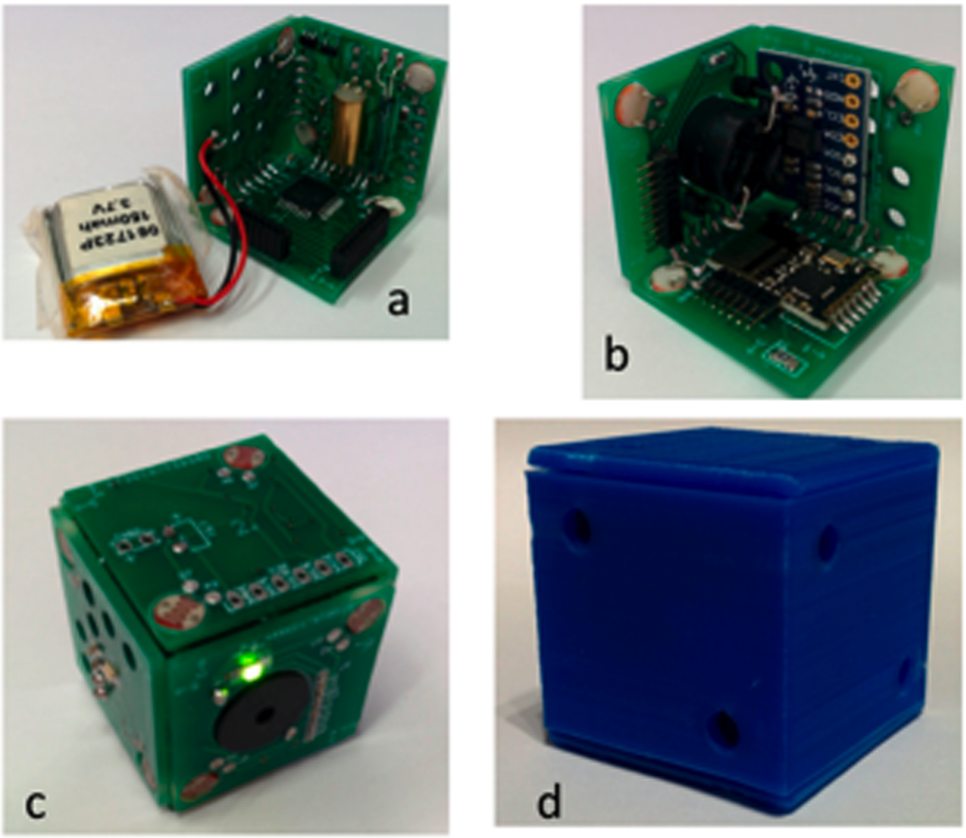
The EDUCERE project aims to create a whole ecosystem of compatible smart toys that provide information that helps child development professionals to detect potential developmental problems. The first toys created for that purpose were a set of cubes that can be stacked to build a tower [4], but other compatible toys, such as pegboards, rattles, or balls, were designed. As a design consideration, we established that the cube must be safe for child interaction. Hence, the cubes cannot open while the toddler plays with them to avoid smaller pieces inside the block from causing harm to the toddler.

All toys include the ATMEL ATmega328p microcontroller by Atmel Corporation (San Jose, CA, USA). This controller is integrated with Arduino boards and is compatible with them, offering an easy and fast prototyping platform to develop the toy software through the Arduino integrated drive electronics. Each toy includes the sensors that add the needed functionality in each case. Every toy uses a NRF24 radio frequency adapter for communications with a data collector device. The NRF24L01+ by Nordic Semiconductor ASA (Oslo, Norway) is an ATmega328p-compatible 2.4 GHz radio frequency transmitter/receiver chosen because of its size, low cost, and low consumption.

Specifically, the stackable cubes consist of the preceding components and:

1. A 3.7 lithium-ion polymer battery. The main limitation for the battery was the maximum size of the cubes (2.5 cm per side), so 150 milliamp hour was the maximum capacity to be fitted in the available space. In addition, a protection circuit was included to ensure that the battery never offers less than 3.2 volts.
2. A set of 12 (two per cube face) light-dependent resistor sensors from Silonex Inc (Montreal, QC, Canada). These sensors allow for determining which face of the cube is covered at each moment.
3. A MPU-9150 InvenSense by Sunnyvale (San Jose, CA, USA) with a three-axis accelerometer and gyroscope. This sensor provides speed, acceleration, and shaking level values for each cube movement.
4. A tilt-based switch that enables a sleep mode to decrease power consumption.
5. Three light-emitting diodes (LEDs) and a buzzer for a visual and auditory user interface.

Figure 1. The cube printed circuit board construction (a-c) and the 3D printed external case (d).



XSL•FO
RenderX

The very strict size constraint in cube size (no more than 2.5 cm for each face) required designing a specific printed circuit board (PCB) shaped as a cube itself. The PCB is divided into six square faces and each face contains part of the printed circuit. Figure 1 shows how the faces are welded together at 90 degrees creating two pieces that are assembled together as a cube. Finally, a plastic 3D printed case was designed to cover the circuit board. A prototype set of 10 cubes was built for the laboratory experiments initially, but more cubes were assembled allowing more experiments to be performed in parallel in different schools.

Cubes transmit the gathered data to a collector module that gives format and stores and encrypts the information. The collector module has been deployed in a Raspberry Pi board with a NRF24L01+ adapter that allows it to connect by itself to the toys. The collector provides a RESTful application program interface through a Wi-Fi access point that allows for managing the experiments and the data files obtained in these experiments using a native app in a tablet or a phone. The data files are cyphered using Advanced Encryption Standard (AES) on Cipher Block Chaining (CBC) mode with a one-use 128-bit key. The key is randomly generated for each experiment and is also encrypted with the RSA DDSS public key to ensure the information is only accessible in a secure Web server.

The EDUCERE Developmental Delay Screening System

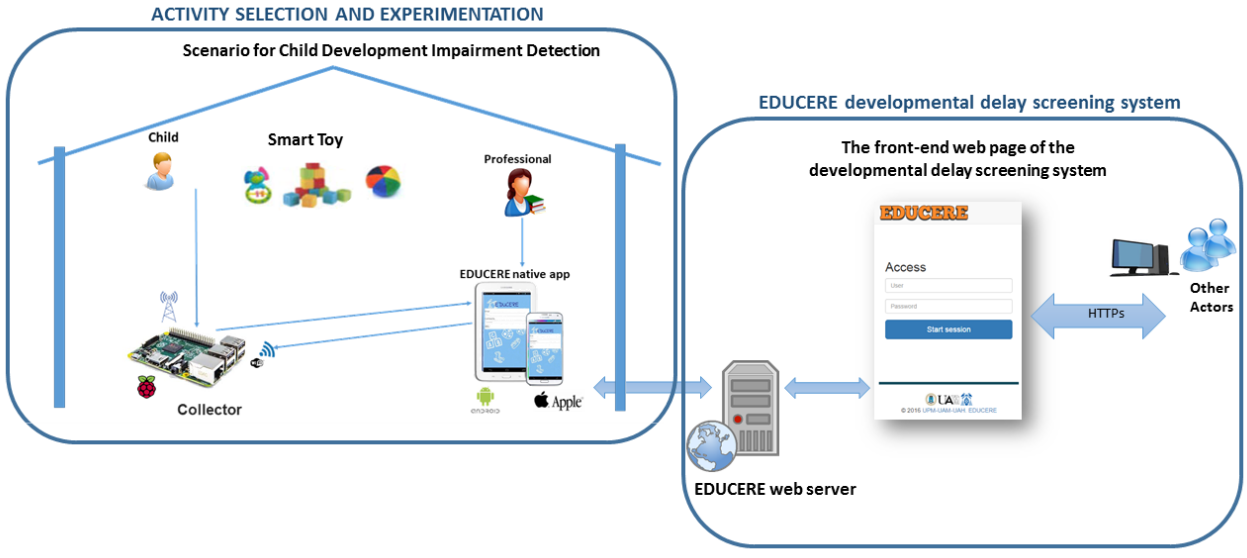
Figure 2 shows the general component architecture of the EDUCERE DDSS. The left side (activity selection and experimentation) shows the necessary elements for the experiment: the child playing with the toy, the smart toy, a professional to assist the child, a Raspberry Pi (used as collector to obtain and save the received data from the smart toy with the rest of experiment information), and a tablet with a mobile app to interact with the collector (for starting, finalizing, repeating, and storing the experiments). The right part (EDUCERE DDSS) shows the components for registry, consultation, and modification of information about children, professionals, and experiments. Professionals can also perform the analysis of the results obtained by the children with the smart toys interaction.

The DDSS of EDUCERE contains all administrative tasks to securely register children, professionals, and activities used during the experimentation scenario. The registry process is done in two steps:

1. Before beginning the process of experimentation with the explicit smart toy, a user with an administrative role will access the EDUCERE system to store the following information: the specific smart toy for the experimentation, the activities that can be performed with that smart toy, and the professional who will carry out the experiment.
2. After that, the administrative user reports to the professional so that they can begin the experimentation process. The professional must be authenticated in the EDUCERE DDSS and enter the information requested for each of the children who will be performing the experiment.

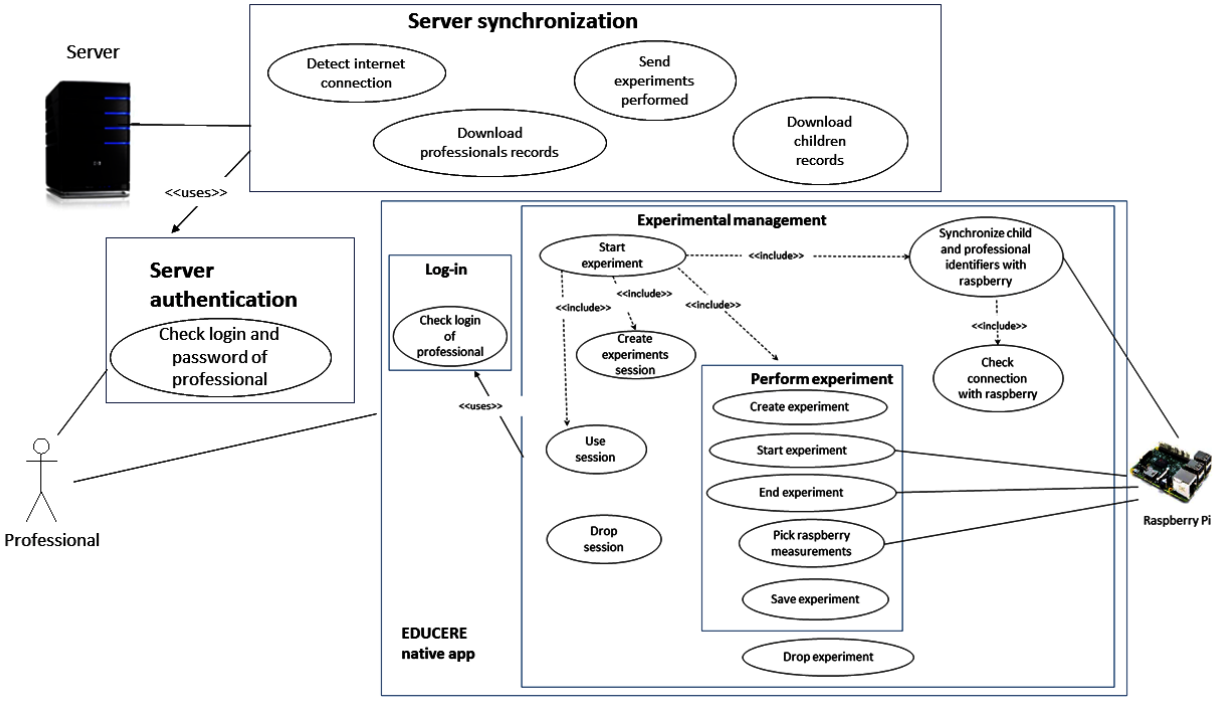
The person with the professional role must upload information about the children who will carry out the experiment with the smart toy. The professional enters name, date of birth, gender, the name of the professional who will carry out the experiment, and the name of the center where this experiment will be done. The professional must also record in EDUCERE the information about the activities that the child could do with the smart toy.

Figure 2. Components of the general architecture in EDUCERE developmental delay screening system.



XSL•FO
RenderX

Figure 3. EDUCERE mobile app use case diagram.



Several children participate at each screening session. In addition, the experimentation session can last for several days with the whole group. Figure 3 shows the use case diagram with the main functionality of the mobile app developed for the tablet.

Use cases of the mobile app consist of three main stages:

1. Experimental management: includes all use cases related to the experimentation process. Users log in to the mobile app to access, such as a known professional. In this case, only the log-in of the professional is required. Before starting experiments, the mobile app (previously connected to the EDUCERE Wi-Fi generated by the Raspberry Pi collector) synchronizes professionals and children with the collector. To start experiments, it is required to create or use an experiment session because all experiments will belong to a concrete session. The professional could also drop a session when the session is over. Once the professional has selected a session, the professional should configure the experiment, choosing a registered child and the cube activity. The experiment starts when the professional presses the "start experiment" button. The experiment stops when he or she presses the "stop experiment" button. Next, the professional can refuse it or store it. If the professional chooses to store it, the collector sends all data collected during the experiment in an encrypted file form. This file is saved in the internal storage of the mobile or tablet.
2. Server synchronization: there are two ways of synchronization between the mobile app and the EDUCERE DDSS server. Before the experiment starts, the mobile app connects with the EDUCERE DDSS server to download information about registered children and professionals. The mobile app uses this information to synchronize the professionals and children with the collector. After performing experiments, the mobile app connects with the EDUCERE DDSS server to upload experiment files with the collected measurements to the server.
3. Server authentication: to synchronize the mobile app with the EDUCERE DDSS server, the professional must be authenticated in the app (submitting log-in and password) to access to the server. This log-in and password are sent to the server to start synchronization.

The mobile app has been implemented in HTML5-Javascript for cross-platform development, using jQuery and the responsive Web app development framework Bootstrap. Through Cordova, a well-known open source mobile development framework, the wrapper is generated to run the app on different mobile platforms, such as Android or iOS. The result is a hybrid app executed via Web views. Cordova also provides a series of plug-ins to access the functionalities of the mobile device, such as internal storage, connection detection, and vibration, all of which are necessary for the development of the mobile app.

Figure 4 shows an example of two interfaces of the mobile EDUCERE app. Part A displays the interface to start a concrete experiment with a child by using one of the smart toys developed. Through this user interface, professionals can choose the child and the type of experiment. The "start experiment" button sends the signal "start" to the collector to start storing data from the experiment. Part B contains information of all performed experiments by the professional. This screen includes the button "send all experiments to the server" to upload json experiment files to the EDUCERE DDSS.

XSL•FO
RenderX

Figure 4. EDUCERE mobile app interfaces (in Spanish). The professional can select the toy for experiment (a) and show data from experiments performed (b).

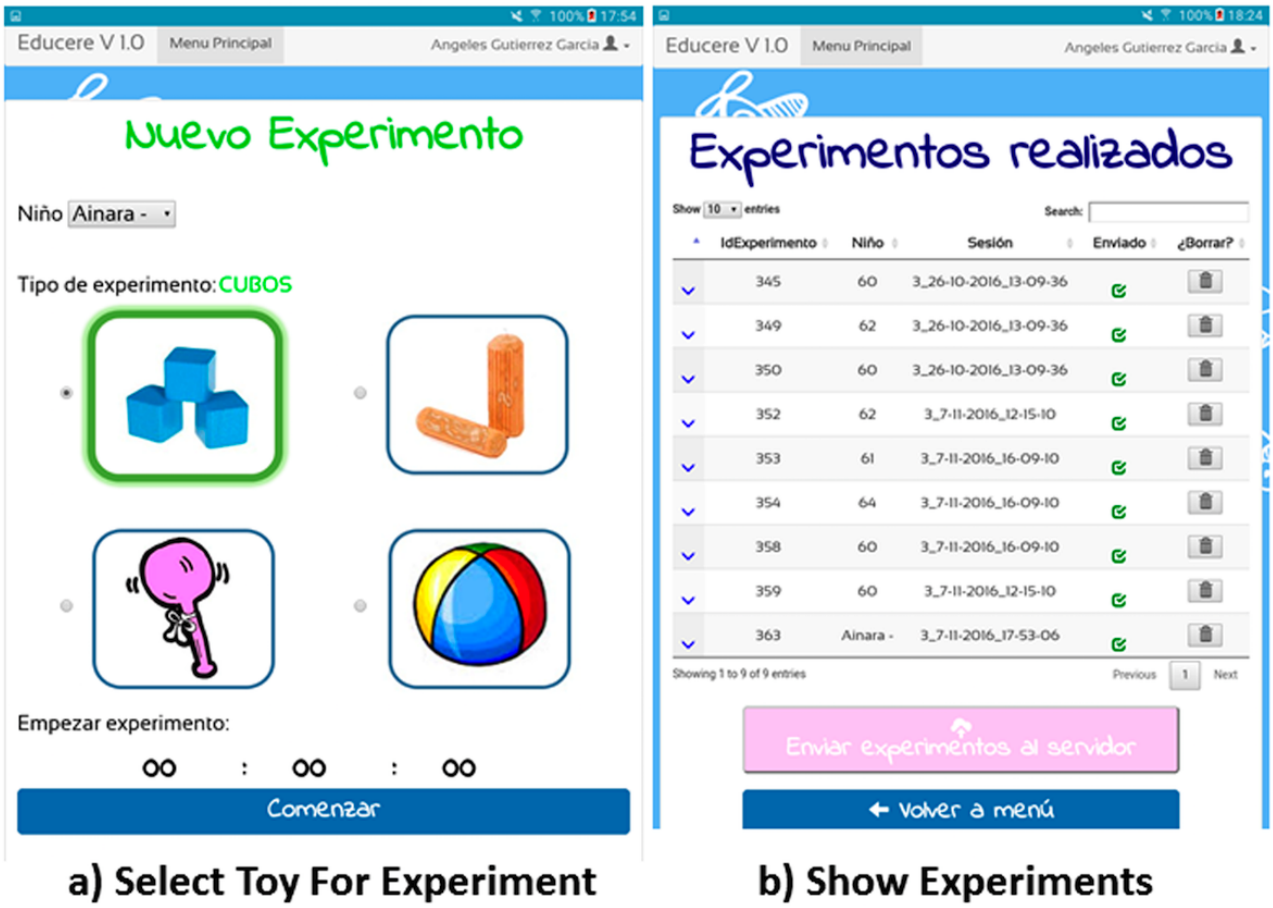
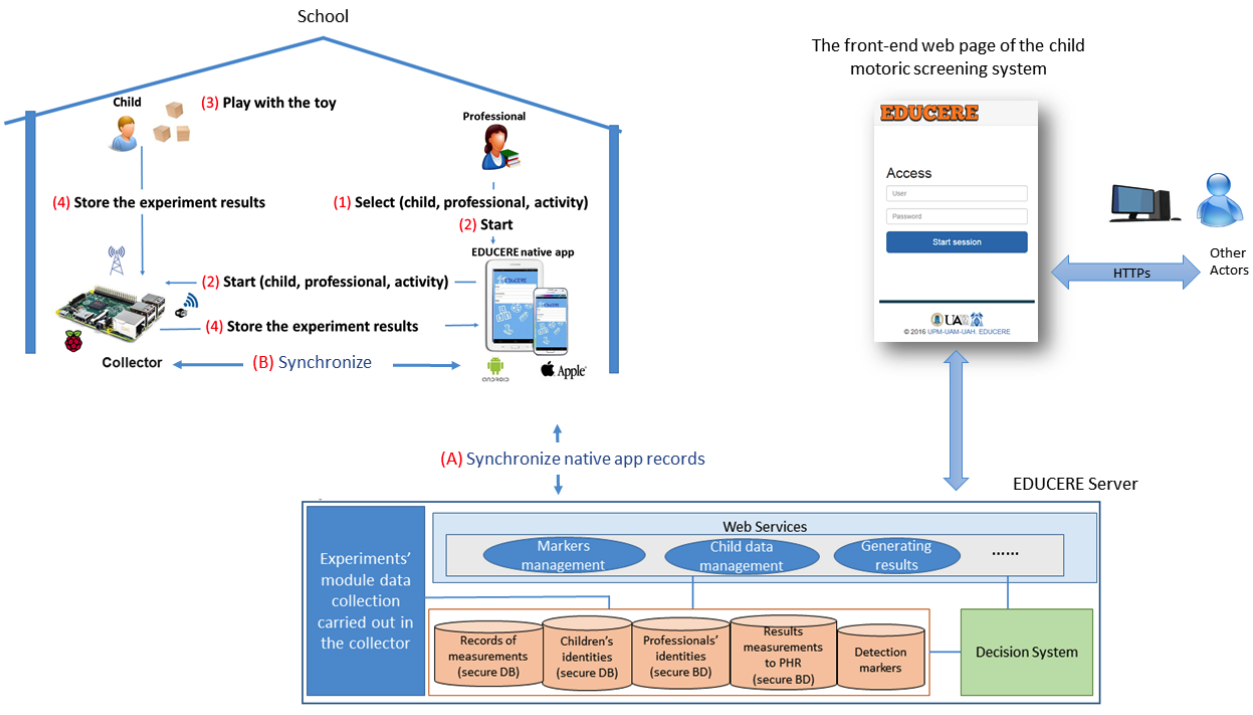


Figure 5. Steps for interaction with the EDUCERE developmental delay screening system.



XSL•FO
RenderX

During the experimentation scenario, the records of the activities performed by the child with the smart toy are collected and sent to the EDUCERE DDSS to be persistently stored (see Figure 5). In addition, the EDUCERE DDSS checks the results obtained by each child with the smart toy. Thus, the information stored in the EDUCERE system can provide professionals with useful information about early detection of a child's motor difficulties.

Both in the school or home scenario, the child registered in the EDUCERE DDSS performs the activity indicated by the professional with the specific smart toy (rattles, balls, and cube towers). In this task, the professional assesses the child during the experimentation process and takes responsibility for the experiment. He/she uses a tablet to perform the process of starting the experiment, finalizing the experiment, and storing it. This is the way in which communication occurs between the professional and the collector, who manages the activity to be performed with the smart toy.

As Figure 5 outlines, the professional should take the following steps to guide the child in the interaction process with the game:

1. When opening the app, the displayed home screen helps to identify the professional who will perform the experiment. Once the professional provides identification, he or she selects the child who will perform the experiment. After that, the professional chooses the cube tower activity, which is the activity to be performed in which the child plays with a set of cubes.
2. Start experiment. At this point, it is necessary to synchronize the collector with the identifiers that are in the tablet corresponding to the child and professional (steps A and B in Figure 5). Then, the professional chooses a session already created or he/she starts with the creation of a new session. This distinction is necessary because the experimentation process with a set of children can be done over several days and it may be necessary that all the experiments belong to the same session, although they take place on different dates.
3. After that, the child begins interacting with the toy and the information generated by the toy's sensors is stored in the collector. When the child completes the activity with the smart toy, the professional indicates on the tablet that the experiment has ended.
4. At this step, the experiment data in the collector are transmitted to the tablet and securely stored.

Finally, when the mobile app detects a known Wi-Fi, it connects to it to synchronize with the EDUCERE DDSS server (always prevails over all Wi-Fi raspberry, to which you must connect to start the experiment). The tablet connects to a specific Web service in the server to synchronize and transfer information between the experiments stored in the tablet and the EDUCERE system databases on the server (step A in Figure 5). The EDUCERE databases contain information that identifies registered professionals in the system and the children who have been discharged by each professional, and data from all experiments uploaded by the mobile app.

Methods

Participants

A total of 65 toddlers (32 boys and 33 girls) from a public nursery school aged between 23 and 37 months (mean 29.02, SD 3.81 months) took part in a pilot trial in which they had to build a tower with five stackable cubes. The professional did not have clinical information about the toddlers who participated in the experiment. Parents signed an informed consent sent by EDUCERE project.

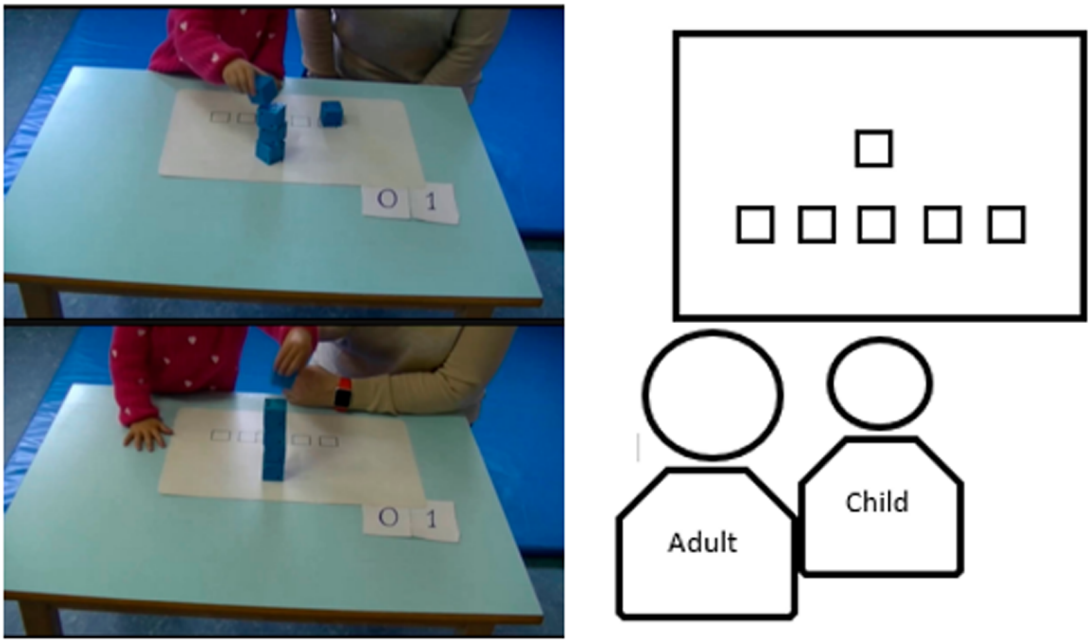
Apparatus and Materials

The toddlers were given five EDUCERE stackable cubes placed in a row on a template (Figure 6). Thus, the initial positions of the smart cubes and the place to build the tower were marked to align the trial.

The observer sat to the left of the child who was in front of the table in the middle. During the experiment, the experimenter needed the following elements: (1) a collector module (described in Smart Toys Development subsection) to format, store, encrypt, and transmit information obtained from the cubes; (2) a tablet with a mobile app (as shown Figure 4) to interact with the collector for starting, finalizing, repeating, and storing the experiments; and (3) a video camera to record all trials for later viewing.

XSL•FO
RenderX

Figure 6. Experimental scenario. Initially the five cubes are placed in a row on the template and the child is to build a tower with the cubes on the square in front.



Procedure

The set up of the trials was fixed on the tables used by the toddlers. The professional made the tower with the stackable cubes in the marked position while saying, “Look, I’m going to make a tower with these cubes right here” (see Multimedia Appendix 1, Table S1). After a few seconds and making sure that the toddler looked at the tower, the professional put the cubes back in their initial positions. Then the adult asked the child to make the tower (“Now I would like you to make a tower just like I did”) and waited for the child to make the tower. If the child dropped the pieces, the professional suggested putting them back on the tower. Children could make as many attempts as they wanted and all data were recorded but, once the video was visualized, the first attempt was selected for analysis. Finally, all toddlers received a sticker as a reward.

Four experts in child psychomotor development (one developmental psychologist, one physiotherapist, and two educators) viewed the recordings and selected an analyzable fragment and scored each child trial on a scale of 1 to 10 according to their performance. The experts had a meeting prior to viewing the videos in order to agree on assessment criteria (see Multimedia Appendix 1, Table S2).

Each expert randomly viewed half of the videos and two experts individually rated each video. In this way, two experts scored each child’s performance. The experts made their assessments without having contact with one another to ensure their independence of judgment. Because the experts came from different professional backgrounds, it was expected that their assessments would not be identical, despite having reached agreement on the criteria. Multimedia Appendix 1 (Table S3) shows the professional profiles of the four experts.

The selected video fragments had their corresponding data recorded by the collector module. These data were included, together with the expert scores, in the statistical analysis.

Design

The set of variables for each experiment were measured by using the sensors included in the toys (see Smart Toys Development subsection). The toy processes these variables starting from the values obtained from the sensors and stores them for each movement. The variables are the maximum acceleration during the movement, the maximum and mean speed, the time at which the maximum speed is reached within the movement, the time the movement took, and the number of shakings detected. The shakings are calculated from the graph of instantaneous accelerations by considering a shaking as data between two minimum values. There were four levels of shakings. In order to determine the level of shaking, the number of samples that fit the “mound” in the accelerations graph was taken into account. Using this classification scheme, the first level represented the smaller shakings (ie, the shorter “mounds”, where only one sample from the sensor was received before and after the maximum acceleration value) and the fourth level represented the bigger shakings (where four or more samples were received). This measurement and classification has been explained in more detail previously [4].

Once all the experiments were performed, the stored per-movement data were summarized in a per-experiment data file. The variables used in the analysis were these summarized values, including the performance scores determined by experts, which are detailed in Table 1.

XSL•FO
RenderX

Table 1. Summary of variables used in the analysis.

Variable name	Meaning	Dimensions/Range	How it is calculated?
Performance scores	Scores of children performing the activity	1-10 (10 being the best possible score)	It is assigned by experts while reviewing the experiment
Number of movements	Total number of movements made with all the cubes during an experiment	1-n (ideally five, one movement per cube)	A movement is any period of time in which the cube accelerometer sensor gives (after velocity calculation) a value high enough to determine the cube is moving (see [4] for a detailed explanation of the calculations)
Mean time of movement	Mean of the duration of each movement during an experiment	Milliseconds (msec)	The period of each movement is detected and stored and then the mean value of all these time values is calculated
Mean speed of movement	Mean of all the mean speed values measured during an experiment in meters per second.	Meters per second (m/s)	The speed values during a movement are calculated by integrating the values obtained by the cube accelerometer; with all the instant values within a movement, the mean speed is calculated and this value is the mean of these means for all the experiment
Mean of maximum speed	Mean of all the maximum speed values	m/s	For each speed value obtained during a movement, the maximum value is stored, then the mean of these values is calculated for the entire experiment
Highest maximum speed	The maximum value of the maximum speeds	m/s	For all the maximum values stored during an experiment, the maximum value is selected
Lowest maximum speed	The minimum value of the minimum speeds	m/s	For all the maximum values stored during an experiment, the minimum value is selected
Maximum acceleration of movement	Mean of the maximum acceleration values	m/s^2	The accelerations are calculated directly from the values obtained in the accelerometer; the maximum value obtained for a movement is stored and, for this variable, the mean of these maximum values is calculated
Highest maximum acceleration	The maximum value of the maximum accelerations	m/s^2	This variable represents the highest value of the maximum accelerations stored during an experiment
Lowest maximum acceleration	The minimum value of the maximum accelerations	m/s^2	This variable represents the lowest value of the maximum accelerations stored during an experiment
Mean of shaking (level 1)[a]	Mean of the number of shaking of level 1	1-n	Given the previous definition of shaking, this variable represents the mean of the level 1 shakings measured for each movement
Mean of shaking (level 2)[a]	Mean of the number of shaking of level 2	1-n	Given the previous definition of shaking, this variable represents the mean of the level 2 shakings measured for each movement
Mean of shaking (level 3)[a]	Mean of the number of shaking of level 3	1-n	Given the previous definition of shaking, this variable represents the mean of the level 3 shakings measured for each movement
Mean of shaking (level 4)[a]	Mean of the number of shaking of level 4	1-n	Given the previous definition of shaking, this variable represents the mean of the level 4 shakings measured for each movement

[a] The first level represents the smaller shakings (ie, the shorter “mounds”, where only one sample from the sensor is received before and after a maximum acceleration value) and the fourth level represents the bigger shakings (where four or more samples are received).

Results

In order to analyze interrater agreement, a reliability analysis was conducted in IBM SPSS Statistics 23. We computed a intraclass correlation coefficient (ICC) analysis, following the Model ICC (1,k) in SPSS, in which 1 denotes that each participant is assessed by a different set of randomly selected raters, and k is the number of raters for every score. In this experiment, two experts out of four were randomly assigned half of the videos to rate. The other half of the videos were rated by the two remaining experts. This way, each of 65 videos had two ratings.

XSL•FO
RenderX

In the model, reliability was calculated by taking the mean of the two raters' measurements across the 65 scores. The ICC for single measures was 0.961 (95% CI 0.937-0.976; $F_{64,64}$=50.39, P<.001) and the ICC for mean measures was 0.980 (95% CI 0.967-0.988; $F_{64,64}$=50.39, P<.001). The ICC of 0.980 for the mean measures indicates that 98% of the variance in the mean of these raters was "real." The 95% CI (0.967-0.988) suggests that the process was successful to separate different levels of performance. Because the reliability among the judges' assessment was high, we used the mean of the expert's scores as a variable of performance in the subsequent data analysis.

A factor analysis was conducted to group similar variables into dimensions [18]. This analysis does not distinguish between independent and dependent variables, but it was useful to reduce the number of variables in the predictive regression model. We needed a predictive model to build an automatic system to support detection of developmental delays. This approach was effective for redesigning the initial prototype to be more efficient by, for example, reducing the number of sensors that focus on collecting the main relevant data.

The factor analysis showed that the first three factors together accounted for 76.784% of the total variance. Table 2 includes the rotated factor loadings, which represent both how the variables were weighted for each factor, but also the correlation between the variables and the factor. The extraction method was principal axis factoring and the rotation method was varimax with Kaisser normalization.

Table 2. Rotated component matrix.

Variance and variables	Component		
	1	2	3
Variance explained	31.386%	24.788%	20.616%*
Variable, correlation estimate			
Number of movements	–.049	.294	–.782*
Mean time of movement (msec)	.983*	–.048	.003
Mean speed of movement (m/s)	.015	.840*	.199
Mean of max speed (m/s)	–.024	.943*	–.009
Highest maximum speed (m/s)	–.078	.723*	–.572
Lowest maximum speed (m/s)	.035	.264	.800*
Maximum acceleration of movement	–.083	.809*	.139
Highest maximum acceleration	–.090	.642*	–.597
Lowest maximum acceleration	–.044	.447	.784*
Mean of shaking 1	.747*	–.021	–.229
Mean of shaking 2	.896*	–.120	–.047
Mean of shaking 3	.892*	–.024	.173
Mean of shaking 4	.728*	.022	.225

* Strongest correlations between variables and components (factors). Those in component 1 make up "trembling" factor, those in component 2 make up "speed" factor, and those in component 3 make up "accuracy" factor.

In Table 2, the most important correlations between variables and components (factors) are marked. We assigned a name to each of these factors to represent the variables that are part of them. Based on factor loadings, we think the factors represent the following concepts:

1. Component 1 presents high correlations with the variables mean time of movement and mean of shaking (1, 2, 3, and 4). We call this factor "trembling."
2. Component 2 indicated high correlations with the variables mean speed of movement, mean maximum speed, highest maximum speed, maximum acceleration of movement, and highest maximum acceleration. We call this factor "speed."
3. Component 3 links high correlations with the variables number of movements, lowest maximum speed, and lowest maximum acceleration. However, the correlation with number of movements was negative, as can be observed. This means that the number of movements varies in the opposite direction to that of the other significant variables of the component and, of course, opposite the factor. We call this factor "accuracy."

In order to design the EDUCERE automatic system for the detection of delays in toddlers' psychomotor development using the smart stackable cubes, it was necessary to describe an algorithm that included the significant factors, or independent variables, and a dependent variable "performance."

Two multiple regression analyses were carried out to predict (1) the value of the variable performance based on the value of the three components obtained in the factor analysis, trembling, speed, and accuracy, and (2) the value of the variable "age" based on the same three components.

Table 3 presents the model summaries for all multiple regression analyses. Table 4 presents the coefficients for all multiple regression analyses.

XSL•FO
RenderX

Table 3. Multiple regression analyses: model summary.

Model	Dependent	Predictor	R	R^2	Adjusted R^2	SE of the estimate
1	Performance	Accuracy, speed, trembling	.517	.267	.231	1.556
2	Age (months)	Accuracy, speed, trembling	.362	.131	.089	3.637

Table 4. Multiple regression analyses: coefficients.

Model	Unstandardized coefficient, B (SE)	Standardized coefficient, beta	t_{61}	P
1				
(Constant)	7.662 (0.193)		39.698	<.001
Trembling	0.050 (0.194)	0.028	0.257	.80
Speed	–0.630 (0.194)	–0.355	–3.239	.002
Accuracy	0.665 (0.194)	0.375	3.419	.001
2				
(Constant)	29.015 (0.451)		64.315	<.001
Trembling	–0.152 (0.455)	–0.040	–0.334	.74
Speed	0.003 (0.455)	0.001	0.007	.99
Accuracy	1.372 (0.455)	0.360	3.018	.004

Based on Table 3, the equation for the regression line in model 1 was: performance = 7.662 + 0.05(trembling) – 0.630(speed) + 0.665(accuracy), with the standard error of the estimate=1.556.

The coefficient for trembling (0.50) was not significantly different from zero (P=.80), but the coefficient was positive, which would indicate that higher trembling is related to better performance (not what we would expect). The coefficient for accuracy was positive and significantly different from zero, which means with higher accuracy there is better performance, as expected. Conversely, the coefficient for speed was significantly negative, which means that at higher speed there is worse performance.

The equation for the regression line in model 2 was: age = 29.015 – 0.152(trembling) + 0.03(speed) + 1.372(accuracy), with the standard error of the estimate=3.637.

Only the accuracy coefficient (1.372) was significant (P=.004). The coefficients for trembling (P=.74) and speed (P=.99) were not significantly different from zero. The accuracy coefficient was positive, which would indicate that higher accuracy was related to age (what we would expect).

Discussion

The regression equation obtained for the variable “performance” is the algorithm that will be the basis of the automatic detection of developmental delays. In order to obtain a design as efficient as possible, the design of the smart toy must be adjusted by reducing the amount of data from the collector module taking into account the nonsignificant results obtained in the statistical analyses.

From the factor analysis, we conclude that trembling explains the greater percentage of the variance (31.38%), but considering it as a possible predictor of the performance in the regression analysis, the results show a lack of significance, so variance is unrelated to performance. This implies that the sensors that provide measurements for the variables mean time of movement and mean of shaking (1, 2, 3, and 4) have to be reconfigured to obtain only the data of interest, those that the automatic system needs to classify toddlers’ psychomotor performances. Although the prediction power of this reduced set would be slightly lower than the obtained using the original one, the reduction is probably not enough to justify maintaining the complex design of the devices.

On the other hand, speed accounts for 24.78% of the variance, but it is also negatively related to performance. For this reason, the sensors that allow the collector module to obtain measurements of the mean speed of movement, the mean maximum speed, the highest maximum speed, the maximum acceleration of movement, and the highest maximum acceleration, must be kept in the smart toy because they were in the original design [4].

Finally, accuracy accounts for 20.61% of variance and is a predictor of performance, so sensors that collect data to measure the number of movements, the lowest maximum speed, and the lowest maximum acceleration should also be kept in the smart toy because they were in the original design [4].

Chronological age is not a direct indicator of the level of psychomotor development, although it is related to it. According to the results obtained in the regression analysis, the only factor that predicts children’s chronological age is accuracy. This result agrees with those of a study on dysgraphia [19] in which it was concluded that poor writers were less accurate.

In this research, we have detected at least three factors of interest to describe the level of psychomotor development of children: trembling, speed, and accuracy. This will be the starting point for further research that will focus on exploring the relationships

XSL•FO
RenderX

of these factors to a set of motor behaviors of children in their natural settings, school, home, and playground [19,20].

The system is not designed to predict the age of the children, but it was of theoretical interest to know how age is related to the other variables observed. However, it is relevant to know that the variable accuracy, as described in this research, is related to the level of psychomotor development of children and is one of the aspects to be observed in predicting possible difficulties in school, such as dysgraphia [19].

The EDUCERE DDSS could benefit from larger sample sizes to "learn" to detect and classify delays of psychomotor development. Therefore, the following research work for the smart toy should be tested with more toddlers, with and without developmental delays, diagnosed or not. Furthermore, the data obtained from these new experiments will be used for a further validation of the results of the analysis presented in this paper because this new set of data could guarantee that the predictive power of the algorithm stays the same when separating training and test data.

Consequently, the next phase of the research will focus on the objective to establish the criteria for classification of psychomotor development delays and to describe the actions to be performed in each case, relying on the previously validated smart toys. These criteria do not correspond exactly to the conventional diagnostic criteria because the interest of this investigation is to detect slight delays that are usually unnoticed in the standardized tests, which is the added value of this research.

The final mission of the EDUCERE DDSS is starting to be achieved by providing parents, educators, psychologists, and pediatricians with accurate data about potential delays detected. Results obtained initially triggered some advisement to follow-up in some cases leading to messages such as, "the movements and child´s interaction are OK" or "let the child keep playing but visit the specialist in 3 months."

Acknowledgments

This paper is part of research conducted under the EDUCERE project (Ubiquitous Detection Ecosystem to Care and Early Stimulation for Children with Developmental Disorders; TIN2013-47803-C2-1-R), supported by the Ministry of Education and Science of Spain through the National Plan for R+D+I (research, development, and innovation). Besides, we would like to thank the aid received for PhD students contracts of the Ministry of Economy of Spain (BES-2014-067912).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptions relevant to the experimental procedure.

[PDF File (Adobe PDF File), 27KB - jmir_v19i5e171_app1.pdf]

References

1. Roid G, Sampers JL. Merrill-Palmer-Revised Scales of Development. Wood Dale, IL: Stoelting Company; 2004.
2. Valero MA, Lindén M, Velasco JR, Björkman M. Personalisation of intelligent homecare services adapted to children with motor impairments. In: Proceedings of the 8th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAml). Personalisation and User Adapted Services Lecture Notes in Computer Science. 2014 Presented at: 8th International Conference on Ubiquitous Computing and Ambient Intelligence; Dec 2-5, 2014; Belfast, UK p. 476-479. [doi: 10.1007/978-3-319-13102-3_76]
3. Nunez-nagy S, Alarcos B, Valero M, Gallego T, Ramirez R, Serrano C. Smart toys for detecting children development delays. 2015 Presented at: Proceedings of the International Conference in Recent Advances in Neurorehabilitation (ICRAN); Jun ; Valencia, Spain; 2015; Valencia p. 9-12. [doi: 10.13140/RG.2.1.4392.520]
4. Rivera D, García A, Alarcos B, Velasco JR, Ortega JE, Martínez-Yelmo I. Smart toys designed for detecting developmental delays. Sensors (Basel) 2016 Nov 20;16(11):1-22 [FREE Full text] [doi: 10.3390/s16111953] [Medline: 27879626]
5. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics 2006 Jul;118(1):405-420. [doi: 10.1542/peds.2006-1231] [Medline: 16818591]
6. Cools W, Martelaer KD, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. J Sports Sci Med 2009 Jun 01;8(2):154-168 [FREE Full text] [Medline: 24149522]
7. Bayley N, Reuner G. Bayley Scales of Infant and Toddler Development. San Antonio, TX: Harcourt Assessment Psych Corporation; 2006.

XSL•FO
RenderX

8. Focaroli V, Taffoni F, Parsons SM, Keller F, Iverson JM. Performance of motor sequences in children at heightened vs low risk for ASD: a longitudinal study from 18 to 36 months of age. Front Psychol 2016;7:724 [FREE Full text] [doi: 10.3389/fpsyg.2016.00724] [Medline: 27242630]
9. Mari M, Castiello U, Marks D, Marraffa C, Prior M. The reach-to-grasp movement in children with autism spectrum disorder. Philos Trans R Soc Lond B Biol Sci 2003 Feb 28;358(1430):393-403 [FREE Full text] [doi: 10.1098/rstb.2002.1205] [Medline: 12639336]
10. Forti S, Valli A, Perego P, Nobile M, Crippa A, Molteni M. Motor planning and control in autism. A kinematic analysis of preschool children. Res Autism Spect Dis 2011 Apr;5(2):834-842. [doi: 10.1016/j.rasd.2010.09.013]
11. Martin-Ruiz ML. Foundations of a smart toy development for the early detection of motoric impairments at childhood. Int J Pediatr Res 2015 Dec 31;1(2):1-5. [doi: 10.23937/2469-5769/1510011]
12. Marschollek M, Gietzelt M, Schulze M, Kohlmann M, Song B, Wolf K. Wearable sensors in healthcare and sensor-enhanced health information systems: all our tomorrows? Healthc Inform Res 2012;18(2):97. [doi: 10.4258/hir.2012.18.2.97]
13. Albinali F, Goodwin MS, Intille S. Detecting stereotypical motor movements in the classroom using accelerometry and pattern recognition algorithms. Pervasive Mob Computing 2012 Feb;8(1):103-114. [doi: 10.1016/j.pmcj.2011.04.006]
14. Taffoni F, Focaroli V, Keller F, Iverson JM. A technological approach to studying motor planning ability in children at high risk for ASD. In: Conf Proc IEEE Eng Med Biol Soc. 2014 Presented at: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE; Aug 26-30, 2014; Chicago, IL p. 3638-3641. [doi: 10.1109/EMBC.2014.6944411]
15. Wang W, Kuo V, King C, Chang C. Internet of toys: an e-Pet overview and proposed innovative social toy service platform. : Proceedings of the IEEE International Computer Symposium (ICS); 2010 Presented at: Computer Symposium (ICS), 2010 International; Dec 16-18, 2010; Tainan, Taiwan p. 265-269. [doi: 10.1109/COMPSYM.2010.5685507]
16. Calipso Project. European Community's 7th Framework Programme: CALIPSO-Connect All IP-based Smart Objects URL: http://www.ict-calipso.eu/ [accessed 2017-02-16] [WebCite Cache ID 6oJjAQulO]
17. Vukadinovic V, Glaropoulos I, Mangold S. Enhanced power saving mode for low-latency communication in multi-hop 802.11 networks. Ad Hoc Networks 2014 Dec;23:18-33. [doi: 10.1016/j.adhoc.2014.06.001]
18. Tucker L, MacCallum R. Exploratory Factor Analysis. 1993. URL: http://www.unc.edu/~rcm/book/factornew.htm [accessed 2017-02-16] [WebCite Cache ID 6oJhRov3z]
19. Smits-Engelsman B, Van Galen G. Dysgraphia in children: lasting psychomotor deficiency or transient developmental delay? J Exp Child Psychol 1997:164-184. [doi: 10.1006/jecp.1997.2400]
20. Riquelme Agulló I, Manzanal González B. Factors influencing motor development in children with Down syndrome. Int Med Rev Down Syndr 2006 Jul;10(2):18-24. [doi: 10.1016/S2171-9748(06)70063-0]

Abbreviations

DDSS: developmental delay screening system
EDUCERE: Ecosistema de Detección Ubicua, atenCión y Estimulación tempRana para niños con trastornos del dEsarrollo [Ubiquitous Detection Ecosystem to Care and Early Stimulation for Children with Developmental Disorders]
ICC: intraclass correlation coefficient
IoT: Internet of things
PCB: printed circuit board

Edited by G Eysenbach; submitted 20.02.17; peer-reviewed by V Traver Salcedo, C Costa, M Drobics; comments to author 21.03.17; revised version received 23.03.17; accepted 09.04.17; published 19.05.17.

Please cite as:
Gutiérrez García MA, Martín Ruiz ML, Rivera D, Vadillo L, Valero Duboy MA
A Smart Toy to Enhance the Decision-Making Process at Children's Psychomotor Delay Screenings: A Pilot Study
J Med Internet Res 2017;19(5):e171
URL: http://www.jmir.org/2017/5/e171/
doi:10.2196/jmir.7533
PMID:28526666

©María Angeles Gutiérrez García, María Luisa Martín Ruiz, Diego Rivera, Laura Vadillo, Miguel Angel Valero Duboy. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 19.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of

XSL•FO
RenderX

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

XSL•FO
RenderX

Viewpoint

Noncommunicable Disease Risk Factors and Mobile Phones: A Proposed Research Agenda

Adnan A Hyder[1,2], MPH, MD, PhD; Adaeze C Wosu[3], MPH; Dustin G Gibson[1], MS, PhD; Alain B Labrique[1], MHS, MS, PhD; Joseph Ali[1,2], JD; George W Pariyo[1], MBChB, PhD

[1]Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD, United States

[2]Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, United States

[3]Johns Hopkins Bloomberg School of Public Health, Department of Epidemiology, Baltimore, MD, United States

Corresponding Author:
Adnan A Hyder, MPH, MD, PhD
Johns Hopkins Bloomberg School of Public Health
Department of International Health
615 North Wolfe Street, Suite E8132
Baltimore, MD, 21205
United States
Phone: 1 410 502 8947
Fax: 1 410 614 1419
Email: ahyder1@jhu.edu

Abstract

Noncommunicable diseases (NCDs) account for two-thirds of all deaths globally, with 75% of these occurring in low- and middle-income countries (LMICs). Many LMICs seek cost-effective methods to obtain timely and quality NCD risk factor data that could inform resource allocation, policy development, and assist evaluation of NCD trends over time. Over the last decade, there has been a proliferation of mobile phone ownership and access in LMICs, which, if properly harnessed, has great potential to support risk factor data collection. As a supplement to traditional face-to-face surveys, the ubiquity of phone ownership has made large proportions of most populations reachable through cellular networks. However, critical gaps remain in understanding the ways by which mobile phone surveys (MPS) could aid in collection of NCD data in LMICs. Specifically, limited information exists on the optimization of these surveys with regard to incentives and structure, comparative effectiveness of different MPS modalities, and key ethical, legal, and societal issues (ELSI) in the development, conduct, and analysis of these surveys in LMIC settings. We propose a research agenda that could address important knowledge gaps in optimizing MPS for the collection of NCD risk factor data in LMICs and provide an example of a multicountry project where elements of that agenda aim to be integrated over the next two years.

(J Med Internet Res 2017;19(5):e133) doi:10.2196/jmir.7246

KEYWORDS

mHealth; noncommunicable disease; mobile phone; research agenda; survey

Background

The World Health Organization (WHO) estimates that 67% of the 56 million deaths that took place globally in 2012 were due to noncommunicable diseases (NCDs), in particular, cardiovascular diseases (42.6%), cancers (21.7%), chronic respiratory diseases (10.7%), and diabetes (4.0%) [1]. About 75% of all NCD deaths took place in low- and middle-income countries (LMICs) where nearly half of all deaths occurred in persons aged under 70 years [1]. The WHO projects a rise in the number of NCD deaths from 36 million in 2008 to 55 million by 2030 if effective steps are not taken to curb the epidemic [2]. Furthermore, deaths from NCDs for people younger than 70 years old is projected to increase from 10.8 million in 2010 to 15.4 million in 2050 [3]. Four key risk factors responsible for a majority of NCDs are tobacco use, unhealthy diet, sedentary lifestyle, and excessive use of alcohol—all behavioral and largely modifiable [4]. These factors, as well as loss of employment due to NCD-related disability, combined with the long duration and complexity of NCD treatment pose additional challenges to poverty reduction and sustainable development [5].

Key to global efforts to prevent and control NCDs is national surveillance. At the national level, surveillance has three primary

XSL•FO
RenderX

components: exposures (behavioral, physiologic, metabolic, social determinants), outcomes (NCD-specific morbidity and mortality), and health system capacity and response (infrastructure, policies, plans, access to health care, and partnerships) [4]. The WHO STEP-wise approach to NCD risk factor surveillance (STEPS) provides a standardized framework and toolkit for monitoring national levels of major NCD risk factors (exposures) [6]. STEPS consists of three levels of data collection: questionnaires, physical measurements, and biochemical measurements [6]. Similar to other household surveys, the STEPS questionnaires are conducted through face-to-face interviews, thus requiring significant time and resource commitments.

Targets and Indicators

In 2013, the WHO developed an NCD global action plan and monitoring framework, establishing 9 voluntary targets and 25 indicators that have to be achieved to attain a 25% reduction in premature mortality from the four major NCDs (ie, cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes) by 2025 [7]. A summary of these targets and indicators is provided in Table 1. Targets include a 25% relative reduction in the risk of premature death from the four major NCDs, a 30% relative reduction in the prevalence of current tobacco use, and a 30% relative reduction in mean population intake of salt or sodium [1,2,7]. The WHO NCD global action plan also calls for surveillance efforts to incorporate cross-sectorial engagement [1]. Despite substantial progress over the past decade, reliable, high-quality data on the prevalence, incidence, and consequences of NCDs remain less available in some countries [1].

One promising approach increasingly being explored for public health surveillance involves mobile phones. An emergent field, mobile health (mHealth), describes medical and public health activities that leverage the global proliferation of cellular networks and mobile phone ownership (or access) to improve population and clinical health outcomes. Device functionalities such as voice calls and short message services (SMS)—often referred to as text messages—global positioning system (GPS), as well as other wearable and connected devices may be harnessed to provide data and insight on individuals and populations [8,9]. Currently, there are over 7.4 billion wireless subscriptions globally, with the majority (78%) in LMICs [10]. The UN International Telecommunications Union estimates 2016 global phone ownership at 99.7 mobile phone subscriptions per 100 persons, a significant increase compared with 41.7 mobile phone subscriptions per 100 persons a decade ago [10,11]. It is this ubiquitous connectivity to cellular networks that potentially makes large proportions of a population accessible through their mobile phones.

In view of the increasing disease burden, the intersecting need for NCD data in LMICs and the near-universal population access to mobile phones in a growing number of countries presents an opportunity for public health. As cellular technologies leapfrog the need for traditional landline infrastructure, public health programs could conduct remote, population-level surveys by reaching out to community members through their own cell phones. Can such a widely dispersed technology have the potential to assist in the achievement of the WHO NCD global action plan targets? How can mobile phone surveys (MPS) be used for population-level NCD risk factor surveillance? What critical technology, sampling, and methodological questions need to be addressed before leveraging this potential? In this paper, we present the potential for MPS to collect such data, review key research issues, and introduce a multicountry effort that seeks to partly respond to this public health challenge. We draw on a selected review of literature, key global reports, and experience of the authors to propose a research agenda and report on efforts underway to address it. It is believed that this paper will facilitate a global dialog and action to enhance the use of MPS for NCD, and potentially other public health risk factor surveillance.

XSL•FO
RenderX

Table 1. Summary of the World Health Organization (WHO) 9 voluntary targets and 25 indicators for global noncommunicable disease (NCD) monitoring [1].

Element	Targets	Relevant indicators
Morbidity and mortality		
Premature mortality from NCD	25% reduction in premature mortality from NCD	Unconditional probability of death between ages 30 and 70 years from the 4 main NCDs
		Cancer incidence, by type of cancer, per 100,000 population
Risk Factors		
Behavioral risk factors		
Harmful use of alcohol	10% reduction in harmful use of alcohol	Total (recorded, unrecorded) alcohol per capita (in ≥15 years old) consumption within a calendar year in liters pure alcohol
		Age-standardized prevalence of heavy episodic drinking among adolescents and adults
		Alcohol-related morbidity and mortality among adolescents and adults
Physical inactivity	10% reduction in prevalence of physical inactivity	Prevalence of insufficiently active adolescents, defined as <60 min of moderate or vigorous intensity activity daily
		Age-standardized prevalence of insufficiently active persons, aged ≥18 years (defined as <150 min moderate activity per week, or equivalent)
Salt or sodium intake	30% reduction in mean population intake of salt or sodium	Age-standardized mean population intake of salt (sodium chloride) per day in grams in persons ≥18 years old
Tobacco use	30% reduction in prevalence of tobacco use	Prevalence of current tobacco use among adolescents
		Age-standardized prevalence of current tobacco use among persons ≥18 years old
Biological risk factors		
Raised blood pressure	25% reduction in prevalence of raised blood pressure or contain the prevalence of raised blood pressure	Age-standardized prevalence of raised blood pressure among persons ≥18 years old (defined as systolic blood pressure ≥140 mmHg and/or diastolic blood pressure ≥90 mmHg); and mean systolic blood pressure
Diabetes and obesity	0% increase in diabetes or obesity	Age-standardized prevalence raised blood glucose or diabetes among ≥18 years old (fasting plasma glucose ≥7.0 mmol/L (126 mg/dL) or on medication for raised blood glucose)
		Prevalence overweight or obesity in adolescents (WHO growth reference for school-aged children and adolescents)
		Age-standardized prevalence of overweight or obesity ≥18 years (body mass index ≥25 kg/m²; overweight ≥30 kg/m² obesity)
Additional indicators		Age-standardized mean proportion of total energy intake from saturated fatty acids in persons ≥18 years old
		Age-standardized prevalence (≥18 years old) consuming less than five total servings (400 g) of fruit or vegetables per day
		Age-standardized prevalence raised total cholesterol aged ≥18 years (total cholesterol ≥5.0 mmol/L or 190 mg/dL)
National Response Systems		
Drug therapies to prevent heart attacks and strokes	≥50% coverage in eligible individuals in drug therapy and counseling	Proportion of eligible persons (defined as ≥40 years old with a 10-year cardiovascular risk of ≥30%, including those with existing CVD) receiving drug therapy and counseling (including glycemic control) to prevent heart attacks and strokes
Essential NCD medicines and basic technologies to treat major NCDs	80% coverage for technologies and essential NCD medicines to treat NCD in public and private facilities	Availability and affordability of quality, safe and efficacious essential NCD medicines, including generics, and basic technologies in both public and private facilities

XSL•FO
RenderX

Element	Targets	Relevant indicators
Additional indicators		Access to palliative care assessed by morphine-equivalent consumption of strong opioid per death from cancer
		Adoption of national policies that limit saturated fatty acids and virtually eliminate partially hydrogenated vegetable oils in the food supply
		Availability of cost-effective and affordable vaccines against human papillomavirus
		Policies to reduce the impact on children of marketing of foods and nonalcoholic beverages high in saturated fats, trans fatty acids, free sugars, or salt
		Vaccination coverage hepatitis B by number of third doses of Hepatitis-B vaccine (HepB3) administered to infants
		Proportion of women 30-49 years screened for cervical cancer at least once

[a]NCD: noncommunicable disease.

[b]mmol: millimole.

[c]mg: milligram.

[d]dL: deciliter.

[e]WHO: World Health Organization.

[f]CVD: cardiovascular disease.

Potential of Mobile Phone Surveys for NCD Data

Access to mobile phones has increased globally, reaching more than 100% saturation in some LMICs [10,11]. For the period 2010-2015, for example, mobile phone subscriptions per 100 people increased 30-85% in the following countries: Bangladesh (45 to 83), Ghana (85 to 130), Lebanon (66 to 87), Papua New Guinea (28 to 47), Uganda (38 to 50), and Zambia (41 to 74) [11]. mHealth technology is already being incorporated into public health data collection due to its potential to save time and costs, as well as its reliability across both highly developed urban centers and remote, rural areas. A recent review described 12 common mHealth strategies, from surveillance and training to point-of-care diagnostics, registries or vital events tracking, as well as commodities management [12]. Some advantages of mHealth-empowered surveillance include instantaneous collection and sharing of data (across large samples), enabling analysis of results by demographic and spatial characteristics, significant savings in the amount of time needed to obtain data from remote geographic areas (without the need to go there like traditional household surveys), and potential cost savings from reductions in human resources needed to administer questionnaires and enter data [12].

It was found that 83% of the 112 WHO member states responding to a 2014 survey reported having one or more mHealth initiatives; the most frequently reported categories were call centers or helplines, emergency toll free telephone services, and mobile telemedicine (reported by 49-59% of countries). Least frequent uses of mHealth were health surveys, surveillance, awareness, and decision support (reported by 19-26% of countries) [8]. The WHO noted that these findings appeared to be incongruous with peer-reviewed literature, which tends to report more frequent use of mHealth for data collection and disease surveillance. Possibly, those responding to the survey may not have been aware of mHealth programs in the nongovernmental or private sectors, or it is possible that nongovernmental organizations (NGO) or government programs are slow to share activities and findings in the academic literature [13,14]. To promote better integration, scale-up, and sustainability of mHealth initiatives, the mHealth Assessment and Planning for Scale (MAPS) toolkit provides detailed self-assessment for project teams to measure the scalability of their mHealth products, as well as planning and guidance strategies for surmounting barriers to scale [15].

Limited evidence exists on the comparative effectiveness of MPS modalities in LMICs though a variety of options are available. Three frequently employed modalities are interactive voice response (IVR), SMS, and computer-assisted telephone interview (CATI). IVR involves prerecorded audio messages or instructions given over the phone by a computer application; CATI involves one person interviewing another via the phone; and SMS involves communication between devices using short text messages [16]. Although studied in high-income countries, there is limited empirical evidence on the factors that influence MPS response, completion, and attrition, as well as ways to maximize these key metrics in LMIC settings across any of these modalities [17,18].

In one comparative study, 630 Lebanese adults selected through multistage stratified cluster sampling were respondents for a face-to-face (40-min) survey adapted from the WHO STEPS, and a mobile phone interview (abridged version of NCD questions). The study reported high percent agreements (89.5-95.6%) and kappa statistics (κ=.79 to .91) for past-year alcohol consumption, ever smoking, and current cigarette smoking; and over 90% agreement for diagnoses of hypertension, diabetes, hyperlipidemia, and heart disease (κ=.87 for diabetes; κ=.66 to .68 for hypertension, hyperlipidemia, and heart disease). Mobile phone interviews also saved about US $14 per person interviewed, when compared with face-to-face methods [19]. The World Bank compared MPS modalities in Peru and Honduras to collect frequent and nationally representative data by sending out 677 invitations on SMS, 383 for IVR, and 384 for CATI. Attrition rates were high for IVR

XSL•FO
RenderX

(75-81%) and SMS (70-79%) compared with CATI (49-61%); and modalities had different cost per interview including face-to-face (US $40), CATI (US $25), IVR (US $17), and SMS (US $8). In addition, answers to SMS and IVR were significantly different from face-to-face answers, with CATI responses being almost the same as face-to-face responses [16].

The evidence base around the efficacy and impact of mHealth programs is growing rapidly, with increasing numbers of projects being evaluated with methodological rigor [20,21]. However, a number of issues related to the optimization and conduct of MPS have been highlighted in the literature and appear to be largely unaddressed. These include concerns around equity, the optimal timing of surveys, appropriate financial incentives, and the quality and accuracy of responses [22,23]. As a nascent field, many mHealth strategies are still being tested and have not all been carefully evaluated; for example, a 2012 review of SMS interventions for NCD prevention in LMICs found that only 5 of 34 projects included in the review provided an evaluation of impact [22].

The proliferation of mobile phones, and the critical gaps raised by WHO and the literature highlight both the potential for MPS and the need for rigorous investigations of the feasibility, validity, and comparative effectiveness of different MPS modalities. In addition, there is a need to understand the ethical, legal, and societal issues (ELSI) around use of mobile phones for collecting NCD risk factors. These issues give rise to a rich research and development agenda proposed below.

NCD Risk Factors and Mobile Phone Survey: A Research Agenda

We propose a research and development agenda for NCD risk factors and MPS. The eventual purpose of the proposed agenda is to help standardize operating procedures for MPS, which will allow for comparisons of NCD risk factors within and across sites, and over time. Within this overall purpose, we propose specific goals. First, to answer key questions on how to design and deliver robust MPS to collect NCD risk factors; second, to assess the comparative effectiveness and costs of MPS modalities (eg, IVR, CATI, SMS); and third, to determine key ELSI in the development, in conduct, analysis, and reporting of MPS in LMICs. Each is further described below with proposed research objectives. These goals and specific objectives are also summarized in Table 2.

Under the *first goal* (to answer key questions on how to design and deliver robust MPS on NCDs), one research objective is to assess the usability of different ways of delivering MPS (eg, voice calls, recorded voice surveys—commonly referred to as interactive voice response (IVR)—or by text message) and community perceptions and willingness to complete MPS. This will probably require qualitative methods such as key informant interviews, focus group discussions, and user-group testing. Participants' perceptions and willingness to complete and ability to navigate MPS delivery modalities such as IVR, SMS, and CATI can be assessed in order to provide a user-designed MPS [24]. These formative efforts can help inform translation and wording of questions, and suggest ways to improve MPS response rates. Results can be used to improve MPS platform design and questionnaires.

A second objective is to assess the role of different incentive amounts in MPS. Quasi-experimental or randomized controlled trial designs can be used to assess the effect of incentives on standardized contact, response, completion, and refusal rates in a MPS [25]. Studies will need to assess how different monetary incentive amounts; timing of incentive delivery (eg, beginning of MPS vs end); and nature of incentive structure (eg, fixed amount vs lottery) affect MPS metrics. These studies will need specific sample size calculations and calculation of key survey metrics most likely stratified by key demographic characteristics. Important issues around sampling, weighting, and stratification also remain understudied for MPS under conditions of varying mobile phone ownership and/or access [26].

A third objective will be to assess the impact of MPS introductory messages. This will, for example, necessitate studies that assess the effect of MPS introductory message on contact, response, completion, refusal, and attrition rates. Research could further examine the introduction content (motivational vs informational), voice type (male vs female), and the introduction's character (casual voice vs formal) on key survey metrics by demographic characteristics. Exploring different sampling frames of mobile phone numbers is a fourth objective to assess the effect of the sampling frame on representativeness, response rate, completion rate, and costs of MPS. For example, the benefit of using random digit dialing (RDD) compared with a list of preexisting phone numbers, for example, from mobile network operators or previous surveys has not been characterized. Finally, the impact of NCD-specific questions (in modules) on survey attrition needs to be studied. NCD modules in existing surveys are groups of common-themed risk factor questions, such as diet and alcohol, and can be delivered in different order (or randomized) to assess survey impact by question number and content (also tests cultural sensitivities) by key demographics. The research agenda under this first goal will help optimize the delivery of MPS.

Under the *second goal* (to assess the comparative effectiveness of MPS modalities), it is important to understand the role and impact of different channels of communication with phone owners. One objective is to assess the effect of MPS modalities (eg, CATI vs IVR) on key metrics and performance indicators, for example, by administering the same survey content to a list of participants using two or more modalities. The costs of modality setup, deployment, and maintenance also need to be assessed to perform intermodal cost-comparisons. A second objective is to compare national or subnational estimates of NCD risk factor indicators between MPS modalities—this could be done by comparing NCD risk factor distributions captured using different MPS modalities. Variability in response to different questions, modules and risk factor prevalence captured by multiple modalities can be ascertained.

A third objective is to document MPS intermodal reliability and response consistency for NCD risk factors. This could be done by allocating (or randomizing) participants into two or more arms (eg, IVR and CATI) to establish congruency in responses for the same questions. Participants in one arm could receive

XSL•FO
RenderX

one modality followed by another modality a few days later. Crossover designs might also be possible for an assessment of response consistency, adjusted for the risk of priming after exposure to the prior modality. Differences in the reported prevalence of NCD risk factors can also be evaluated to quantify possible incongruences in responses between MPS modalities. These objectives under the second research goal will answer questions on comparative effectiveness assessment in MPS in LMIC settings, as well as questions as to which modality provides better information on key survey metrics.

Under the *third goal* (determine key ELSI) relevant issues need to be explored in the development, conduct, analysis, and reporting of MPS in LMICs. Objectives under this goal are to (1) determine key ethical issues encountered in the conduct of MPS; (2) identify common and preferred practices for obtaining individual consent or permission for MPS to ensure voluntary, informed data collection; (3) document commonly encountered regulatory complexities and challenges that arise when conducting MPS and how they have been addressed; and (4) identify key societal goals and values that are supportive of MPS and how these are balanced against other important interests (eg, expectations of privacy and confidentiality).

A systematic review of the literature is needed to collect the most common ELSI and collate them according to broad themes. Data collection with researchers, public health practitioners, programmers, policymakers, and other stakeholders of MPS may help define and assess the key ELSI, how they have been addressed in practice, and how well they are reflected in existing reports and theoretical frameworks identified by global stakeholders. Additional conceptual work may be required in this space to understand which ethics principles and values are being satisfied and which may be tested using MPS. Equally important is the critical examination of factors that influence whether, and to what extent, various ELSI apply across different types of mHealth and digital health activities. This research goal will provide crucial perspectives and data to inform the responsible implementation of MPS in, by and for LMICs. The three research goals and possible objectives described above comprise an initial proposed research agenda for MPS and NCD risk factor data collection with a special focus on LMICs (Table 2).

Integration Into Data Projects: An Opportunity

Implementing the proposed research agenda will require both resources and creativity, either as standalone activities or integrated into other agendas. As such, we introduce in this section, an opportunity for integrating this agenda into a stream of work around data for health and NCD. The Bloomberg Data for Health Initiative (BD4HI) is supported by Bloomberg Philanthropies and the Australian Government's InnovationXchange to improve the health of populations through strengthening public health data [27]. It involves multiple partners working in several sites including at national and subnational levels (Table 3). The project has an arm focused on strengthening vital statistics including birth and death registration in several countries. It also has an NCD arm implementing both nationally representative WHO STEPS survey and MPS using different mHealth platforms to generate comparable data across countries. The MPS are being designed to use or closely approximate the standardized STEPS survey with questions on demographics, tobacco use, alcohol consumption, diet, physical activity, and history of blood pressure and diabetes diagnoses and medications [6]. Finally, the project also has an arm focused on strengthening data impact and use of data for policies in countries.

It is in the context of this project that we proposed to integrate an initial research agenda for MPS (Table 2). Key to success of our research agenda is an understanding that this research would seek to inform MPS activities at the national level by feeding results to implementing partners and governments through our in-country collaborating institutions. In order to make this a realistic program of work, we decided to (1) initially focus on a few (2-4) selected countries in either the African, Middle Eastern, or Asian regions; (2) select two MPS modalities only for testing—IVR and CATI for now; and (3) work only at the subnational level. These initial research efforts will help inform future national level rollout in other countries considering adopting use of MPS. Table 2 includes a summary of our project-specific research agenda based on these decisions; this agenda incorporates elements of each of the three goals described above.

XSL•FO
RenderX

Table 2. Proposed research agenda for mobile phone survey (MPS) and noncommunicable disease (NCD) risk factor data collection

Goals	Objectives	BD4HI
Goal 1: Answer key questions on how to design and deliver robust MPS on NCDs.	Exploring the usability or technical requirements of MPS modalities, community perceptions, and willingness to take MPS.	The project will (1) identify lessons learned and challenges from individuals with experience of MPS (via key informant interviews); (2) understand community acceptance and willingness to respond to IVR (via focus group discussions); and (3) examine and refine the usability of an NCD risk factor survey delivered through an IVR platform (via semi-structured interviews).
	Exploring the impact of incentive amounts, incentive timing, and incentive structure on key MPS metrics.	Test impact of different incentive factors on key survey metrics; participants to receive different incentive (1) amounts (including none), (2) timing (prepaid to poststudy), and (3) structures (fixed, lottery).
	Exploring the impact of the content, type, and modality of introductory messages on key MPS metrics.	Exploring the impact of different content (informational, motivational) and voice (male, female) of IVR introduction.
	Exploring the impact of different sampling frames on key MPS metrics.	Assess the benefit of using random digit dialing (RDD) compared with a list of preexisting phone numbers
	Exploring the impact of specific questions and their order on key MPS metrics.	Examine impact of different orders of NCD modules (eg, diet, tobacco use, alcohol) on key survey metrics.
Goal 2: Assess the comparative effectiveness of MPS modalities	Assess impact of MPS modality on key metrics, performance characteristics, and costs of MPS.	Participants will respond to either CATI or IVR using the same questionnaire; and response characteristics will be studied.
	Compare national or subnational indicators between MPS modalities.	Responses of participants to IVR and CATI surveys will be compared.
	MPS intermodal reliability and response consistency.	Participants will be randomized to one of two arms: IVR then CATI, or CATI then IVR. The questionnaires used in both study arms will be the same. Crossover design will allow for an assessment of response consistency
Goal 3: Explore ethical, legal and societal issues (ELSI) in the development, conduct, analysis, and reporting of MPS	Determine key ethical issues in the conduct of MPS.	Conduct of a systematic review of the literature to collect the most common ELSI and collate them according to broad themes.
	Describe common and preferred practices for obtaining individual consent or permission for MPS.	Conduct a survey of researchers, programmers, users, and stakeholders of MPS on and from LMICs to help define the prevalence of key ELSI; and how they have been addressed.
	Document commonly encountered regulatory complexities when conducting MPS and how have they been addressed.	
	Identify key societal goals and values that are supportive of MPS and how these are balanced against other important interests.	

[a]BD4HI: Bloomberg Data for Health Initiative.

[b]MPS: mobile phone surveys.

[c]NCDs: noncommunicable diseases.

[d]IVR: interactive voice response.

[e]CATI: computer-assisted telephone interview.

[f]ELSI: ethical, legal, and societal issues.

[g]LMICs: low- and middle-income countries.

XSL•FO
RenderX

Table 3. Bloomberg Data for Health Project—key goals, components, and partners.

Key features	Description
Overall goal	To improve the health of populations through strengthening collection and use of public health data
Components	Civil Registration and Vital Statistics (CRVS) Improvement Noncommunicable Disease Risk Factor Surveillance Strategic Use of Data for Policy and Planning
Partners	Centers for Disease Control and Prevention, USA Johns Hopkins Bloomberg School of Public Health, USA University of Melbourne, Australia Vital Strategies World Health Organization
Donors	Bloomberg Philanthropies Australian Government

Discussion

Currently, there are over 7 billion mobile phone subscriptions around the globe [10]. The majority of subscriptions are in LMICs, which are undergoing demographic and epidemiologic transitions where NCDs are catching up to infectious diseases as leading causes of death. To address this challenge, countries need cost-effective, timely, and systematic means of conducting surveillance and tracking progress toward globally agreed targets for NCD prevention and control [6,28]. Mobile phones present an opportunity to conduct rapid risk factor surveillance that can be used to inform resource allocation, as well as evaluate NCD prevention and control efforts and policies [29]. Addressing unresolved questions about maximizing key metrics for MPS, comparative effectiveness of MPS, and ELSI in the development, conduct, analysis, and reporting of MPS, would aid collection of timely and quality information about NCD risk factors [17,18]. The eventual goal of the proposed research agenda is to help standardize operating procedures for MPS, which will allow for comparisons of NCD risk factors within and across sites, and over time.

Despite the many advantages they present, it is important to note that NCD risk factor MPS probably would not obviate the need for in-person data collection. Physical measurements such as blood pressure, weight or height, and waist circumference, or obtaining blood samples for analyses (eg, for lipids, glucose), need to be conducted in-person. The MPS can complement an NCD surveillance system that includes a STEPS or similar survey done less frequently, which together can inform prevention and policy decisions [30]. The surveys, of course, have to be coupled with other aspects of the WHO surveillance framework including outcomes (ie, disease rates, deaths), health system capacity, and response (ie, infrastructure, policies or plans, access to care or medicines, partnerships) [4].

The BD4HI offers a unique opportunity to test MPS technology and delivery options in real world settings, while enhancing our understanding of the norms and values that are likely to influence global implementation. Proof of concept on the use of this technology for NCD risk factor data collection could help countries to complement existing mechanisms for monitoring progress in their interventions toward achieving commitments for NCD global targets. We welcome the global NCD and mHealth communities to provide further input on the proposed research agenda and hope it will inform research projects in the coming years.

Acknowledgments

The studies described as part of the research agenda are funded by the Bloomberg Philanthropies. This funding agency had no role in the preparation of this manuscript.

Conflicts of Interest

None declared.

References

1. Global status report on noncommunicable diseases 2014. Geneva, Switzerland: World Health Organization; 2014.
2. Global action plan for the prevention and control of noncommunicable diseases 2013 - 2020. Geneva, Switzerland: World Health Organization; 2013.
3. Alwan A, Maclean DR, Riley LM, d'Espaignet ET, Mathers CD, Stevens GA, et al. Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. Lancet 2010 Nov 27;376(9755):1861-1868. [doi: 10.1016/S0140-6736(10)61853-3] [Medline: 21074258]
4. Global status report on noncommunicable diseases 2010. Geneva, Switzerland: World Health Organization; Apr 2011.

XSL•FO
RenderX

5. Bloom D, Cafiero E, Jane-Llopis E, Abrahams-Gessel S, Bloom L, Fathima S, et al. The global economic burden of noncommunicable diseases. Geneva, Switzerland: World Economic Forum; 2011.
6. WHO. Geneva, Switzerland: World Health Organization; 2003. STEPS: A framework for surveillance URL: http://www.who.int/ncd_surveillance/en/steps_framework_dec03.pdf [accessed 2017-03-24] [WebCite Cache ID 6pCQmc8ou]
7. NCD global monitoring framework: Ensuring progress on noncommunicable diseases in countries. Geneva, Switzerland: World Health Organization; 2013. URL: http://www.who.int/nmh/global_monitoring_framework/en/ [accessed 2017-03-16] [WebCite Cache ID 6ozsQBP5u]
8. mHealth: New horizons for health through mobile technologies. Geneva, Switzerland: World Health Organization; Jun 07, 2011.
9. Mehl G, Labrique A. Prioritizing integrated mHealth strategies for universal health coverage. Science 2014 Sep 12;345(6202):1284-1287. [doi: 10.1126/science.1258926] [Medline: 25214614]
10. International Telecommunication Union. 2016. Statistics URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2016-12-12] [WebCite Cache ID 6mgkxF0WI]
11. The World Bank. Mobile cellular subscriptions (per 100 people) URL: http://data.worldbank.org/indicator/IT.CEL.SETS.P2 [accessed 2016-11-12] [WebCite Cache ID 6lyVl5ftK]
12. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
13. Brinkel J, Krämer A, Krumkamp R, May J, Fobil J. Mobile phone-based mHealth approaches for public health surveillance in sub-Saharan Africa: a systematic review. Int J Environ Res Public Health 2014 Nov 12;11(11):11559-11582. [doi: 10.3390/ijerph111111559] [Medline: 25396767]
14. Labrique A, Vasudevan L, Chang LW, Mehl G. H_pe for mHealth: more “y” or “o” on the horizon? Int J Med Inform 2013 May;82(5):467-469 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.11.016] [Medline: 23279850]
15. The MAPS toolkit: mHealth assessment planning for scale. Geneva, Switzerland: World Health Organization; 2015.
16. Ballivian A, Azevedo J, Durbin W. Using mobile phones for high-frequency data collection. In: Toninelli D, Pinter R, de Pedraza P, editors. Mobile Research Methods: Opportunities and challenges of mobile research methodologies. London, England: Ubiquity Press; 2015.
17. Greenleaf A, Gibson DG, Khattar C, Labrique AB, Pariyo GW. Building the evidence base for remote data collection in low- and middle-income countries: comparing reliability and accuracy. J Med Internet Res 2017;19(5):e140 [FREE Full text] [doi: 10.2196/jmir.7331]
18. Gibson DG, Pereira A, Farrenkopf BA, Labrique AB, Pariyo GW, Hyder AA. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. J Med Internet Res 2017;19(5):e139 [FREE Full text] [doi: 10.2196/jmir.7428]
19. Mahfoud Z, Ghandour L, Ghandour B, Mokdad A, Sibai A. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? Field Methods 2015;27(1):39-54.
20. Higgs E, Goldberg A, Labrique A, Cook S, Schmid C, Cole C, et al. Understanding the role of mHealth and other media interventions for behavior change to enhance child survival and development in low- and middle-income countries: an evidence review. J Health Commun 2014;19(Suppl 1):164-189. [doi: 10.1080/10810730.2014.929763]
21. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363. [doi: 10.1371/journal.pmed.1001363]
22. Déglise C, Suggs LS, Odermatt P. Short message service (SMS) applications for disease prevention in developing countries. J Med Internet Res 2012;14(1):e3 [FREE Full text] [doi: 10.2196/jmir.1823] [Medline: 22262730]
23. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
24. Principles for Digital Development. The Principles URL: http://digitalprinciples.org/ [accessed 2016-12-12] [WebCite Cache ID 6mgks116N]
25. American Association for Public Opinion Research. 2016. Standard definitions: final dispositions of case codesoutcome rates for surveys URL: http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2017-03-16] [WebCite Cache ID 6p0u5VjxF]
26. Labrique AB, Blynn E, Ahmed S, Gibson DG, Pariyo GW, Hyder AA. Health surveys using mobile phones in developing countries: automated active strata monitoring and other statistical considerations for improving precision and reducing biases. J Med Internet Res 2017;19(5):e121 [FREE Full text] [doi: 10.2196/jmir.7329]
27. Bloomberg MR, Bishop J. Understanding death, extending life. Lancet 2015 Oct 17;386(10003):e18-e19. [doi: 10.1016/S0140-6736(15)00400-6] [Medline: 26434355]
28. Pan American Health Organization. 2016. Noncommunicable Diseases Surveillance URL: http://www.paho.org/hq/index.php?option=com_content&view=article&id=1512%3A2009-chronic-disease-surveillance&catid=1384%3Acncd-surveillance&Itemid=1663&lang=en [WebCite Cache ID 6p0uytRbb]

XSL•FO
RenderX

29. Pariyo GW, Wosu AC, Gibson DG, Labrique AB, Ali J, Hyder A. Moving the agenda on noncommunicable diseases: policy implications of mobile phone surveys in low- and middle-income countries. J Med Internet Res 2017;19(5):e115 [FREE Full text] [doi: 10.2196/jmir.7302]
30. Thacker SB, Stroup DF, Rothenberg RB. Public health surveillance for chronic conditions: a scientific basis for decisions. Stat Med 1995;14(5-7):629-641. [Medline: 7792453]

Abbreviations

BD4HI: Bloomberg Data for Health Initiative
CATI: computer-assisted telephone interview
CRVS: Civil Registration and Vital Statistics
CVD: cardiovascular disease
ELSI: ethical, legal, and societal issues
GPS: global positioning system
IVR: interactive voice response
LMIC: low- and middle-income countries
MAPS: mHealth Assessment and Planning for Scale
MPS: mobile phone surveys
NCD: noncommunicable disease
NGO: nongovernmental organization
RDD: random digit dialing
SMS: short message service
STEPS: WHO STEP-wise approach to NCD risk factor surveillance
WHO: World Health Organization

Edited by E Rosskam; submitted 02.01.17; peer-reviewed by P Devarsetty, C Hoe; comments to author 18.02.17; revised version received 27.02.17; accepted 28.02.17; published 05.05.17.

Please cite as:
Hyder AA, Wosu AC, Gibson DG, Labrique AB, Ali J, Pariyo GW
Noncommunicable Disease Risk Factors and Mobile Phones: A Proposed Research Agenda
J Med Internet Res 2017;19(5):e133
URL: http://www.jmir.org/2017/5/e133/
doi:10.2196/jmir.7246
PMID:28476722

©Adnan A Hyder, Adaeze C Wosu, Dustin G Gibson, Alain B Labrique, Joseph Ali, George W Pariyo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Moving the Agenda on Noncommunicable Diseases: Policy Implications of Mobile Phone Surveys in Low and Middle-Income Countries

George W Pariyo[1], MBChB, PhD; Adaeze C Wosu[2], MPH; Dustin G Gibson[1], MS, PhD; Alain B Labrique[1], MHS, MS, PhD; Joseph Ali[1,3], JD; Adnan A Hyder[1,3], MPH, MD, PhD

[1]Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD, United States

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

[3]Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
George W Pariyo, MBChB, PhD
Johns Hopkins Bloomberg School of Public Health
Department of International Health
615 North Wolfe Street E 8648
Baltimore, MD, 21205
United States
Phone: 1 410 502 5790
Fax: 1 410 614 1419
Email: gpariyo1@jhu.edu

Abstract

The growing burden of noncommunicable diseases (NCDs), for example, cardiovascular diseases and chronic respiratory diseases, in low- and middle-income countries (LMICs) presents special challenges for policy makers, due to resource constraints and lack of timely data for decision-making. Concurrently, the increasing ubiquity of mobile phones in LMICs presents possibilities for rapid collection of population-based data to inform the policy process. The objective of this paper is to highlight potential benefits of mobile phone surveys (MPS) for developing, implementing, and evaluating NCD prevention and control policies. To achieve this aim, we first provide a brief overview of major global commitments to NCD prevention and control, and subsequently explore how countries can translate these commitments into policy action at the national level. Using the policy cycle as our frame of reference, we highlight potential benefits of MPS which include (1) potential cost-effectiveness of using MPS to inform NCD policy actions compared with using traditional household surveys; (2) timeliness of assessments to feed into policy and planning cycles; (3) tracking progress of interventions, hence assessment of reach, coverage, and distribution; (4) better targeting of interventions, for example, to high-risk groups; (5) timely course correction for suboptimal or non-effective interventions; (6) assessing fairness in financial contribution and financial risk protection for those affected by NCDs in the spirit of universal health coverage (UHC); and (7) monitoring progress in reducing catastrophic medical expenditure due to chronic health conditions in general, and NCDs in particular. We conclude that MPS have potential to become a powerful data collection tool to inform policies that address public health challenges such as NCDs. Additional forthcoming assessments of MPS in LMICs will inform opportunities to maximize this technology.

(J Med Internet Res 2017;19(5):e115) doi:10.2196/jmir.7302

KEYWORDS

NCDs; policy; mHealth; policy analysis; surveys

Introduction

The growing burden of noncommunicable diseases (NCDs) such as hypertension, diabetes mellitus, obesity, asthma, and chronic obstructive pulmonary disease presents special challenges for health policy in low- and middle-income countries (LMICs) [1]. These challenges include priority setting and resource allocation in the context of resource constraints and lack of reliable and timely data for evidence-based policy and decision-making for NCD prevention and control. In LMICs, NCDs grew as a share of the major causes of death from 59% in 1990 to 64% in 2010 and were estimated at 67% in 2015

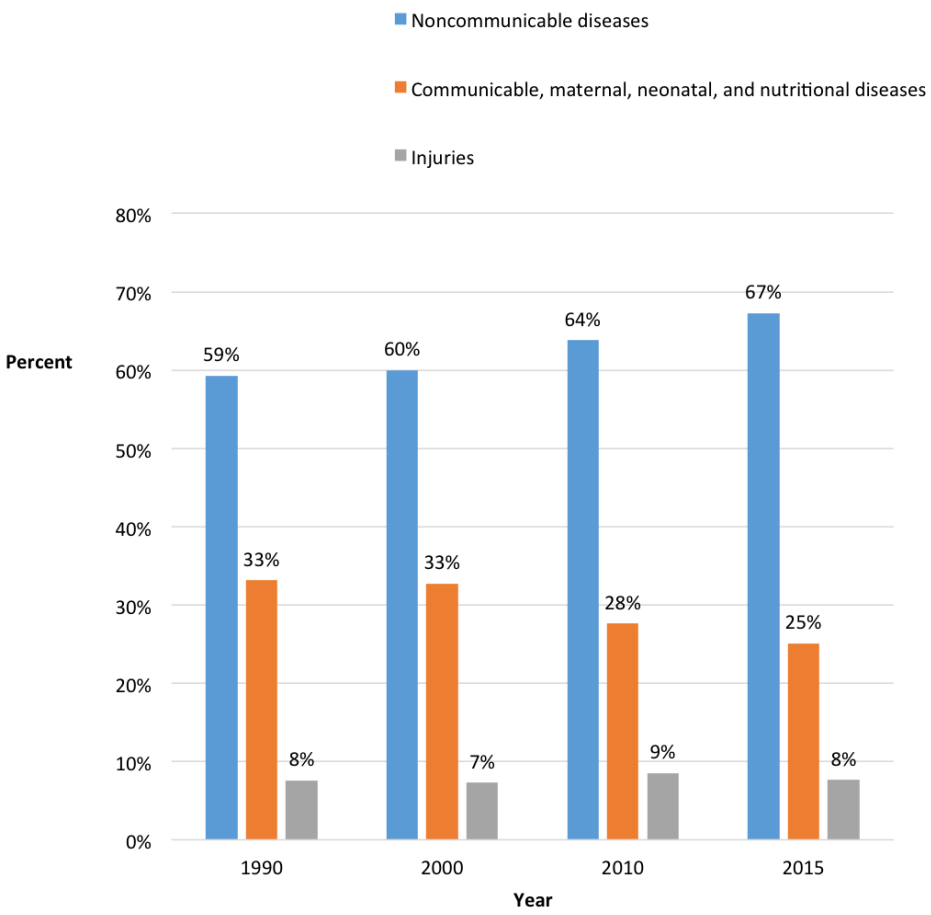
XSL•FO
RenderX

(calculations based on *Global Burden of Disease* report; see Figure 1) [1-2].

To assess the burden and prevalence of NCDs and risk factors in order to inform policy, programming, and to track progress, the World Health Organization (WHO) supports countries with the implementation and analysis of the STEPwise Approach to Surveillance (STEPS) of NCDs survey [3-4]. The STEPS survey contains 3 steps, or components, of NCD surveillance: (1) self-reported risk factor questionnaire, (2) physical measurement, and (3) biochemical measurement, all through face-to-face contact with respondents [4].

This paper highlights aspects of NCD-related health policy that decision-makers in LMICs face. These include identifying which risk factor-related priority interventions to implement, target groups to cover, how to ensure equity of coverage, and financial risk protection, among others. Decision-makers could benefit from newly available capabilities to conduct mobile phone surveys (MPS) in practically all settings, taking advantage of the growing ubiquity of mobile phones.

Figure 1. Changes in major causes of death in low- and middle-income countries (age-standardized): 1990-2015.

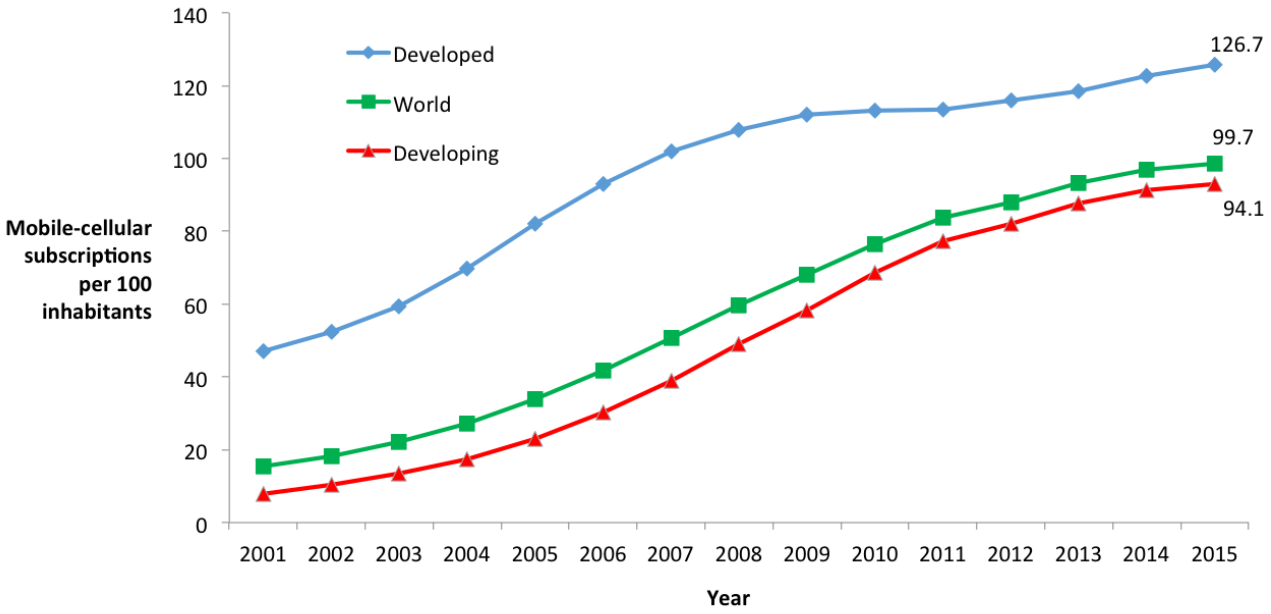


Rising Mobile Phone Subscriptions Present an Opportunity

To reduce the high costs and time requirements associated with conducting household surveys, higher income countries have developed and employed telephone and MPS to collect population-level estimates of health and demographics [5-6]. To date, such surveys are not widely used in LMICs. The global increase in mobile phone ownership and access has created an unprecedented opportunity to leverage mobile phones to revolutionize current methods of public health data collection in LMICs [6-9]; see Figure 2 [10]. Instead of relying only on face-to-face interviews conducted in respondents' homes, it is now possible to conduct interviews on a range of public health topics remotely by delivering short surveys and interviewing respondents over their mobile phones. Options for survey modalities include the use of short message, service (SMS) or text message interactive voice response (IVR), and computer-assisted telephone interviews (CATI), collectively referred to as MPS.

XSL•FO
RenderX

Figure 2. Mobile-cellular subscriptions per 100 inhabitants, 2001-2015.



Global Commitments on NCDs

In 2011, the United Nations (UN) declared NCDs a public health emergency and called on member states to implement NCD prevention and control strategies. These include reducing the harmful use of alcohol, implementing the Framework Convention on Tobacco Control, promoting healthy lifestyles such as healthy diets (eg, rich in fruit and vegetables), increased physical activity, and legislation to control and reduce the sodium content of processed foods [11].

In 2012, the UN committed to achieving universal health coverage (UHC) through a resolution adopted by member states [12]; however, it will not be possible to achieve UHC without addressing challenges posed by NCDs. Through a resolution of the UN General Assembly in 2015, the UN member states adopted the sustainable development goals (SDGs) to replace the millennium development goals (MDGs) [13]. The third goal in the SDGs states "Ensure Healthy Lives and Promote Well-Being for All at All Ages" and calls for reducing premature mortality from NCDs through prevention, treatment, and promotion of mental health and well-being [13]. Noting an inadequate capacity for NCD surveillance, the WHO has called for such surveillance systems to be developed urgently and, where they exist, strengthened for monitoring NCDs, and their risk factors [4,14]. Innovative solutions are needed to strengthen national NCD policies and their implementation, monitoring, and evaluation, and to ensure equity [15].

The WHO has developed an NCD global action plan with six objectives to (1) raise the priority accorded to prevention and control, (2) strengthen national capacity, leadership, governance, multisectoral action, and partnerships for prevention and control, (3) reduce modifiable risk factors and underlying social determinants, (4) strengthen and orient health systems to address prevention and control, and underlying social determinants, (5) promote and support national capacity for high-quality research and development for the prevention and control, and (6) monitor the trends and determinants of NCDs and evaluate progress in their prevention and control [14]. The basis for monitoring progress on implementation of the global action plan is the WHO NCD monitoring framework which has 25 indicators, and proposes a set of 9 voluntary global targets that member states can choose to track and report on an annual basis [3].

Getting NCDs on the National Policy Agenda

Having signed onto global commitments for UHC and the SDGs, countries have to find practical ways to translate these commitments into action. A comprehensive policy and practice framework for NCD surveillance, prevention, and care is essential to meeting national commitments to UHC and SDGs. Countries need to assess how and on what basis NCD policies are being formulated. Studies in a number of LMICs show that NCDs are a growing burden while health services are still largely inadequate to meet the needs of the population [16-18].

Translating commitments into action involves governments exercising stewardship and governance in the health system. One of the primary goals of health systems is to improve population health [19]. Currently, there exists limited good quality evidence on the state of NCD risk factors and effectiveness of NCD prevention and control efforts in LMICs [14,20]. Increasingly, countries are developing national policies and action plans for NCD prevention and control. For instance, the WHO reports that in 2014, of a total of 194 member states, 160 (82%) had regulations on age limits for the sale of alcohol, 76 (39%) had a written national policy on alcohol, and 52 (27%) had taken steps to implement such policies. In 2013, 95 (49%) of WHO member states had implemented at least one of four key tobacco control interventions [21].

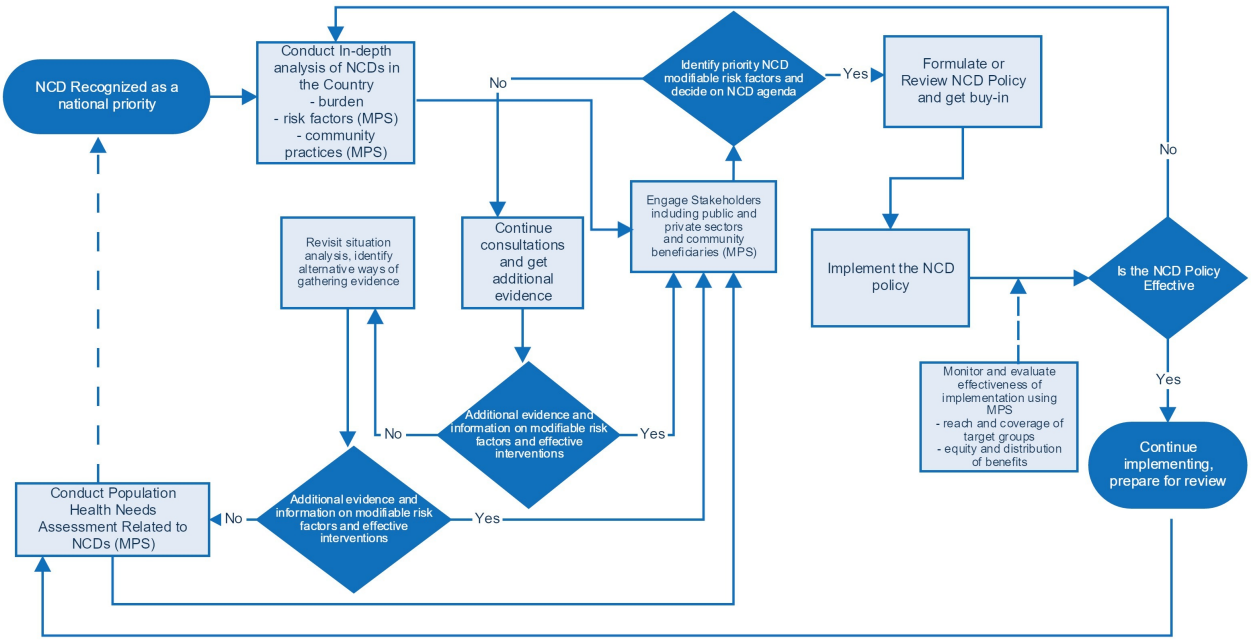
Apart from countries that have implemented STEPs surveys and have population-based data on NCDs, many LMICs still largely rely on data collected at health facilities. Such data are

XSL•FO
RenderX

inadequate given numerous limitations in LMIC settings including incompleteness, inaccuracies, poor quality, and lack of resources for reliable data storage, retrieval, analysis, and reporting [22,23]. Gathering data on population-based risk factors and existing NCD-related behaviors is greatly needed.

According to the paradigm of punctuated equilibrium periods of stasis in public policy are shaken from time to time by innovations or political events, such as change of government, which presents a window of opportunity and accords new priority to an issue or changes the dynamics [24,25]. Following years of relatively limited action, the MPS innovation represents a potential game-changer for national NCD policy-making in LMICs. MPS provides an opportunity for strengthening national agenda setting [24,25], including taking evidence-based action in the areas of NCD prevention and control. The technology can help generate new information rapidly, which can indicate the effectiveness or lack of a national policy, and thereby help to challenge the status quo of equilibrium [25,26]. The pathways through which MPS may influence NCD policy are highlighted in Figure 3.

Figure 3. Pathways for mobile phone surveys to influence noncommunicable disease policy.



Once NCDs are recognized as a current or emerging priority, a country may commission a situation analysis that would provide an in-depth assessment of the NCD burden, risk factors, and existing policy and programmatic responses. Surveys conducted using mobile phones could potentially play a role in generating population level evidence particularly as it pertains to behavioral risk factors and service utilization. Such an analysis would be useful in engaging key stakeholders from the public and private sectors, professional groups, civil society, academics and researchers, and health development partners. Such engagement could result in a clearer identification of the modifiable risk factors in an agenda setting decision-making. The policy formulated in this process has to be implemented. Again MPS could play a key role in monitoring population level reach, acceptability, and effectiveness of interventions. Information collected rapidly from MPS during implementation would help inform monitoring of progress as well as contribute to program evaluation. These activities would provide new input into the identification of population level needs and a repetition of the policy cycle.

In using MPS-generated data for decision-making, it should be kept in mind that no tool or approach is without shortcomings. MPS are not free of bias. The information collected from target populations will be a representative of the population that owns mobile phones [27]. It is also still to be explored how the findings from MPS can be interpreted, taking into account people likely to be unreached or underrepresented, such as people who do not own mobile phones or who live in areas where network coverage is patchy or nonexistent. A number of research and development aspects related to the use of MPS still need to be adequately addressed [28]. Similarly, a number of MPS development challenges such as increasing survey completeness, representativeness and response rates, as well as the potential role for use of incentives are to be kept in mind and are discussed elsewhere [29].

Identifying Priority NCD Modifiable Risk Factors

Strategies to collect population level NCD-related risk factor data through face-to-face household surveys, such as the WHO STEPs surveys, demand time and financial resources even though they provide a comprehensive picture every 5-7 years. Although MPS are not a substitute for current surveys administered through face-to-face interactions, they offer promise as interim and rapid assessment tools. They can be useful in identifying population level changes in modifiable risk factors such as physical activity levels, dietary intake, smoking, and harmful use of alcohol; and help inform equitable coverage of interventions such as NCD-related health messaging or access to care.

XSL•FO
RenderX

MPS can assist policy makers in identifying high-risk groups, such as groups with above-average prevalence or clustering of behavioral risk factors, for better targeting of interventions. During priority setting, the community—a key stakeholder—often is difficult to engage to get their perceptions and inputs [30-33]. With MPS, this barrier can potentially be overcome, for example, by polling people at the population level on their policy or resource allocation preferences, or the extent to which particular services are reaching them. As shown in other sectors, with MPS, the time for gathering population-based data, as well as time for analysis and data sharing can be considerably shortened compared with traditional approaches of data collection. For example, the use of mobile phones for rapidly collecting data to inform policies has been documented by the World Bank-supported *Listening to Africa* (L2A) and *Listening to Latin America and the Caribbean* (LAC) programs, as well as others. The L2A and LAC are pilot programs that recruit panels of respondents from the community level, and upon obtaining consent, call them regularly to collect community views and experiences on a wide range of topics including provision of services such as health, education, water and sanitation, and emergency assistance following a crisis. For instance, in Tanzania, the World Bank partners with Twaweza, a local nongovernmental organization. Twaweza's "sauti za wananchi," Swahili for "voices of citizens," collected data monthly from a representative population sample on topics such as education, health, a new tax, and the new national constitution [8,9,34,35]. These experiences suggest that, although not yet widely used in public health policy, there is strong potential for wider use for improved citizen engagement in the public policy space, which would benefit public health.

Schmidt and Barnhill [15] argue that, in order to achieve the SDGs, it will be important to take into account equity dimensions in implementing interventions for NCDs. The concern for improving health equitably is central to public health policies. MPS offer a possibility to generate timely population-based data at regular intervals thus permitting a review of existing NCD priorities, identifying which groups are being reached or not reached, and taking into account community perspectives.

Engaging Stakeholders in NCD Policy Formulation

Public health policy formulation and analyses have traditionally tended to be dominated by technocrats. Policy beneficiaries, while one of the key groups of actors in the policy triangle made up of content, context, and process [36], are rarely, if ever, consulted on their real needs, perceptions, preferences, and utilization of existing facilities. Even if an intervention is technically sound, lack of involvement of intended beneficiaries may result in formulating a less appropriate policy and decrease its effectiveness along the coverage continuum [37,38]. However, providing information such as may be collected from MPS is a necessary though not sufficient input into the process. Engaging communities also requires an enabling environment and democratic culture of consultation and public dialog, which go beyond the realm of public health, and have implications on governance and the stewardship role of government on one hand as well as strong civil society on the other. Success in formulating an adequate public health policy and implementing it depends on concerted action at various levels, involving partnerships in the public and private sectors, community representatives, political leaders, technical experts such as researchers and academics, and representatives of professional groups [36]. Beneficiary communities are not simply passive recipients of policy interventions; they are among the key actors in the uptake of policy interventions. Their acceptance or rejection of policies can make the difference between success and failure. Using MPS to identify or track trends in key NCD risk factors can be a catalyst to initiate engagement or strengthen the involvement of various stakeholders in order to more adequately address the NCD challenge. Identifying a trend showing increases in binge drinking through rapid surveys on harmful use of alcohol among teenagers, for instance, could act as a stimulus to engage with youth groups, parents, educators, political leaders, health professionals, and behavior change advocates to identify the most effective messaging and strategies to address the issue.

There is potential for MPS to provide timely new information that leads to building a coalition of support during the process of NCD policy reviews [36]. New actors or policy entrepreneur [39] could emerge from any of the stakeholder groups and help strengthen the NCD policy development process. In this way, beneficiaries become central actors rather than passive bystanders, in a way that is consistent with global movements toward patient- and client-centered health management. Although this may demand new skills to incorporate such data in meaningful ways, potential exists to facilitate better planning and delivery of interventions to mitigate NCD risk factors. The negotiation between what is technically feasible and politically and practically realistic is an essential aspect of policy development and implementation [40,41].

NCD Policy Implementation and Monitoring

Successful policy implementation needs more than a technically sound intervention. Frieden [42] proposes the following six components that are essential for effective public health program implementation: (1) innovation to develop evidence base for action; (2) technical package of a limited number of evidence-based interventions that will have a large impact; (3) managing performance, especially through rigorous real-time monitoring and evaluation leading to improvement in program implementation; (4) partnerships and coalitions (eg, between public and private sectors); (5) communication of accurate timely information to decision-makers, implementers, and the public; and (6) political commitment to obtain resources and support for effective action. MPS can contribute in generating timely evidence and speeding up the availability of information to various actors in the policy implementation process. Implementation of policy in real life is also influenced by perceptions of a situation during the interface between implementers and beneficiaries. The final step in the implementation chain is dependent on the interpretations of the policy by those functionaries at the interface with users. Lipsky [43] called these final implementers as street-level bureaucrats and posits that this effectively makes them policy makers. The implementers, for example, nurses in a clinic, or community health workers (CHWs) in a village, have the ability to influence perceptions of the users and communities of the intended policy

XSL•FO
RenderX

beneficiaries. The potential to regularly assess the coverage, frequency, and effectiveness of the interface with beneficiaries during implementation using MPS gives the policy maker higher up in the system an unprecedented means to gauge what is happening on the ground in reality (Table 1). This can be particularly useful in informing policy makers in LMICs on what services are being offered, who is delivering them, and how national level policies are received. This is important especially as these countries tend to have a mix of formal public and private sector providers as well as a significant role played by other informal actors including quacks and traditional and complementary medicine practitioners. The use of MPS can help flag emerging needs and opportunities for course correction before too much time has passed and potentially avoid unnecessary wastage of scarce resources on ineffective interventions.

Policy makers can use MPS to monitor (1) whether services are reaching intended beneficiaries or not and (2) other important aspects as shown in Table 1. With rapid monitoring, ineffective interventions or groups not being reached with current levels of effort and resource allocation can be detected early in the course of policy implementation. Policies often have unintended and unwanted consequences [36], the effects of which often become known after a long period of implementation and after considerable resources have been deployed. MPS can help to rapidly identify unintended or unwanted consequences so that the managers can take remedial action. In Malawi, for example, feedback from communities was used by health care managers in deciding to change the mix of personnel delivering services through a task sharing arrangement [17]. In South Africa, CHWs used mobile phones to increase screening for cardiovascular diseases at the community level, effectively increasing screening rates and reducing errors [44]. MPS can help to rapidly collect data on how much people are spending on NCD-related health care costs and monitor for the risk of catastrophic NCD-related health expenditures, especially given that many people are outside the health care financial safety system, even in the era of UHC, particularly in LMICs. Htet et al [45] found that, in Myanmar, up to 60% of out-of-pocket expenditure for NCDs was for medicines. For personnel, delivering care such as CHWs, MPS can help empower them with access to rapid survey-based information to target their activities. Mayosi [46] makes the case for the need to integrate surveillance and care of NCDs.

Evaluating NCD Policy Effectiveness

Finally, every policy should be evaluated to assess its effectiveness, revise the policy, or change direction altogether. Sabatier [26] estimated that typically one needs up to a decade for a reasonable evaluation of a policy's impact. Part of the long lead-time is because policies need time to take root and be effectively implemented. Some of the delay, however, is due to the time taken to implement, analyze, and report findings from surveys using traditional methods. For some of the delay due to long lead-time of traditional surveys, MPS have the potential to shorten the cycle by making information more frequently available for use in monitoring trends and for evaluation of policy effectiveness and impact. Managers, for instance, can assess service coverage using a continuum to include (1) availability: which services are available to the community, (2) accessibility: how accessible are the services, (3) acceptability of services, (4) actual utilization, and (5) effectiveness of utilization [38]. Such assessments would help inform policy and programs, for example, extent of catastrophic expenditure arising from NCD-related health care and its causes may lead to resource re-allocations to better address population needs.

Table 1 summarizes some of the main topics relevant to NCD policy formulation, monitoring implementation, and policy evaluation, and the potential role of using MPS to generate the information needed.

XSL•FO
RenderX

Table 1. Examples of ways mobile phone surveys can contribute to noncommunicable disease (NCD) policy formulation, monitoring implementation, and evaluation (based on authors' assessments and expectations).

Topic or issue	Examples of information needed for policy and program management	Potential contribution from using MPS[a] (high, medium, low)		
		Formulation	Monitoring	Evaluation
Service delivery and utilization	Services most needed at the community level; groups most affected by different risk factors; where people seek care for NCD[b]-related services; frequency of contact between providers and users	High	High	High
Equity	Whether service delivery is equitable; who is being reached with interventions or not?	High	High	High
NCD care benefit packages	Informing selection, for example, priority target groups to benefit from NCD-related services and financial subsidies; tracking achievements of targets; assessing household care utilization; and economic implications	Low	Medium	High
Public-private partnership	NCD-related services being accessed through the public or private sector; effectiveness of contracted providers in reaching beneficiaries	High	Medium	High
Continuity of care	Coverage of continued care in the community, for example, people with hypertension who have their blood pressure monitored close to where they live	Low	High	High
Access to essential medications	Access to medications close to where people live	High	High	High
Behavior change efforts	Source and uptake of behavior change communication messages; role of incentives and disincentives to facilitate healthy behavior, for example, increasing physical activity	Medium	High	High
Fairness in financial contribution and financial risk protection	Public spending and subsidies aimed at the poor—if reaching the intended beneficiaries and preventing catastrophic medical expenditure; costs of seeking care for NCD-related conditions and source of payments	Low	High	High
Health system responsiveness	Whether services are responsive to people's expectations, user-satisfaction with existing NCD services	Medium	High	High
Health management information systems	Triangulating data from routine facility-based information systems with population-level data, for example, on characteristics of service users	Low	Medium	High
Universal health coverage	Coverage of the population in scope and reach of NCD[b] services	Low	High	Low
Pharmaceutical policies related to NCDs	Drugs to allow for use at community level; how to monitor safe use; rational drug use	Low	Medium	High

[a]MPS: mobile phone surveys.

[b]NCD: noncommunicable disease.

Discussion

NCDs present special challenges for health policy, especially in LMIC settings. How can recent advances in mobile phone technology (eg, ability to conduct national surveys more frequently than traditional public health surveys and at lower cost), be leveraged to strengthen formulation, implementation, and evaluation of health policy? The coming together of problems, alternative solutions, and politics can create windows of opportunity for action [25]. We advocate for going beyond current efforts of collecting NCD-related data. Current efforts are based on passive identification of persons already affected by NCDs who seek care at health facilities and costly population-based household surveys. MPS can help to speed-up population-based risk factor surveillance, monitoring of trends, identification of high-risk groups, and monitoring perceptions and effectiveness of interventions at population level more frequently than existing face-to-face surveys. The high costs and time commitments of conducting face-to-face surveys on NCD risk factors means that they are conducted infrequently, about every 5-7 years. Although they generate data that are useful for long-term planning and evaluations, these surveys are costly and take time to plan, fund, and conduct. Consequently, national policy makers and program managers, who wish to track progress and do course correction in a more timely fashion, are left with a gap. With scientifically sound methodology and robustness, MPS potentially provide a more cost-efficient and timely mechanism to collect data on NCD risk factors and other public health priorities.

It is our hope that the near future will see an increase in studies that examine the role of MPS in health in LMICs. A review by Peiris [47] found that only a few studies have examined the role and potential for mHealth to improve health, especially in LMICs. Bloomfield et al [48] found limited evidence on use or

XSL•FO
RenderX

benefit for health of mHealth in sub-Saharan Africa, despite the growing use of mobile devices. Such evidence points to mHealth being an underexploited resource [48]. During the process of scanning for policy alternatives, MPS can help to zoom in on those requiring a closer look and thus promote more relevant options [49].

The potential role of public-private partnerships and ethical, legal, and societal issues that arise in using MPS to serve the public good are still to be investigated [50]. Furthermore, MPS can at best serve as a complement to household surveys—face-to-face interactions are still needed for obtaining measurements such as body weight and height, and blood samples for glucose and lipid tests, at least for the foreseeable future.

Despite some limitations, MPS can facilitate engaging communities to better understand their perspectives, concerns, preferences for service delivery, and continuity of care, in order to develop policy options that are accessible to communities, for example, a drug policy that ensures people access appropriate medications for NCDs close to where they live and establishing public-private partnerships with drug shops and clinic operators.

The issues we have highlighted in this paper are neither comprehensive nor exhaustive. There are certainly other key gaps in knowledge that will be relevant for formulation or implementation of policies on NCDs relevant for prevention, risk factor surveillance, treatment and control, and rehabilitation. We have tried to focus on those which, in our view, are amenable to being informed by MPS.

It is imperative that prevention strategies be developed which address NCDs and monitor trends of risk factors [4,51]. In this paper, we have attempted to argue that MPS represent a disruptive innovation [52] that can help in this endeavor taking advantage of a convergence of opportunities. The recognition of NCDs as a global public health emergency, availability of interventions, and improved capabilities to generate timely information using MPS represent emerging multiple streams [25]. The time has come to take advantage of these to benefit NCD prevention and control policy.

Conclusions

MPS offer a powerful tool for collecting population-based data that can help inform policies to address key public health challenges such as NCDs. MPS can be developed for use at national and subnational levels. This paper has laid out some of the key NCD-related policy areas and processes that could benefit from the emerging possibility of high-quality MPS in LMICs. More forthcoming assessments of MPS in LMICs will help to provide real-world experiences.

If LMICs are going to deal effectively with the existing and growing burden of NCDs, they cannot afford to continue using a passive approach of waiting for people to fall sick with NCDs and then attempting to treat. Furthermore, MPS have potential to promote improved accountability and transparency in policy processes through regularly sharing results based on the most current population-based data, which will be increasingly important in the context of global attention on NCDs and national efforts in the push for UHC.

Acknowledgments

This paper was prepared as part of background to the research agenda funded by the Bloomberg Philanthropies. This funding agency had no role in the preparation of this manuscript.

Conflicts of Interest

None declared.

References

1. Murray Christopher J L, Barber RM, Foreman KJ, Abbasoglu OA, Abd-Allah F, Abera E, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition. Lancet 2015 Nov 28;386(10009):2145-2191 [FREE Full text] [doi: 10.1016/S0140-6736(15)61340-X] [Medline: 26321261]
2. GBD Results Tool. 2017. Institute for Health Metrics and Evaluation URL: http://ghdx.healthdata.org/gbd-results-tool [WebCite Cache ID 6nHHoQq2f]
3. Riley L, Guthold R, Cowan M, Savin S, Bhatti L, Armstrong T, et al. The World Health Organization STEPwise approach to noncommunicable disease risk-factor surveillance: methods, challenges, and opportunities. Am J Public Health 2016 Jan;106(1):74-78. [doi: 10.2105/AJPH.2015.302962] [Medline: 26696288]
4. WHO. 2003. STEPS: A framework for surveillance URL: http://www.who.int/ncd_surveillance/en/steps_framework_dec03.pdf [WebCite Cache ID 6nF2cctgW]
5. Barr ML, van Ritten JJ, Steel DG, Thackway SV. Inclusion of mobile phone numbers into an ongoing population health survey in New South Wales, Australia: design, methods, call outcomes, costs and sample representativeness. BMC Med Res Methodol 2012;12:177 [FREE Full text] [doi: 10.1186/1471-2288-12-177] [Medline: 23173849]
6. Liu B, Brotherton JM, Shellard D, Donovan B, Saville M, Kaldor JM. Mobile phones are a viable option for surveying young Australian women: a comparison of two telephone survey methods. BMC Med Res Methodol 2011 Nov 24;11:159 [FREE Full text] [doi: 10.1186/1471-2288-11-159] [Medline: 22114932]

XSL•FO
RenderX

7. Mtema Z, Changalucha J, Cleaveland S, Elias M, Ferguson HM, Halliday JE, et al. Mobile phones as surveillance tools: implementing and evaluating a large-scale intersectoral surveillance system for rabies in Tanzania. PLoS Med 2016 Apr;13(4):e1002002 [FREE Full text] [doi: 10.1371/journal.pmed.1002002] [Medline: 27070315]
8. Ballivian A, Azevedo J, Durbin W. Using Mobile Phones for High-Frequency Data Collection. In: Toninelli D, Pinter R, de Pedraza P, editors. Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies. London: Ubiquity Press; 2015:21-39.
9. Croke K, Dabalen A, Demombynes G, Giugale M, Hoogeveen J. openknowledge.worldbank. 2013. Collecting high-frequency data using mobile phone: do timely data lead to accountability? URL: https://openknowledge.worldbank.org/bitstream/handle/10986/17046/NonAsciiFileName0.pdf?sequence=1 [accessed 2017-01-11] [WebCite Cache ID 6nAo7AHo0]
10. Statistics. 2016. International Telecommunication Union URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [WebCite Cache ID 6mgkxF0WI]
11. UN. 2011. High-level meeting on non-communicable diseases URL: http://www.un.org/en/ga/president/65/issues/ncdiseases.shtml [accessed 2017-01-02] [WebCite Cache ID 6nECzX06R]
12. Universal Health Coverage Day. 2012. UN Resolution: United Nations General Assembly Sixty-seventh Session Global Health and Foreign Policy Adopted 12 December 2012 URL: http://universalhealthcoverageday.org/un-resolution/ [accessed 2016-11-15] [WebCite Cache ID 6m30u4E3J]
13. UN. 2015. Resolution Adopted by the General Assembly on 25 September 2015 URL: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E [accessed 2017-01-02] [WebCite Cache ID 6nDu4FW9V]
14. WHO. 2013. 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?ua=1 [WebCite Cache ID 6nF45h3Mu]
15. Schmidt H, Barnhill A. Equity and noncommunicable disease reduction under the sustainable development goals. PLoS Med 2015 Sep;12(9):e1001872 [FREE Full text] [doi: 10.1371/journal.pmed.1001872] [Medline: 26349050]
16. Zaman MM, Bhuiyan MR, Karim MN, MoniruzZaman, Rahman MM, Akanda AW, et al. Clustering of non-communicable diseases risk factors in Bangladeshi adults: An analysis of STEPS survey 2013. BMC Public Health 2015 Jul 14;15:659 [FREE Full text] [doi: 10.1186/s12889-015-1938-4] [Medline: 26169788]
17. Lupafya PC, Mwagomba BL, Hosig K, Maseko LM, Chimbali H. Implementation of policies and strategies for control of noncommunicable diseases in Malawi: challenges and opportunities. Health Educ Behav 2016 Apr;43(1 Suppl):64S-69S. [doi: 10.1177/1090198115614313] [Medline: 27037149]
18. Guwatudde D, Mutungi G, Wesonga R, Kajjura R, Kasule H, Muwonge J, et al. The epidemiology of hypertension in Uganda: findings from the National Non-Communicable Diseases Risk Factor Survey. PLoS One 2015;10(9):e0138991 [FREE Full text] [doi: 10.1371/journal.pone.0138991] [Medline: 26406462]
19. WHO. 2000. The world health report 2000 - health sytems: improving performance URL: http://www.who.int/whr/2000/en/whr00_en.pdf?ua=1 [WebCite Cache ID 6nDubMExp]
20. Samb B, Desai N, Nishtar S, Mendis S, Bekedam H, Wright A, et al. Prevention and management of chronic disease: a litmus test for health-systems strengthening in low-income and middle-income countries. Lancet 2010 Nov;376(9754):1785-1797. [doi: 10.1016/S0140-6736(10)61353-0]
21. WHO. 2014. Global status report on noncommunicable diseases 2014 URL: http://www.who.int/nmh/publications/ncd-status-report-2014/en/ [WebCite Cache ID 6nF4Ar1rO]
22. Mutale W, Chintu N, Amoroso C, Awoonor-Williams K, Phillips J, Baynes C, Population Health ImplementationTraining – Africa Health Initiative Data Collaborative. Improving health information systems for decision making across five sub-Saharan African countries: implementation strategies from the African Health Initiative. BMC Health Serv Res 2013;13(Suppl 2):S9 [FREE Full text] [doi: 10.1186/1472-6963-13-S2-S9] [Medline: 23819699]
23. Wagenaar BH, Gimbel S, Hoek R, Pfeiffer J, Michel C, Manuel JL, et al. Effects of a health information system data quality intervention on concordance in Mozambique: time-series analyses from 2009-2012. Popul Health Metr 2015;13:9 [FREE Full text] [doi: 10.1186/s12963-015-0043-3] [Medline: 25821411]
24. Baumgartner F. Agenda setting in comparative perspective. Perspect polit 2016 Jun 13;14(02):456-460. [doi: 10.1017/S1537592716000141]
25. Baumgartner F. John Kingdon and the evolutionary approach to public policy and agenda-setting. In: Zahariadis N, editor. Handbook of public policy agenda-setting. Cheltenham, UK: Edward Elgar Publishing; 2016.
26. True JL, Jones BD, Baumgartner FR. Punctuated-equilibrium theory: explaining stability and change in public policymaking. In: Sabatier P, editor. Theories of the policy process. Boulder, Colorado: Westview Press; 2007.
27. Labrique AB, Blynn E, Ahmed S, Gibson DG, Pariyo GW, Hyder AA. Health surveys using mobile phones in developing countries: automated active strata monitoring and other statistical considerations for improving precision and reducing biases. J Med Internet Res 2017;19(5):e121 [FREE Full text] [doi: 10.2196/jmir.7329]
28. Hyder A, Wosu A, Gibson D, Labrique A, Ali J, Pariyo G. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246]
29. Gibson D, Pereira A, Farrenkopf B, Pariyo G, Labrique A, Hyder A. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. J Med Internet Res 2017;19(5):e139 [FREE Full text] [doi: 10.2196/jmir.7428]

XSL•FO
RenderX

30. Shayo EH, Mboera LE, Blystad A. Stakeholders' participation in planning and priority setting in the context of a decentralised health care system: the case of prevention of mother to child transmission of HIV programme in Tanzania. BMC Health Serv Res 2013 Jul 12;13:273 [FREE Full text] [doi: 10.1186/1472-6963-13-273] [Medline: 23849730]
31. Kamuzora P, Maluka S, Ndawi B, Byskov J, Hurtig A. Promoting community participation in priority setting in district health systems: experiences from Mbarali district, Tanzania. Glob Health Action 2013 Nov 25;6:22669 [FREE Full text] [Medline: 24280341]
32. Hipgrave DB, Alderman KB, Anderson I, Soto EJ. Health sector priority setting at meso-level in lower and middle income countries: lessons learned, available options and suggested steps. Soc Sci Med 2014 Feb;102:190-200. [doi: 10.1016/j.socscimed.2013.11.056] [Medline: 24565157]
33. Bolsewicz AK, Hipgrave D, Jimenez-Soto E. Public engagement in health priority setting in low- and middle-income countries: current trends and considerations for policy. PLoS Med 2013 Aug;10(8):e1001495 [FREE Full text] [doi: 10.1371/journal.pmed.1001495] [Medline: 23940462]
34. Worldbank. 2016. Listening to Africa URL: http://www.worldbank.org/en/programs/listening-to-africa [WebCite Cache ID 6nB4t07vu]
35. El-Jardali F, Makhoul J, Jamal D, Ranson MK, Kronfol NM, Tchaghchagian V. Eliciting policymakers' and stakeholders' opinions to help shape health system research priorities in the Middle East and North Africa region. Health Policy Plan 2010 Jan;25(1):15-27 [FREE Full text] [doi: 10.1093/heapol/czp059] [Medline: 19948770]
36. Walt G, Gilson L. Reforming the health sector in developing countries: the central role of policy analysis. Health Policy Plan 1994 Dec;9(4):353-370. [Medline: 10139469]
37. Tugwell P, Robinson V, Grimshaw J, Santesso N. Systematic reviews and knowledge translation. Bull World Health Organ 2006 Aug;84(8):643-651 [FREE Full text] [Medline: 16917652]
38. Tanahashi T. Health service coverage and its evaluation. Bull World Health Organ 1978;56(2):295-303 [FREE Full text] [Medline: 96953]
39. Shearer J. Policy entrepreneurs and structural influence in integrated community case management policymaking in Burkina Faso. Health Policy Plan 2015 Dec;30(Suppl 2):ii46-ii53 [FREE Full text] [doi: 10.1093/heapol/czv044] [Medline: 26516150]
40. Lipsky M. Street-level bureaucrats as policy makers (1970). In: Shafritz JM, Layne KS, Borick CP, editors. Classics of public policy. New York: Pearson Longman; 2005:51-61.
41. Stone D. The art of political decision making (1997). In: Shafritz JM, Layne KS, Borick CP, editors. Classics of public policy. New York: Pearson Longman; 2007:62-70.
42. Frieden TR. Six components necessary for effective public health program implementation. Am J Public Health 2014 Jan;104(1):17-22. [doi: 10.2105/AJPH.2013.301608] [Medline: 24228653]
43. Lipsky M. Street Level Bureaucracy: Dilemmas of the Individual in Public Service. 30th Anniversary Expanded Edition. New York: Russell Sage Foundation Publications; 2010.
44. Surka S, Edirippulige S, Steyn K, Gaziano T, Puoane T, Levitt N. Evaluating the use of mobile phone technology to enhance cardiovascular disease screening by community health workers. Int J Med Inform 2014 Sep;83(9):648-654 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.008] [Medline: 25002305]
45. Htet S, Alam K, Mahal A. Economic burden of chronic conditions among households in Myanmar: the case of angina and asthma. Health Policy Plan 2015 Nov;30(9):1173-1183 [FREE Full text] [doi: 10.1093/heapol/czu125] [Medline: 25468454]
46. Mayosi BM, Flisher AJ, Lalloo UG, Sitas F, Tollman SM, Bradshaw D. The burden of non-communicable diseases in South Africa. Lancet 2009 Sep 12;374(9693):934-947. [doi: 10.1016/S0140-6736(09)61087-4] [Medline: 19709736]
47. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
48. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014;10:49 [FREE Full text] [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]
49. Etzioni A. Mixed-scanning: a “third” approach to decision-making. Publ Admin Rev 1967 Dec;27(5):385-392. [doi: 10.2307/973394]
50. Ali J, Labrique AB, Gionfriddo K, Pariyo GW, Gibson DG, Pratt B, et al. Ethics considerations in global mobile phone-based surveys of noncommunicable diseases: a conceptual exploration. J Med Internet Res 2017;19(5):e110 [FREE Full text] [doi: 10.2196/jmir.7326]
51. Boutayeb A. The double burden of communicable and non-communicable diseases in developing countries. Trans R Soc Trop Med Hyg 2006 Mar;100(3):191-199. [doi: 10.1016/j.trstmh.2005.07.021] [Medline: 16274715]
52. Christensen CM, Baumann H, Ruggles R, Sadtler TM. Disruptive innovation for social change. Harv Bus Rev 2006 Dec;84(12):94-101, 163. [Medline: 17183796]

Abbreviations

CATI: computer-assisted telephone interviews

XSL•FO
RenderX

CHW: community health worker
IVR: interactive voice response
LAC: Latin America and the Caribbean
L2A: Listening to Africa
LMICs: low- and middle-income countries
MDGs: millennium development goals
mHealth: mobile health
MPS: mobile phone survey
NCD: noncommunicable diseases
SDGs: sustainable development goals
SMS: short message service
STEPs: WHO STEPwise Approach to Surveillance of Noncommunicable Diseases
UHC: universal health coverage
UN: United Nations
WHO: World Health Organization

Edited by E Rosskam; submitted 11.01.17; peer-reviewed by G Bloom, D Peiris; comments to author 21.02.17; revised version received 02.03.17; accepted 05.03.17; published 05.05.17.

Please cite as:
Pariyo GW, Wosu AC, Gibson DG, Labrique AB, Ali J, Hyder AA
Moving the Agenda on Noncommunicable Diseases: Policy Implications of Mobile Phone Surveys in Low and Middle-Income Countries
J Med Internet Res 2017;19(5):e115
URL: http://www.jmir.org/2017/5/e115/
doi:10.2196/jmir.7302
PMID:28476720

©George W Pariyo, Adaeze C Wosu, Dustin G Gibson, Alain B Labrique, Joseph Ali, Adnan A Hyder. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Ethics Considerations in Global Mobile Phone-Based Surveys of Noncommunicable Diseases: A Conceptual Exploration

Joseph Ali[1,2], JD; Alain B Labrique[2], MHS, MS, PhD; Kara Gionfriddo[1], MPP; George Pariyo[2], MBChB, PhD; Dustin G Gibson[2], MS, PhD; Bridget Pratt[2,3], MIH, PhD; Molly Deutsch-Feldman[4], MSc; Adnan A Hyder[1,2], MPH, MD, PhD

[1]Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, United States

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

[3]Nossal Institute for Global Health, School of Population and Global Health, The University of Melbourne, Melbourne, Australia

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

Corresponding Author:
Joseph Ali, JD
Berman Institute of Bioethics
Johns Hopkins University
Rm 208
1809 Ashland Avenue
Baltimore, MD, 21205
United States
Phone: 1 410 614 5370
Fax: 1 410 614 5360
Email: jali@jhu.edu

Abstract

Mobile phone coverage has grown, particularly within low- and middle-income countries (LMICs), presenting an opportunity to augment routine health surveillance programs. Several LMICs and global health partners are seeking opportunities to launch basic mobile phone–based surveys of noncommunicable diseases (NCDs). The increasing use of such technology in LMICs brings forth a cluster of ethical challenges; however, much of the existing literature regarding the ethics of mobile or digital health focuses on the use of technologies in high-income countries and does not consider directly the specific ethical issues associated with the conduct of mobile phone surveys (MPS) for NCD risk factor surveillance in LMICs. In this paper, we explore conceptually several of the central ethics issues in this domain, which mainly track the three phases of the MPS process: predata collection, during data collection, and postdata collection. These include identifying the nature of the activity; stakeholder engagement; appropriate design; anticipating and managing potential harms and benefits; consent; reaching intended respondents; data ownership, access and use; and ensuring LMIC sustainability. We call for future work to develop an ethics framework and guidance for the use of mobile phones for disease surveillance globally.

(J Med Internet Res 2017;19(5):e110) doi:10.2196/jmir.7326

KEYWORDS

ethics; mobile phone survey; mHealth; noncommunicable diseases; research ethics; bioethics

Introduction

Routine public health surveillance, a continuous process of collecting and analyzing health-related data, is critical for monitoring disease epidemiology and implementing public health programs [1]. Traditionally, active surveillance has often been carried out through face-to-face household surveys or contact with health care providers to obtain relatively reliable information [1,2]. This manner of data collection can be considerably resource-intensive [2]. In recent years, interest in alternative and streamlined approaches has grown [3]. In particular, some have begun to explore the use of basic mobile phone surveys (MPS) to augment data collection, anticipating several potential advantages over traditional methods, and leveraging existing and emerging mobile survey platforms [3,4]. MPS may provide opportunity for data collection more quickly, with fewer staff and lower overall programmatic costs; though approaches are only now being optimized and compared with traditional surveillance methodologies [5-7].

XSL•FO
RenderX

There is particular interest in using MPS in low- and middle-income countries (LMICs) where household surveys can be challenging to implement and frequent data collection is critical to effective monitoring of rapidly changing health behaviors and disease burdens [5]. Interest in MPS technology utilization in LMICs has also grown in large part due to the expansion of mobile phone use within such countries over the past decade. In sub-Saharan Africa, mobile phone ownership among individuals aged over 15 years reached 69% in 2015 [8]. Such increases present a considerable opportunity to access previously hard-to-reach populations; however, research has also documented some persistent inequities in mobile phone ownership, access, and use along axes of gender and income level, which may complicate the collection of representative data and make it difficult to ensure the equitable distribution of potential harms and benefits from research or surveillance [9-11].

In most wealthy countries where smartphones and reliable high-speed data networks are prevalent, mobile phone apps are also increasingly being used for health promotion, management, and surveillance [12]. However, the use of these technologies and approaches is more limited in LMICs where smartphone ownership and use is relatively low; in 2015, 37% of adults in low-income countries reported owning a smartphone, compared with 68% of adults in wealthy countries [8]. Challenges such as poor access to the Internet, cost, as well as lack of familiarity with technology limit the utility of smartphone-based health assessments across diverse LMIC populations [13,14]. Thus, emerging health surveillance programs in many LMICs typically have sought to conduct surveys using approaches conducive to simple cellular phones and networks, such as short message service (SMS), interactive voice response (IVR), and computer-assisted telephone interviews (CATI).

A relatively small number of studies have been conducted to determine the effectiveness of mobile phone–based disease surveillance in LMICs. These studies have focused on both infectious and chronic diseases, the latter of which is of particular interest as rates of noncommunicable diseases (NCDs) are on the rise within LMICs [15,16]. Researchers conducted pilot studies on the penetration of mobile phones and the use of mobile health (mHealth) tools for NCD surveillance and care in several countries, including Bolivia and South Africa [13,16]. Although many advocate for increased use of mHealth for NCD monitoring and care, rigorous research is necessary to inform practice. A research framework, developed by Bloomfield and colleagues [16] for advancing mHealth technology, to help address NCDs specifically in sub-Saharan Africa can help guide future activities as interest expands.

Additional recent efforts are ongoing to identify and navigate the corresponding ethics-related issues that arise when planning and implementing MPS in LMICs. Previous work has detailed several of the central ethics issues facing mHealth in general [11]. Vayena and colleagues [17] have specifically begun to map the ethical issues in “digital disease detection” or “digital epidemiology,” particularly when “big data” present the opportunity to aggregate digital information from multiple existing sources to, for example, identify potential disease outbreaks. However, much of the focus of emerging digital and mHealth health ethics literature is on the challenges associated with complex data systems, and the use of smartphones, tablets, and other more advanced mobile technologies.

Efforts have certainly also been made to guide “traditional” public health surveillance programs in LMICs, including within the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH) Network, a collaboration of health and demographic surveillance systems across 49 field sites in 20 countries [18]. INDEPTH has published a Resource Kit in an effort to advance best practices for surveillance; however, the kit only briefly touches on ethics considerations [19]. Thus, despite the growing use of mobile phone–based surveys and surveillance programs in LMICs, including for NCDs, in-depth analyses of relevant ethics challenges are quite limited.

With support from the Bloomberg Philanthopies’ Data for Health Initiative, we initiated a project to critically examine the ethical, legal, and societal issues associated with the use of MPS for NCD risk factor surveillance in LMICs [20]. The project involves activities to optimize mobile phone surveys and platforms in order to expand the capabilities of NCD risk factor data collection [4], presenting a unique opportunity to explore, both conceptually and empirically, the ethics challenges first hand.

In this paper, we provide an initial conceptual review of several central ethics challenges that ought to be considered when formulating and administering MPS of NCD risk factors in LMICs. The issues discussed in this paper reflect both key questions that have emerged during our own initial programmatic efforts, as well as those that are likely to become relevant as MPS efforts expand. The ethics issues raised in this manuscript were reviewed by diverse group of approximately 30 global experts in ethics, mHealth, social science, health policy, MPS technology, and regulatory oversight during a technical workshop. Workshop participants provided valuable feedback that helped to refine our analyses and focus efforts on the central challenges. The ethics-related issues that we discuss are mainly presented based on when they occur in the MPS process—before data collection, during data collection, and after data collection—acknowledging that some of the issues raised (such as stakeholder engagement and risk-benefit assessment) are cross-cutting (see Table 1). We discuss relevant considerations for each ethics issue, explore broader societal issues, and make recommendations for future work.

XSL•FO
RenderX

Table 1. Suggested ethics consideration in mobile phone survey (MPS) of noncommunicable disease (NCD) risk factors by phase.

Issues	Key considerations or questions	Available options (indicative)
Before data collection		
Defining the activity	Does the activity constitute research, M&E, and/or surveillance?	Nature of activity is defined by independent persons
	Who defines the nature of the activity?	Criteria to distinguish types of MPS[b] are established and applied
	What ethical requirements follow?	
	IRB[a] or other review?	
	Disclosures or authorizations?	
	Transformation of data into policy, practice?	
Stakeholder engagement	What are the ethical goals of engagement?	Identify local norms and practices associated with the use of mobile phones
	Who should be engaged?	Determine the social, cultural, legal, and public health significance of the information being collected
	How and when should stakeholders be engaged?	Identify stakeholders, for example, technical experts, community representatives, health system agents, donor agency representatives
Appropriate design	How should the survey be delivered? What qualities of survey delivery will affect the respondent and data validity?	Align delivery method with population characteristics, for example, if low literacy, IVR may be preferred
	What are the consequences of: IVR[c] or CATI[d] or SMS[e] or mixed modality?	
	Will the survey reach minority, disadvantaged, marginalized groups?	
	Does the delivery method influence social desirability bias?	Use all major languages commonly used in other traditional surveys
	Linguistic strategy: how many and which languages ought to be used to administer the survey?	
Anticipating potential harms and benefits	What are the potential burdens or risks of participating in a MPS program (at the individual, community, national levels)?	User-testing, interviews, and focus groups to identify key sensitive issues and design survey taking these into account
	For example, Sensitive information and stigma	
	What are the potential benefits or advantages of participating in a MPS program (at the individual, community, national levels)?	
	For example, More frequent, less costly information, and better NCD[f] care	
	What strategies will minimize harms and maximize benefits?	
	How will harms or benefits be balanced for disadvantaged and marginalized groups?	
During data collection		
Consent	What information about the MPS should be *disclosed* to respondents?	Disclose information consistent with *basic* consent elements
	How should respondents *authorize* their participation?	Review local practices, population preferences, and legal permissibility of different opt-in or opt-out approaches
	Should types of disclosure and authorization differ based on the MPS delivery method (IVR or CATI or SMS or mixed modality) and population?	
Reaching intended respondents	Which MPS sampling and recruitment methods best reduce unnecessary burdens upon intended and unintended MPS recipients?	Take stock of other existing survey approaches in designing MPS delivery strategies
	What approach to MPS incentive delivery is most appropriate?	Conduct follow up surveys by human caller on limited sample to verify

XSL•FO
RenderX

Issues	Key considerations or questions	Available options (indicative)
	Are existing laws likely to make it difficult to reach respondents using MPS?	Consult local regulatory authorities to identify and address potential areas of concern
After data collection		
Data ownership, access and use	To whom do the data belong?	Establish advance agreements to guide terms of data ownership, access, and use
	What data can be shared?	Consider restrictions on data access and use for purposes that do not advance public health
	Who should have access to the data?	
	How should the data be shared?	
	To what uses can the data be put?	
Ensuring LMIC[g] sustainability	What are different actors' responsibilities to promote and advance local sustainability?	Agree upon sustainability and capacity development commitments in advance
	Who is responsible for deciding who is allocated which strategies to undertake?	Identify core areas of local need and strength
	Should ongoing MPS efforts be supported only in the original data collection area or more broadly?	Integrate with existing programs and approaches where available

[a]IRB: institutional review board.

[b]MPS: mobile phone survey.

[c]IVR: interactive voice response.

[d]CATI: computer-assisted telephone interview.

[e]SMS: short message service.

[f]NCD: noncommunicable disease.

[g]LMIC: low- and middle-income country.

Before Data Collection

Defining the Activity

MPS are being rolled out across LMICs to determine their potential for collecting NCD risk factor information, with a long-term goal of generating reasonably reliable and valid population representative data. Therefore, in addition to collecting individuals' responses to questions about NCD risk factors such as diet, physical activity, tobacco or alcohol use and the like, the surveillance approach is itself being actively monitored, studied, and improved. Understanding the circumstances under which MPS activities constitute research, program monitoring and evaluation (M&E) and/or public health surveillance can be important to determining the nature of ethics and regulatory oversight typically required.

It can be difficult, however, to characterize the nature of particular MPS public health data collection activities, especially where multiple overlapping purposes exist and activities are implemented with significant involvement of many different types of stakeholders, including public health agencies and policy makers. Moreover, in contrast with research, ethical practice norms for M&E and surveillance are less established in general, let alone in the context of MPS. Therefore, even once MPS approaches for monitoring NCDs and other risk factors have been reasonably optimized through research and data can be considered valid, questions about ethics, and oversight requirements for ongoing public health use of MPS will likely remain. It is furthermore imperative to understand how the intent of any MPS, regardless of how surveyors define it, is perceived by respondents. Many mobile phone users may associate phone surveys with marketing and commercial data collection.

We suggest that defining the nature of the activity encompasses three central questions: (1) what features are relevant to determining whether a MPS activity constitutes research, M&E, and/or surveillance; (2) who defines the nature of the activity; and (3) what ethical requirements follow for each type of MPS activity [21]? Although it is beyond the scope of this paper to answer these questions fully, some believe that the proper identification of the nature of the activity is highly relevant to, among other things, determining whether the activity undergoes institutional review board (IRB) or other forms of prospective ethics review, the nature of disclosures and authorizations provided to and obtained from respondents, and the extent of the obligation to transform data into policy and practice [22]. On the other hand, others may argue that the tendency to link key ethics and regulatory oversight requirements to "labels" (such as research, M&E, and surveillance) can at times distract from the need to consider the nature of potential harms and benefits, and their distribution across society, irrespective of what the activity is called. It seems appropriate, particularly at early stages of MPS development and utilization for NCDs, that care be taken not simply to rely on interested parties to define the nature of applicable ethics and regulatory requirements. Independent consult from ethics and regulatory experts should

XSL•FO
RenderX

be sought, especially from within countries where MPS will be implemented.

Engagement With Local Stakeholders

Although MPS of NCD risk factors can be rolled out relatively easily and from a single or cyber location to the remotest regions of the globe, this does not in any way decrease the need for local engagement and indeed partnership. In fact, robust engagement seems imperative for the effective use of MPS for NCD risk factor surveillance in LMICs given the significant potential for over-utilization of MPS capabilities, sensitivity (legal, political, financial, and social) to perceived mobile phone "spam," and broader goals relating to local recognition and uptake of data. Three key questions related to stakeholder engagement in NCD MPS programs require attention: (1) what are the ethical goals of engagement in a MPS program; (2) who should be engaged and why; and (3) how and when should stakeholders be engaged?

On the *first* question, the *ethical goals* of MPS engagement may be conceived to include (1) understanding local norms and practices associated with the use of mobile phones and aligning approaches as best as possible with local expectations to minimally disturb respondents and other stakeholders such as mobile network operators; and (2) identifying the social, cultural, legal, and public health significance of the NCD risk factor information being collected to anticipate and mitigate avoidable informational risks and maximize potential benefits. Pursuit of these goals can help ensure local desirability, relevance, representativeness, and sustainability of the program. They imply a continuous engagement process inclusive of collaborative planning, implementation, and capacity strengthening.

It quickly becomes evident that attention to the *goals* of engagement in the context of a particular MPS effort is also critical to the process of identifying *who* to engage and *how* to do so. For example, ensuring *relevance* and *representativeness* of an NCD MPS effort would encompass the goals of deciding collaboratively on which diseases, and risk factors and from whom to gather data. The former would entail assessing which NCDs are of high burden or priority in the particular country or locale being surveyed and ensuring that the diseases surveyed included those of burden to disadvantaged and marginalized groups within the area. The latter would entail deciding what subpopulations to rely on to collect data and ensuring that they include better-off *and* disadvantaged areas or groups. If the population from which data are collected is not broadly representative or does not include sufficient numbers of different marginalized groups, the capability to identify the *distribution* of the burden of disease will be restricted. Key questions associated with access and demographics, such as who has the hardest time owning mobile phones, who has the hardest time accessing mobile phones, and what norms promote inequity of access to mobile phones, are crucial to answer in advance, when possible [11]. Those who are likely to be underrepresented by MPS can potentially be included through other more suitable means of data collection identified through direct engagement with those populations or their representatives. Both informal and formal engagement processes can serve these purposes.

Similarly, local *desirability* and *sustainability* considerations might require that individuals pursuing new MPS programs in low-resource settings also consider the local interest in and capacity to carry out the program. Are existing technologies or programs within the country being appropriately leveraged? What types of data for specific groups may be captured through MPS that have not been captured with household surveys or other surveillance methods? Who should coordinate and implement the program locally, alongside external or government actors? Is the program being externally imposed, using an infrastructure that cannot be supported with local resources?

The comparative ease with which technology and mobile devices can now be used to collect and transmit data has also sparked debate globally regarding the proper use of such technology and information for public and private purposes. Negative public sentiments toward electronic surveillance in other arenas (eg, for national security, intelligence or commercial purposes) may transfer to disease surveillance and require concerted efforts to overcome. This is a critical societal issue which requires broader discussions around the perceived and actual risks of collecting and using data acquired through information technology to improve the public's health. Establishing awareness campaigns and fora for open discussion may increase understanding and trust in MPS, particularly in locales where the approach is new or where it has previously been "abused." Addressing any such perceptions and questions is important for ensuring that the MPS program is accepted and operates for the benefit of all stakeholders.

Appropriate Design

It is important to consider the various survey delivery approaches (eg, SMS, IVR, and CATI) that can be used for MPS, and the possible consequences of each, particularly in terms of who they are likely to reach and whether this includes minority, disadvantaged, and marginalized groups within a particular country. As an ethics matter, any MPS ought to be conducted as efficiently and equitably as possible and in a manner that yields unbiased, reliable data. Therefore, one should prospectively consider which format of survey delivery is likely to provide such data in differing LMIC contexts.

SMS and IVR are likely to be lower cost and may reduce social desirability bias that can arise from face-to-face or phone interviews in which participants interact with other persons conducting the surveys [23]. IVR strategies may be more flexible and increase respondent comprehension compared with SMS, particularly in LMIC settings where literacy levels may be an issue [3]. However, both SMS and IVR risk selection bias (groups, such as the elderly, may be poorly represented due to lack of familiarity with the interfaces), and misclassification bias if questions are not explained fully or responses are not entered properly. CATI offers the opportunity for a structured questionnaire to be delivered more personably, that is, by a live person who can offer basic clarifications over the phone; however, it is generally more time- and resource-intensive and can suffer from interviewer bias and the previously mentioned social desirability bias. Additionally, both CATI and IVR may introduce complex and potentially biasing elements between

XSL•FO
RenderX

interviewer and respondent due to cultural or demographic differences as interpreted through voice and lexicon. Weighing the pros and cons of different types of survey delivery methods, including their accessibility to disadvantaged and marginalized groups, is important to the ethical goal of deploying MPS for the benefit of all, or at least not at the *systematic* disadvantage of some. Robust empirical comparisons of these modalities in the LMIC NCD context are therefore essential not only to the goals of surveillance, but also to comprehensive ethical analysis.

Similarly, all MPS programs, regardless of mode of delivery, must consider the proper languages and terminology to use. Selecting the best linguistic strategy is ethically important not only for obtaining representative data, but also for ensuring that participants understand all parts of the survey (including the reason why they are being surveyed). Many countries, particularly LMICs, have multiple languages and dialects; indeed, some countries like Indonesia and Nigeria have hundreds [24,25]. Often times "official" languages are most representative of dense urban areas where health burdens and behaviors differ from more rural areas. Variations between written and spoken language are also very common and the process of developing and verifying survey text and audio recordings in many different languages is potentially time- and resource-intensive. Although similar challenges may apply to traditional face-to-face household surveys of NCDs, the use of visual aids (show cards) and local data collectors with regional language skills can support data collection across varying linguistic areas. Where MPS are distributed using random digit dialing (RDD)—an approach that is of particular interest for nationally representative surveys given relative ease of implementation and statistical advantages—building in adequate linguistic representation into MPS is key.

Anticipating Potential Harms and Benefits

Anticipating potential harms or burdens and benefits or advantages is a critical part of MPS planning. This may involve different types of assessments, and the development of strategies to mitigate risk and maximize potential benefit. Those involved in collecting and storing data should always give due consideration to the risks (to individuals, communities, institutions, and nations) associated with the MPS approach and information being collected. Although the burden of participation may be relatively minimal, informational risks should be of concern to MPS for NCDs. For example, NCD risk factor data about alcohol consumption may be fairly sensitive in a country or region where consumption is culturally or religiously prohibited. Although certainly unacceptable to publicize individual-level data, even data showing particular communities to be associated with greater rates of alcohol consumption can also potentially generate harm.

Both those who have access to raw data and those who report findings have important data privacy and security-related responsibilities. MPS data are likely to be linked (at least in raw form) to individual phone numbers, even where a RDD approach is used. Whereas public telephone databases may not be available in many LMICs, mobile network operators and government authorities who may be engaged in MPS usually have access to sufficient information to potentially identify many individuals, though are unlikely to have reason to do so. Indeed, in some instances, for example, with data collected via SMS, both incoming and outgoing data itself may be automatically recorded by mobile network operators alongside personally identifying information. Planning is needed to develop protocols that define legitimate data use and protect against informational risk. Technical MPS platform developers or intermediaries should also be engaged to support alignment of software capabilities and data management practices with risk mitigation strategies.

Potential harms, however, should not be considered in the abstract and merely prospectively. They must be assessed in relation to potential *benefits*, and the *actual* accrual of harms and benefits must be monitored during MPS and evaluated afterwards. Potential population-level *benefits* of a well-designed and implemented MPS of NCD risk factors may include the opportunity for more frequent, less resource-intensive, and more convenient data collection across large geographic areas (not merely surveillance sites) to rapidly inform NCD policy and care. Another potential benefit of MPS, particularly IVR and CATI, may be its comparative advantage over face-to-face surveys in terms of respondent privacy. Other individuals associated with or in proximity of MPS respondents are unlikely to know the details or context of the survey unless respondents choose to make this known. Still, SMS surveys may suffer from privacy breaches due to the "written format" of the survey, unlike the digital and voice formats of IVR and CATI. Finally, MPS may yield better access to data from population subgroups that are harder to reach with household or other face-to-face survey methods.

Anticipating harms and benefits and evaluating the net risk of MPS therefore generally entails (1) identifying potential harms and benefits, and to whom they might accrue; (2) developing strategies prospectively to minimize harms and maximize benefits; and (3) assessing the balance of harms and benefits, again, including the balance for disadvantaged and marginalized groups. Although we do not describe processes for continuous monitoring and evaluation of harms and benefits in this manuscript, we emphasize its overall importance for MPS. Importantly, we neither believe that there is sufficient evidence to justify a *presumption* of net harm or benefit of MPS, nor do we think it wise to make such a generalization. Rather, the potential harms and benefits of each MPS program should be evaluated, ideally by independent personnel (eg, by an IRB) including individuals who represent those being surveyed. As part of this evaluation, where implementation strategies are unproven, the justification for testing the program in LMICs must also be assessed.

During Data Collection

Consent

The question of how to properly explain and obtain agreement to collect data using mobile technology has been raised in previous literature and is a potential challenge for MPS conducted in low-resource settings [11,26]. Here, we refer to a combination of basic *disclosure* and voluntary *agreement* as "consent" and distinguish it from a theoretical notion of

XSL•FO
RenderX

"informed consent" which typically requires in-depth explanation of the data collection activity and understanding on the part of respondents before voluntary agreement. We suggest that obtaining *basic consent* of respondents is likely to be acceptable (and most practicable) for most forms of NCD risk factor MPS.

Approval of key "gatekeepers," however, may also be important at a group level. For example, if not already centrally involved in initiating the MPS program, representatives from relevant governmental ministries and agencies (eg, Ministries of Health, Science and Technology, Telecommunications, Information Technology) should be fully briefed on the nature of the proposed MPS program and permitted to decide whether it aligns with relevant priorities, is allowed as a legal matter, would be an appropriate use of information communication systems and therefore whether it should proceed within the country. In instances where an MPS asks questions that are, for example, potentially stigmatizing for particular communities, or where data are only being collected from a more limited geographical zone, including local community representatives in decision-making is also likely important.

Although consensus has yet to form in defining the ethically required elements of consent *disclosure* for different types of MPS (eg, IVR, CATI, and SMS), a brief disclosure to respondents of essential MPS information might generally include the purpose, procedures, sponsor, key potential burdens, and benefits including expected duration and whether compensation (eg, airtime credits) will be provided, and the voluntary nature of the MPS. For multilingual MPS, it is important that this information be provided after respondents select a language of choice.

When indicating their *agreement* to participate, potential respondents can be requested to actively or passively o *pt-in*, or actively or passively *opt-out*. Examples of each are provided in Table 2. Of course, under all scenarios, the option to opt-out by not answering or hanging up the phone is available, at least for IVR and CATI. It remains unclear which approaches to authorization are *ethically* preferable for MPS of NCD risk factors and context may matter. If the various approaches to "demonstrating" agreement are determined to be roughly ethically equivalent in a particular context, then surveyors should follow the approach that is likely to yield the best response rate and thereby maximize the chance that data collected will be generalizable to the relevant population or subpopulation. This too has critical ethical importance as respondents are burdened unnecessarily when mHealth data sets are of little use.

Reaching Intended Respondents

NCD risk factor data can be collected using MPS in a matter of hours or days, and as frequently as necessary, at least in principle. If conducted too frequently from the same respondents, public distaste for MPS may increase to a point where adequate completion rates are unachievable. Regardless of frequency, it may be difficult to ensure that respondents actually meet inclusion criteria. This is of particular concern for RDD where anyone with a mobile phone number of a particular prefix can potentially receive a survey. For example, it is conceivable that an adolescent might respond to a MPS of adult risk factors.

It is ethically important to survey administration not to burden individuals unnecessarily. With MPS, one must account for individuals who have moved out of an intended survey locale, but are unintentionally sampled because they have a phone number prefix for the intended area. Furthermore, LMIC mobile networks often have quality and cost issues leading many individuals to obtain subscriber identity module (SIM) cards or phone numbers for multiple networks. This makes it possible that some may receive and perhaps respond to a survey multiple times. Many of these challenges could also cause problems primarily for data reliability and validity, but with careful design, sampling, and statistical analyses, most of these challenges can be managed, as a technical matter [27].

In order to encourage survey completion and accommodate any financial burdens, it is increasingly customary to provide respondents with phone airtime credits. Although many have debated whether and when such incentives can be excessive or inadequate, considerations relating to the timing of incentive delivery may also raise ethical and data quality considerations. Incentive delivery upon MPS completion may increase the likelihood that individuals complete the entire survey; however, some may provide false data, for example, indicate a false age or press random digits, to reach completion. This approach also does not accommodate those who, for reasons not under their control, were only able to partially complete a survey. Providing some airtime credits to all respondents regardless of survey completion status may mitigate these risks and be "fairer," but will likely result in greater overall expense and perhaps decrease the chance that respondents will answer all questions. It remains to be determined empirically whether these particular concerns are, in general, of greater significance to MPS as compared with many face-to-face surveys. Certainly any known measures to reduce these complications should be used.

It may be more difficult to "design around" other issues which have ethics or regulatory implications. For example, it may be challenging to roll out a MPS using RDD in a country where telecommunication law prohibits "robo-calling" except when done by certain agents for the purpose of public service and safety messaging. Other countries may prohibit "masked" or "restricted" phone calls where a call is made through an intermediary and the caller remains hidden. Data collectors may need to work with government authorities in such cases to identify workarounds that still honor the principles embodied in the restriction, for example, establish a call-back number that provides an automated "hotline" with additional information about the MPS.

XSL•FO
RenderX

Table 2. Mobile phone survey (MPS) consent authorization options and examples.

Options	Opt-in	Opt-out
Active	"press 1 if you would like to continue the survey"	"press 2 if you do *not* want to complete the survey"
Passive	"by completing this survey you agree to participate"	"the survey will automatically end if you do not respond to a question within 1 min"

After Data Collection

Data Ownership, Access, and Use

Perhaps the most important postdata collection question relating to the conduct of MPS for NCD risk factor surveillance is, "What will happen to NCD risk factor data once collected?" This can be reduced to at least five component questions: (1) to whom do the data *belong*, (2) *what* data can be shared, (3) who should have *access* to the data, (4) *how* should data be shared, and (5) to what *uses* can the data be put? Most of these questions can be deliberated, agreed upon, and clearly documented in advance in memoranda of understanding or contracts. However, where MPS programs involve multinational donors, LMIC and high-income country teams, and government agencies, such negotiations are likely to have ethical undercurrents relating to power imbalances and the abilities of particular partners to meaningfully engage in contract negotiation. Efforts to guide fair contracting, such as the Council on Health Research and Development (COHRED) Fair Research Contracting initiative may serve as useful references for MPS partnerships [28]. Data sharing agreements and models available from other public health surveillance systems can perhaps also serve as a guide for MPS agreements [29,30]. However, MPS may also give rise to additional opportunities and challenges for data sharing and use.

With MPS, large amounts of electronic data can, in theory, be aggregated, cleaned, analyzed, and shared relatively quickly. Opportunities to validate MPS NCD data against equivalent data should also be sought out. The degree of confidence in the quality and significance of the data will likely serve as an initial filter on what is shared and with whom. When data are found to be of sufficient quality, an obligation likely exists to feed the data back into the local health system so that there is an opportunity for it to inform priority setting and resource allocation. This raises questions such as, Should data be disseminated to all levels of the health system or just national authorities? Should they be disseminated to public and private actors in the health system? Should they be disseminated to those who responded to the MPS or the public more broadly? Do the World Health Organization or other global actors or donors have claims to information access and use?

Although perhaps primarily meant to inform the development of local interventions and policies to reduce NCDs, given the large amount of behavioral data that are likely to be generated through MPS of NCD risk factors, it is expected that various additional groups will be interested in obtaining and learning from the data. Nongovernmental organizations, insurers, large employers, urban planners, biotechnology companies, food manufacturers and distributors, and even alcohol and tobacco companies, to name a few, may all be eager to learn from NCD data and put them to uses that may or may not benefit the public's health [30]. Should data stewards not plan and act in ways that demonstrate careful management of information collected, respondents or groups in LMICs ultimately may demand the ability to opt-in or opt-out of *particular* uses in the future—what is known as "specific consent"—or may refuse to provide their health information altogether [31]. Building and maintaining public trust in MPS is therefore highly critical to realizing the long-term potential of the approach globally. A review of experiences, challenges, and emerging best practices related to genetics and genomics research, biobanking and data sharing in LMICs may provide useful transferable lessons.

Finally, several additional questions relating to the use of data to support intervention and policy decisions are worthy of consideration. How should MPS findings factor into health priority-setting and resource allocation decisions made by policy makers in the context of other data and relevant considerations? Do host country actors (eg, ministries of health, district health officials) have an obligation to use the information generated from MPS surveillance to set or revise priorities and resource allocations within national and district health systems? For example, if the MPS identifies certain high NCD burdens for disadvantaged groups that are very poorly resourced within the district health system relative to other diseases, do district health officials have an obligation to shift resources to those diseases? Even if policy makers have an obligation to use MPS data to help set priorities, burden of disease is not the only (ethical) consideration in health priority-setting and resource allocation; there will be opportunity costs and other considerations involved in shifting resources. The question of how MPS data on NCD risk factors can and should be used by policy makers (and others) is, thus, quite complex and will entail consideration of many contextual features [32]. Where relevant, MPS data should at least feed into existing data-to-policy mechanisms that a given country may have.

Ensuring Low- and Middle-Income Country (LMIC) Sustainability

We underscore the importance of local sustainability of MPS programs. Mobile phone penetration may be rapidly increasing in many LMICs, but the resources and technical capacity needed to independently develop, conduct, and analyze MPS are lagging behind due in part to the many challenges identified here and elsewhere [3]. To be sure, several countries have initiated their own health-related MPS programs with varying degrees of success and longevity. Those engaged in these programs are urged to continue to share case studies describing what is working, what is not, and what is needed to advance sustainability. Assessments of the ethical, legal, and societal dimensions of these programs are welcomed to better understand all contextual factors.

XSL•FO
RenderX

Implementing an unsustainable program is ethically questionable, as is the failure to consider sustainability aspects [33,34]. Determining the concrete contributions that all involved parties can make to promote and support sustained implementation is an important part of ensuring the ethical conduct of MPS. Again, this ethical dimension raises numerous unanswered questions: What are different actors' (funders, survey implementers, local partners, ministries of health, etc) responsibilities to promote and advance sustainability? Do survey implementers have an obligation to initiate discussions with key stakeholders (at national, district, and local levels) about how the program can be sustained and to identify sustainability strategies (eg, advocacy, fund raising, capacity strengthening) in collaboration with those stakeholders? Do different actors then have a responsibility to carry out those strategies and, if so, on what basis? Who should be responsible for deciding who is allocated which strategies to undertake? Ought the program to be sustained in the areas where the initial surveillance implementation is being done or more broadly? These considerations, while easy to overlook, are vital to the success of public health information systems in LMICs.

Recommendations and Conclusions

We have discussed many considerations in this paper which are in need of further conceptual and empirical exploration. These include identifying the nature of the activity; stakeholder engagement; appropriate design; anticipating and managing potential harms and benefits; consent; reaching intended respondents; data ownership; access and use; and ensuring LMIC sustainability. Identification of the degree to which existing ethics guidance in other arenas (eg, research ethics, generally) might support the navigation of ethics challenges associated with MPS for NCD risk factor surveillance in LMICs may be helpful. However, given the current lack of comprehensive ethics guidance for public health surveillance and for MPS in LMICs, several of the issues outlined above are likely to require fresh consideration.

Although we focus mainly on ethics issues in this paper, there is a need for a broad conceptual framework for the ethical, legal, and societal issues associated with MPS for NCD risk factors. Such a framework ought to include thorough analysis of various types of MPS activities (eg, research, M&E, and surveillance) and delivery methods (eg, IVR, CATI, and SMS). It would also likely benefit from empirical testing through application to ongoing MPS in LMICs. Empirical efforts to capture a cross-section of stakeholder perceptions relating to the identified challenges and any additional means to address them in practice would be of additional value.

Practical guidance relevant to the various stakeholders involved in designing, implementing, reviewing, funding, and overseeing MPS could then be formulated and updated as norms continue to develop and technological capabilities advance. It would be particularly useful for guidance documents to identify key issues, outline pros and cons of options available to stakeholders for each issue, review additional points to consider, and, provide references to resources relevant to each issue. In order to begin to address these needs, we hope to establish a global working group inclusive of experts in ethics, mHealth survey implementation, regulatory oversight and policy, public health, social science, and MPS platform development. We welcome opportunities to move forward in addressing these emerging issues collaboratively.

Acknowledgments

We would like to thank workshop participants for their critical feedback on many of the issues discussed in this manuscript. This project is financially supported by the Bloomberg Philanthropies under the Data for Health Initiative (Grant No. 119668). The funding agency had no role in the preparation of this manuscript.

Conflicts of Interest

None declared.

References

1. Nsubuga P, White M, Thacker S, Anderson M, Blount S, Broome C, et al. Public health surveillance: a tool for targeting and monitoring interventions. In: Jamison D, Breman J, Measham A, Alleyne G, Claeson M, Evans D, et al, editors. Disease control priorities in developing countries, 2nd edition. New York: Oxford University Press; 2006:a.
2. Croke K, Dabalen A, Demombynes G, Giugale M, Hoogeveen J. Collecting high frequency panel data in Africa using mobile phone interviews. World Bank Policy Research Working Paper 2012;1:1-27. [doi: 10.1596/1813-9450-6097]
3. Gibson DG, Pereira A, Farrenkopf BA, Labrique AB, Pariyo GW, Hyder AA. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. J Med Internet Res 2017;19(5):e139 [FREE Full text] [doi: 10.2196/jmir.7428]
4. Hyder A, Wosu A, Gibson D, Labrique A, Ali J, Pariyo G. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246]
5. Ballivian A, Azevedo J, Durbin W. Using mobile phones for high-frequency data collection. In: Toninelli D, Pinter R, de Pedraza P, editors. Mobile research methods: opportunities and challenges of mobile research methodologies. London: Ubiquity Press; 2015:e.

XSL•FO
RenderX

6. Moura EC, Claro RM, Bernal R, Ribeiro J, Malta DC, Morais Neto O. A feasibility study of cell phone and landline phone interviews for monitoring of risk and protection factors for chronic diseases in Brazil. Cad Saúde Pública 2011 Feb;27(2):277-286. [doi: 10.1590/S0102-311X2011000200009]
7. Mahfoud Z, Ghandour L, Ghandour B, Mokdad AH, Sibai AM. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? Field Methods 2014 Jul 01;27(1):39-54. [doi: 10.1177/1525822X14540084]
8. Handjiski B. The Brookings Institution: Future Development. Mobile connectivity in Africa has already arrived URL: https://www.brookings.edu/blog/future-development/2015/03/18/mobile-connectivity-in-africa-has-already-arrived/ [accessed 2016-11-28] [WebCite Cache ID 6mMJwCIDl]
9. Tran MC, Labrique AB, Mehra S, Ali H, Shaikh S, Mitra M, et al. Analyzing the mobile "digital divide": changing determinants of household phone ownership over time in rural bangladesh. JMIR Mhealth Uhealth 2015 Feb 25;3(1):e24 [FREE Full text] [doi: 10.2196/mhealth.3663] [Medline: 25720457]
10. GSMA. Bridging the gender gap: mobile access and usage in low-and middle-income countries URL: http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2016/02/Connected-Women-Gender-Gap.pdf [accessed 2016-12-19] [WebCite Cache ID 6msHSuJhF]
11. 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]
12. Krebs P, Duncan DT. Health app use among us mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
13. Kamis K, Janevic MR, Marinec N, Jantz R, Valverde H, Piette JD. A study of mobile phone use among patients with noncommunicable diseases in La Paz, Bolivia: implications for mHealth research and development. Global Health 2015 Jul;11:30 [FREE Full text] [doi: 10.1186/s12992-015-0115-y] [Medline: 26141528]
14. Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. Heterogeneous mobile phone ownership and usage patterns in Kenya. PLoS One 2012 Apr;7(4):e35319 [FREE Full text] [doi: 10.1371/journal.pone.0035319] [Medline: 22558140]
15. Davies J, Yudkin JS, Atun R. Liberating data: the crucial weapon in the fight against NCDs. Lancet Diabetes Endocrinol 2016 Mar;4(3):197-198. [doi: 10.1016/S2213-8587(16)00037-1]
16. Bloomfield G, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez E. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014;10:49.
17. 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]
18. INDEPTH Network. Vision, mission and strategic objectives URL: http://www.indepth-network.org/about-us/vision-mission-strategic-objectives-0 [accessed 2017-03-23] [WebCite Cache ID 6pAl4EF9Y]
19. INDEPTH Network. INDEPTH resource kit for demographic surveillance systems URL: http://www.indepth-network.org/Resource%20Kit/INDEPTH%20DSS%20Resource%20Kit/INDEPTH%20DSS%20Resource%20Kit.htm [accessed 2016-12-20] [WebCite Cache ID 6mu1xouY4]
20. Bloomberg Philanthropies. Data for Health URL: https://www.bloomberg.org/program/public-health/data-health/ [accessed 2016-11-28] [WebCite Cache ID 6mMO3YKjI]
21. Petrini C, Ricciardi G. Ethical issues in public health surveillance: drawing inspiration from ethical frameworks. Ann Ist Super Sanita 2015;51(4):270-276 [FREE Full text] [doi: 10.4415/ANN_15_04_05] [Medline: 26783212]
22. Rhodes R, Baumrin S, Blaser M, Earle W, Indyk D, Jabs E, et al. Public health research on populations. In: Rhodes R, Gligorov N, Schwab AP, editors. The human microbiom: ethical, legal and social concerns. New York: Oxford University Press; 2013:208-246.
23. Gordis L. Epidemiology, 5th edition. Philadelphia: Elsevier Saunders; 2013.
24. United States Central Intelligence Agency. The World Factbook: Indonesia: people and society URL: https://www.cia.gov/library/publications/the-world-factbook/geos/id.html [accessed 2017-03-13] [WebCite Cache ID 6owQuLope]
25. United States Central Intelligence Agency. The World Factbook: Nigeria: people and society URL: https://www.cia.gov/library/publications/the-world-factbook/geos/ni.html [WebCite Cache ID 6mu16O1Wo]
26. Parker M. Ethical considerations related to mobile technology use in medical research. J Mob Technol Med 2012 Sep 12;1(3):50-52. [doi: 10.7309/jmtm.23]
27. Labrique AB, Blynn E, Ahmed S, Gibson DG, Pariyo GW, Hyder AA. Health surveys using mobile phones in developing countries: automated active strata monitoring and other statistical considerations for improving precision and reducing biases. J Med Internet Res 2017;19(5):e121 [FREE Full text] [doi: 10.2196/jmir.7329]
28. Council on Health Research for Development. Fair research contracting URL: http://www.cohred.org/frc/ [accessed 2016-11-28] [WebCite Cache ID 6mMOse9mI]
29. United States Agency of International Development. ADS 579 USAID's policy on development data URL: https://www.usaid.gov/sites/default/files/documents/1868/ADS579FactSheet%202015-02-13.pdf [WebCite Cache ID 6mu28HRbk]
30. Carolan L. Publishing.service. Open data, transparency and accountability: Topic guide URL: https://assets.publishing.service.gov.uk/media/5857fdcb40f0b60e4a0000d6/OpenDataTA_GSDRC.pdf [accessed 2017-03-13] [WebCite Cache ID 6owR58vbu]

XSL•FO
RenderX

31. National Institutes of Health, National Human Genome Research Institute. Genome. Special considerations for genome research URL: https://www.genome.gov/27559024/informed-consent-special-considerations-for-genome-research/ [accessed 2016-12-20] [WebCite Cache ID 6mu0sEZoC]
32. Pariyo GW, Wosu AC, Gibson DG, Labrique AB, Ali J, Hyder A. Moving the agenda on noncommunicable diseases: policy implications of mobile phone surveys in low- and middle-income countries. J Med Internet Res 2017;19(5):e115 [FREE Full text] [doi: 10.2196/jmir.7302]
33. Hyder AA, Pratt B, Ali J, Kass N, Sewankambo N. The ethics of health systems research in low- and middle-income countries: a call to action. Glob Public Health 2014 Aug;9(9):1008-1022 [FREE Full text] [doi: 10.1080/17441692.2014.931998] [Medline: 25104051]
34. London AJ, Borasky DA, Bhan A, Ethics Working Group of the HIV Prevention Trials Network. Improving ethical review of research involving incentives for health promotion. PLoS Med 2012 Mar;9(3):e1001193 [FREE Full text] [doi: 10.1371/journal.pmed.1001193] [Medline: 22479154]

Abbreviations

CATI: computer-assisted telephone interviews
COHRED: Council on Health Research and Development
INDEPTH: International Network for the Demographic Evaluation of Populations and Their Health
IRB: institutional review board
IVR: interactive voice response
LMICs: low- and middle-income countries
mHealth: mobile health
MPS: mobile phone survey
NCDs: noncommunicable diseases
RDD: random digit dialing
SIM: subscriber identity module
SMS: short message service

Edited by E Rosskam; submitted 14.01.17; peer-reviewed by N Unciano, E Bunker; comments to author 07.02.17; revised version received 28.02.17; accepted 28.02.17; published 05.05.17.

Please cite as:
Ali J, Labrique AB, Gionfriddo K, Pariyo G, Gibson DG, Pratt B, Deutsch-Feldman M, Hyder AA
Ethics Considerations in Global Mobile Phone-Based Surveys of Noncommunicable Diseases: A Conceptual Exploration
J Med Internet Res 2017;19(5):e110
URL: http://www.jmir.org/2017/5/e110/
doi:10.2196/jmir.7326
PMID:28476723

©Joseph Ali, Alain B Labrique, Kara Gionfriddo, George Pariyo, Dustin G Gibson, Bridget Pratt, Molly Deutsch-Feldman, Adnan A Hyder. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Policy Proposal

Health Surveys Using Mobile Phones in Developing Countries: Automated Active Strata Monitoring and Other Statistical Considerations for Improving Precision and Reducing Biases

Alain Labrique[1], MHS, MS, PhD; Emily Blynn[1], MPH, MBA; Saifuddin Ahmed[2], PhD; Dustin Gibson[1], PhD; George Pariyo[1], PhD; Adnan A Hyder[1], MD

[1]Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD, United States

[2]Johns Hopkins Bloomberg School of Public Health, Department of Population, Family and Reproductive Health, Baltimore, MD, United States

Corresponding Author:
Alain Labrique, MHS, MS, PhD
Johns Hopkins Bloomberg School of Public Health
Department of International Health
W5501, Johns Hopkins Bloomberg SPH
615 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 4102361568
Fax: 1 4105101055
Email: alabriqu@jhsph.edu

Abstract

In low- and middle-income countries (LMICs), historically, household surveys have been carried out by face-to-face interviews to collect survey data related to risk factors for noncommunicable diseases. The proliferation of mobile phone ownership and the access it provides in these countries offers a new opportunity to remotely conduct surveys with increased efficiency and reduced cost. However, the near-ubiquitous ownership of phones, high population mobility, and low cost require a re-examination of statistical recommendations for mobile phone surveys (MPS), especially when surveys are automated. As with landline surveys, random digit dialing remains the most appropriate approach to develop an ideal survey-sampling frame. Once the survey is complete, poststratification weights are generally applied to reduce estimate bias and to adjust for selectivity due to mobile ownership. Since weights increase design effects and reduce sampling efficiency, we introduce the concept of automated active strata monitoring to improve representativeness of the sample distribution to that of the source population. Although some statistical challenges remain, MPS represent a promising emerging means for population-level data collection in LMICs.

(J Med Internet Res 2017;19(5):e121) doi:10.2196/jmir.7329

KEYWORDS

surveys and questionnaires; sampling studies; mobile health; mobile phone; research methodology

Introduction

Since the filing of Alexander Graham Bell's patent for the telephone in 1876, voice, and eventually, data communications networks have transformed the globe. Hard-wired landline infrastructure was a necessary developmental milestone for communities entering the modern era, rapidly connecting populations across high-income countries and most urban centers of the developing world [1]. By the 1990s, nearly every home in the United States had a fixed landline phone, which was used by national statistical agencies, like the US Centers for Disease Control (CDC) or the Census Bureau, and by polling organizations for conducting household surveys [2].

Until the first mobile phone was introduced in the early 1970s, there was no challenge to the role of the landline telephone as a tool for population-level data collection. As the global mobile phone revolution exploded in the early 2000s, a dramatic shift from landline to cellular networks began to occur. According to the CDC, by early 2005, only 7.3% of US households had shifted to mobile as their only phone connection [3]. By the end of 2015, a little over a decade since the agency began meticulously tracking household-phone ownership, only 8% of homes reported exclusive landline phone access, with 44% of

XSL•FO
RenderX

homes reporting a mobile phone being their only communication access [3]. This transformation has been even more dramatic in the developing world, where landline infrastructure has been leapfrogged by the rapid deployment of mobile networks and affordable cellular telephony [1].

This transition to ubiquitous mobile phone access around the globe has had an important effect on population surveys especially in low- and middle-income countries (LMICs) where the availability of landline phone was rarely universal and surveys were conducted by face-to-face interviews (F2F). However, F2F surveys are expensive, time consuming, and often difficult to conduct in remote or conflict regions. Mobile phone surveys (MPS) are likely to reduce these challenges. In fact, several global agencies and survey firms have begun to leverage mobile phone coverage rates to collect data at random or from panels of respondents [4].

In this paper, we identified some of the key statistical considerations and challenges associated with each stage of mobile-only surveys in LMICs. We propose some novel methodological approaches for improving population representativeness and efficiency of MPS (see Table 1).

Table 1. Key mobile phone surveys (MPS) considerations by survey phase.

Phase	Key considerations	Mitigation
Presampling	Differences between phone owners and nonphone owners	Decreases as mobile penetration increases in LMICs; survey of nonphone owners can help understand bias
Sampling and survey execution	Source of numbers to sample from; obtaining a representative sample; multiple phone or SIM card ownership	Prescreened "valid numbers only" bank of numbers; random digit dialing; automated active strata monitoring
Postsampling	Residual differences between phone owners and nonphone owners; residual differences between respondents and nonrespondents; residual differences between single- and multiple-phone owners	Postsampling weighting

Presampling challenges for MPS

Redefining the Concept of the Sampling Frame

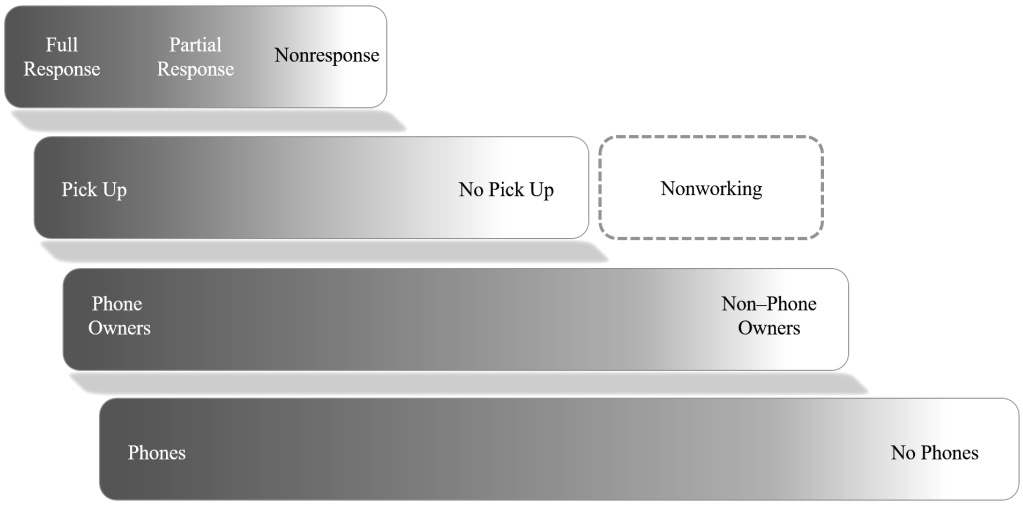
As the level of mobile phone access and ownership reaches saturation (100%, or at least one mobile phone per eligible adult respondent) at a population level, it becomes plausible to consider the entire population of phone owners as elements of a "sampling frame" for an MPS [5]. Although some countries have reached universal level of mobile penetration, most of them have not, thus resulting in some degree of sample misalignment when comparing the theoretical sampling frame to the population at large.

As shown in Figure 1 (adapted from [6]), as country-level mobile penetration and ownership increases, the amount of "white space" in the stratum, representing geographies without mobile phones, should decrease. An MPS begins with the intended target frame of the MPS as the entire population of mobile phone owners, without restriction, but still a subset of the total population until such time that a population is 100% covered (where all eligible respondents have at least 1 phone). In most cases, it is difficult to characterize the relative size and makeup of the non–phone owners, unless representative population data are available on determinants of phone ownership or surveys of characteristics of these two populations are available (or can be conducted by the researchers). It is also important to note that, as the cost of device ownership and airtime decreases, the socioeconomic composition (and associated risk factor behaviors) of phone owners and nonphone owners is also likely to change. These secular trends should be considered when comparing the results of MPS across multiple time points.

In the third layer, we see how this subset of phone owners might respond, by picking up or not, and also a new section, representing numbers that are nonexistent—an artifact of the random digit dialing process, discussed in detail below. The fourth layer depicts the possible response behaviors of the subset of those phone owners who do pick up, with the gradient representing the different possible outcomes of the respondent interaction.

XSL•FO
RenderX

Figure 1. Statistical challenges for each phase of MPS. The figure illustrates the challenges in capturing specific denominators in MPS, where the bottom layer is the complete population, which comprises phone owners and non–phone owners. A cascading gradient has been used to depict the uncertainty and variability, by setting, in the proportions of either group at each layer. Various types of loss and attrition, from nonresponse to invalid numbers, reduce the total number of units sampled successfully and completely.



Differences Between Individuals Who Own and Do Not Own Phones

As populations transition from either fixed or no phone access to widespread cellular networks, heterogeneity in coverage and household ownership can vary across states or districts. These differentials are likely to mirror socioeconomic and rural-urban gradients and could hinder statistically sound estimates of population characteristics and behavior if not carefully understood and managed [3]. Differences between landline and mobile phone (or combined household) respondents have been documented extensively across multiple populations over the past decade [7,8]. Many previous discussions around statistical methods and sampling strategies to mitigate the risks of bias and threats to validity have focused on high-income country settings, in the background of a strong landline or mixed-network legacy systems [9,10]. In most LMICs, landline infrastructure has remained underdeveloped or limited to larger metropolitan areas, leaving mobile phone "only" households as the predominant and more universal denominator. These mobile-only populations remain poorly characterized in most settings, with gaps in ownership/access, sometimes characterized as a facet of the "digital divide," aligned with lower socioeconomic strata or marginalized populations [11]. In some countries of sub-Saharan Africa, where populations are sparse and distributed across a large geographic spread, extreme remoteness and cellular network unavailability may also be determinants of respondent unreachability via MPS.

Sampling Strategies for MPS

Identifying a Sampling Frame and Dealing With Invalid numbers

Apart from addressing challenges in the differences between phone users and nonphone users in a population, the next important hurdle for any phone survey methodology begins with the development of a representative and valid sampling frame, broadly defined as the set of eligible numbers from which the sample of telephone households is selected [9]. The sampling frame has been a critical component of traditional surveys as a major focus of survey representativeness and costs [9]. For telephone surveys, three approaches to sampling frame development have traditionally been used such as random digit dialing (RDD), list-assisted design, and a multiple-frame sampling method that combines the two approaches.

Unlike telecom companies in many high-income countries, in LMICs, providers seldom maintain (or publish) directories of active mobile phone users and their numbers [7]. Without accurate published or network-operator provided lists to select from, RDD remains the most appropriate sampling approach. In landline surveys conducted in the United States, sampling frames are often developed by obtaining known prefixes to which random suffixes are appended. In many countries, mobile network operators (MNOs) are provided prefixes different from each other (eg, 019, 017) by the government, followed by a fixed set of numbers, which helps minimize ineligible random number combinations. As the populations using different MNOs may differ in geographical distribution and other characteristics (eg, income or education level, based on market segments targeted for low cost, entry-level prepaid plans), using known MNO prefixes also helps to ensure proportionate representation in the sample of the market share held by different operators. On the basis of this information, a large pool of possible numbers, an order of magnitude larger than the necessary sample, can be computer generated as the first "pool" of numbers to dial randomly. New pools can be generated, with care taken not to re-create numbers, if a pool is depleted. Although some network operators provide specific feedback (in the form of digital signals) when an invalid number is dialed, this practice is not universal.

Due to invalid sequences that are essentially unavoidable in an RDD, the required sample size should be inflated by dividing with a factor 1-Y, which estimates the proportion of nonworking numbers, to yield the number of random mobile phone numbers to be generated. This factor can be determined by first creating a smaller "test" pool of numbers to determine a likely proportion

XSL•FO
RenderX

of real numbers through a practice RDD round. As the numbers are dialed, if a number is identified as a nonhousehold or definitively found to be nonworking on the first call, then it should be excluded and the next one be dialed. If a call is not answered, it can be redialed a predetermined number of times before it is identified as nonworking and replaced with a new number from the list.

Accelerated Sequential Replacement

To ensure that correct sample size is obtained, traditional landline surveys have employed a process called accelerated sequential replacement. This is an iterative approach that selects numbers from a purposefully expanded sample of random numbers and replaces those definitively identified as nonworking at the end of a given operational stage, such as the day or shift. Three statuses are usually assigned to the outcome of a call such as verified household, verified nonhousehold or nonworking, and unresolved (eg, no answer or strange noise, but not verified as nonworking). Before the beginning of the next stage, those numbers that have been verified as nonworking are replaced with an equal number of new random numbers from the randomly generated list. After a predetermined number of calls, unresolved numbers are assumed to be nonworking and are also replaced. Although traditionally this was a manual process, now it can be easily automated and invalid numbers can be replaced automatically, instead of doing in stages, thereby reducing the total time required to conduct the survey.

An MPS Approach to Quota-Driven Sampling: Automated Active Strata Monitoring

The relatively low cost and automated nature of most MPS technologies, combined with the vast size of the denominator being selected from—effectively every mobile phone owner in the population with an active subscription, connected to a network—allows us to consider an approach that continues to attempt to fill a particular target demographic stratum of a population until that stratum's desired sample size is reached. In this quota-driven sampling procedure, a priori "sample size" is determined with a known statistical precision level, and the sample is selected from an RDD list through a probabilistic sampling procedure. Differences between the composition of the general population and the population of phone owners, in terms of gender, age, socioeconomic status, urban versus rural residence, and geographic origin can be mitigated through establishing target quota, based on the relative proportion of individuals of a particular combination of relevant characteristics in the population at large. Recent census data can be used to assess strata-specific population distribution, or in case where data from a recent census is not available, information from the most recent demographic and health surveys (DHS) may be used. The DHS are conducted in over 90 LMICs and provide nationally representative data on these strata-specific population distributions.

Given the digital nature of MPS, real-time data streams can be monitored, and strata actively "closed" once the required sample size for that subgroup has been met. This process can be automated or monitored by study implementers. The concept, automated active strata monitoring (AASM) also allows many more strata to be chosen to minimize the possible effects of nonprobabilistic sampling from the parent population. In this process, when a participant answers the phone, the first survey questions should establish their demographic, education, and other sociodemographic information of interest to determine stratum contribution. If the required number of targeted respondents has already been reached in their stratum, no further questions are asked, and they will be excused from completing the survey. Conversely, if more respondents are still required in their stratum, they are led through the survey questions.

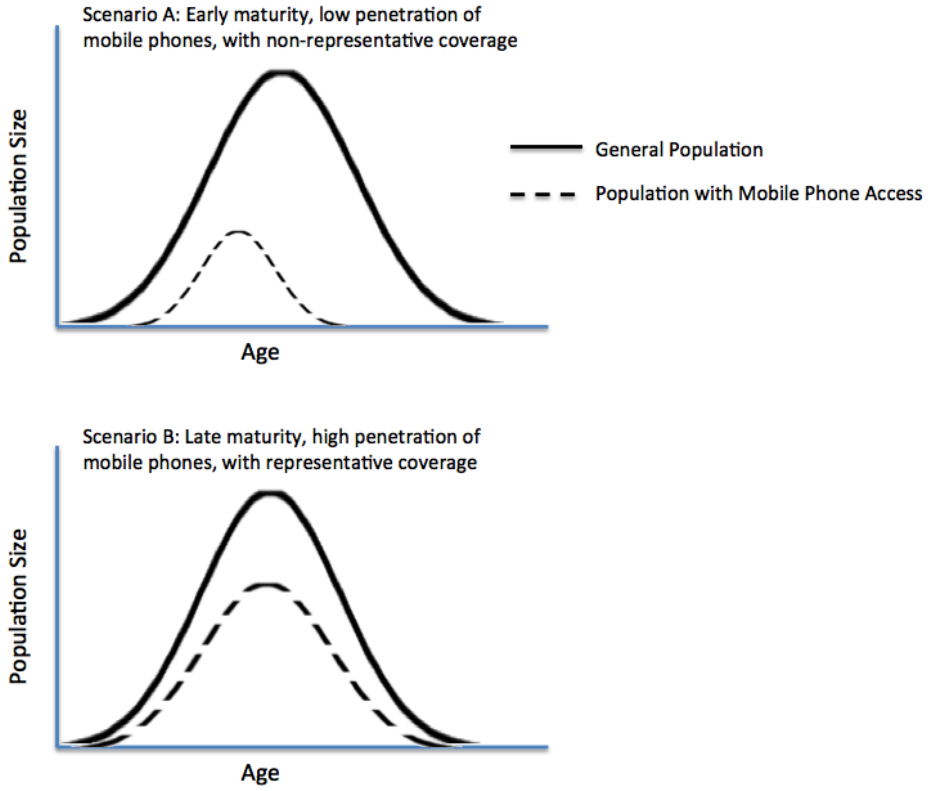
AASM is not plausible in traditional household surveys, as the marginal cost and time required to visit more households, in the effort to complete one or more unfilled strata, becomes prohibitively expensive. With the extremely low cost of MPS, in contrast to traditional F2F methods, and high levels of mobile phone coverage, for the first time in history, the survey denominator is theoretically, in many cases, the entire population (see Figure 2). For example, if the population of mobile phone owners between the ages of 60 and 70 years is a small percentage of the total population, even under less-than-complete mobile phone saturation, the denominator of phone-owning individuals in the population-at-large may number in the tens of thousands or millions. Although a greater number of calls may be required to enroll a sufficient number of these relatively "smaller" population strata, the cost and feasibility of doing so through MPS is much more reasonable. Using AASM, much more granular stratification can be achieved with MPS, ensuring greater representativeness and minimizing potential selection bias due to nonrepresentativeness in select population strata, common with small sample-frame surveys.

For traditional survey methods, an inherent disadvantage of quota-driven sampling is that it is a nonprobabilistic approach, and although certain characteristics of interest have been chosen to recreate a sample reflective of the population as a whole, other unmeasured characteristics have not been accounted for because the underlying population distribution is unknown. In a random sample, the distribution of both measured and unmeasured characteristics withstand a better chance of reflecting the actual population. With AASM in MPS, however, the population distribution strata are preserved through sampling without replacement and quota restrictions to mitigate the oversampling of certain population groups. This method does require, however, access to reliable and recent statistical information regarding the parent-population's characteristics. These may be available from national surveys and other globally standardized surveys (eg, DHS), although recent information may not be readily available for some populations.

Given the large size of the population being sampled from, sampling without replacement from the entire population of phone owners can usually continue until the desired sample size in every sociodemographic stratum is achieved. In some cases, such as scenario B in Figure 2, stopping rules might be useful if certain age quota cannot be met, despite extended efforts, simply because those population groups cannot be reached directly using mobile phones. However, innovative methods similar to snowball sampling could be tried by requesting respondents to hand their phone to a family member, which is fulfilling specific, hard-to-obtain, requisite criteria.

XSL•FO
RenderX

Figure 2. Mobile phone access across two theoretical populations, using age as an illustrative respondent characteristic through which representativeness can be assessed. The figure illustrates a hypothetical population distribution against a distribution of mobile phone ownership, under conditions of low (scenario A) and high (scenario B) mobile penetration. In scenario A, common to populations where mobile phones have recently been introduced, obtaining a representative sample through MPS may not be feasible, even using AASM. As mobile markets mature, the overlap in distributions increases, allowing methods like AASM and random digit dialing to improve the capture of a sample that closely reflects the population-at-large.



Postsampling Weighting

Differences Between Phone Owners and Non-owners

It is well recognized that inequity exists in mobile phone access, and phone ownership is not equally distributed across the population in many countries. Young, male urban populations are more likely to have a mobile phone compared with older, female rural populations. Consequently, a major problem of mobile survey is the selectivity bias due to ownership heterogeneity. A common method is to conduct weighed analyses with poststratification adjustments to reduce the selectivity bias and to improve population representativeness.

Poststratification with weighting, however, is not without problems. Kish has shown that the variance σ^2/n of a weighted estimate is inflated by a factor of $1+CV^2_{wt}$, where CV^2_{wt} is the relative variance of the sampling weights [12]. The design-effect ($deff_{wt}$) of weighted estimate is formally written as shown in Figure 3.

where w_i is the weight (inverse of selection/participation probability) in the i-th stratum As larger is the variability of the weights, the larger is the design-effect, which reduces the efficiency of the sampling design (but increases variance with larger confidence intervals and reduces the ability to reject null hypothesis). A larger deff >1 reflects the loss of precision due to effective reduction in sample size.

A multicountry study in Afghanistan, Ethiopia, Mozambique, and Zimbabwe by the World Bank shows that the design-effects due to weighting were quite large and are 6.3, 11.6, 5.2 and 1.8, respectively [5].

Trimming extreme weights is often suggested to reduce the coefficient of variation (CV) of weight, which may bias the results. We propose, for MPS, reducing the variability in weighting by restricting quota of interviews for each strata to the original sample allocation size. An AASM approach is expected to substantially reduce the CV of weights and thus the *deff* impact of poststratification weighting.

Figure 3. The design-effect (deffwt) of weighted estimate. Here wi is the weight (inverse of selection/participation probability) in the i-th stratum.

$$deff_{wt} = \frac{n\sum_i w_i^2}{\left(\sum_i w_i\right)^2} = 1 + CV_{wt}^2$$

Minimizing the Effects of Nonresponse and Incomplete Surveys

Previous research indicates that dropout rates might be higher with MPS as people are more likely to be occupied than when contacted in person or on a landline and be less available to

XSL•FO
RenderX

complete the full survey [7]. Battery or connectivity/network failure during the survey is also a factor unique to mobile surveys that could contribute to dropouts [7]. However, the response rate could also potentially be higher for MPS than for landline surveys, as the available period to reach respondents is wider, instead of being limited to evenings and weekends when an individual is at home similar to the case with landline surveys.

To minimize the effect on the data from dropouts that do occur, the order of the survey modules could be randomized so that each module has the opportunity to be placed at the beginning of an interview to ensure a sufficient number of responses to each set of questions. Additionally, to keep mobile phone interviews short, questions asked should be limited to "important" key indicators identified in consultation with country policy makers and stakeholders, as well as the literature. Further, to control for the differences between those who respond to the survey and those who refuse, during the survey the number of those who choose not to respond should be recorded. The results should be adjusted postsurvey by nonresponse weighting, using a factor, R_i, expressed formally in Figure 4.

Figure 4. Nonresponse weighting, using a factor, Ri, where f is the nonresponse rate for the i-th group.

$$R_i = \frac{1}{1 - f_i}$$

Multiple Users per Phone

In low socioeconomic populations, mobile phones may be shared among members of a household or among neighbors, which may lead some individuals to be less likely to be surveyed [11,13]. To mitigate this risk, MPS implementations can randomize the time of day when calls are made, to reach individuals who typically use the phone at different times during the day, or purposefully target times of day when members of the family are more likely to be reached (eg, in the evening, when the phone owner returns from work outside the home). AASM, described above, will also aid in ensuring that a sufficient number of those with less frequent use of the phone, for example, women and older individuals, have increased chances of being included in the survey.

Multiple Phones per Person or Household

A concern commonly voiced when considering populations where cellular phone use is high is that people with more than 1 mobile phone have a statistically higher probability of selection from a population, with the relative probability of being selected proportional to the number of phones owned by that user. Although statistically correct, the practical implications of this relatively rare situation, from a population perspective, are negligible. Using RDD in a large population (eg, the denominator of every possible phone number in a country), the individual probability of selection of an individual is very small, even if the user has more than 1 phone.

The selection probability of an individual person can be calculated as follows: (n, sample size targeted × N[p], of phones per person)/(N[T], total number of phones). When the N(p) in numerator is 1, individuals with one phone have the equal selection probability (approximately) to simple random sampling. If N(p) = 2, the selection probability is doubled and so on thus increasing the selection probability proportional to the number of phones a person may have.

Table 2 below illustrates the individual probability of selection, as the proportion of the population with more than one mobile phone increases. As illustrated, in a population of 100 million phone owners, even if 10% of the population owns 3 mobile phones, the probability that one of these individuals is selected is still $2.5 \times e^{-8}$. This probability decreases as the proportion with more than 1 mobile phone increases. Although the theoretical probability of inclusion of those with multiple phones is three times that of someone with only one phone, in practice, the overall likelihood of contacting each participant remains infinitesimally low. Nonetheless, given that individuals with multiple phones are likely different from those with only one phone, negatively weighting responses from those individuals, or otherwise treating their contributed data as different from the majority of respondents may be worth considering. Over the course of our planned work, we aim to further explore this important issue to better elucidate the impact that multiple-phone ownership might have on the data collected in MPS.

Table 2. Illustration of the extremely low individual probability of inclusion of those with 1 phone and those with 3 phones in a theoretical population of 100 million and how these probabilities change as the proportion with multiple phones increases.

One mobile phone			Three mobile phones			Relative probability of inclusion (g) (f)/(c)
Percentage of 100 million population with one mobile phone (a)	Number of mobile phones in group (b) 100,000,000 × (a)	Individual probability of inclusion (c) 1/([b]+[e])	Percentage of 100 million population with 3 mobile phones (d)	Number of mobile phones in group (e) 100,000,000×(d)×3	Individual probability of inclusion (f) 3/([b]+[e])	
100%	100,000,000	1.00E-08	0%	0	3.00E-08	3
90%	90,000,000	8.33E-09	10%	30,000,000	2.5E-08	3
80%	80,000,000	7.14E-09	20%	60,000,000	2.14E-08	3
70%	70,000,000	6.25E-09	30%	90,000,000	1.88E-08	3
60%	60,000,000	5.56E-09	40%	120,000,000	1.67E-08	3
50%	50,000,000	5E-09	50%	150,000,000	1.5E-08	3

XSL•FO
RenderX

In countries where a significant proportion of the population has more than one mobile phone, the selection probability could be adjusted to take this into account. This has been done in landline surveys using an adjustment factor A, expressed formally as:

$A=1/T_i$

where T_i is the number of phones of the *i*-th person [14]. The inverse probability of selection multiplied by A yields the adjusted weight of each respondent in the survey [14].

Unresolved Statistical Challenges for MPS

Geographic Stratification and Representation

Determining the geographic location of respondents is a challenge unique to mobile surveys that is uncommon for landline or household surveys, where the area code or zip code of a respondent tends to be known. As such, if geographic balance is sought, screening questions may be necessary to associate the mobile phone respondent with a geographic location [7]. In population where mobility levels are high, questions should be developed a priori to satisfactorily assign geolocation (eg, where they spend the greatest number of days in a typical week or month) or to be determined by the participant themselves, directly stating a district or region of the country in which they reside.

Cost and Time Required to Obtain a Clean Sample (Removing Nonvalid Numbers)

With all RDD surveys, as phone numbers in the sample are randomly generated, it takes time to exclude nonworking numbers and eventually obtain the necessary sample size. This issue is especially pronounced for mobile surveys. In one example, it took almost 30 hours to remove 6872 nonworking mobile numbers compared with a landline sample, which took only 4.5 hours [7]. There are services that offer verification of phone number samples, but, as mentioned earlier, the associated costs may exceed the cost of the nonproductive call itself. Some active numbers could be also be flagged as nonworking in error [15]. Automated processes expedite the exclusion of nonworking numbers and replacing them, without requiring surveyor time. Notwithstanding, the time and cost necessary to isolate working numbers is still extremely low, when compared with the resources required to perform F2F household surveys. One study found that mobile interviews saved US $14 per participant compared to F2F interviews and took much less time because transportation to the field was not required [16].

Conclusions

The mobile phone revolution has presented an unprecedented opportunity to collect public health-related data directly from populations. Near-universal connectivity has created massive population denominators, which are accessible to researchers interested in supplementing traditional F2F methods with data from MPS. Leveraging large, connected populations for high-quality survey research requires careful consideration of both unique and shared challenges across traditional F2F, landline, mixed population, and mobile-only surveys.

The low cost and automated process of deploying MPS allows for innovative approaches such as AASM to be used to create samples that reflect the population-at-large, acknowledging that nonprobabilistic methods may be accompanied by unmeasurable biases. It is important that researchers working on using MPS methods consider these and, where possible, try to collect data which will improve not only the quality of the study but also our understanding of the strengths and limitations of this method.

We face a unique reality in a growing number of countries, where approaches like RDD now allow virtually every member of a population to be reached and surveyed about important public health issues. In most high-income countries, the over use of mobile networks by telemarketers has reduced our capacity to take advantage of these methods. Robust methods in sampling and design will help maximize the value of MPS data in these countries as a useful approach to population surveillance.

Acknowledgments

The studies described as part of the research agenda are funded by the Bloomberg Philanthropies. This funding agency had no role in the preparation of this manuscript.

Conflicts of Interest

None declared.

References

1. International Telecommunications Union. ITU. 2015 Dec. World Telecommunication/ICT Indicators Database Online URL: http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2016/ITU_Key_2005-2016_ICT_data.xls [accessed 2017-03-30] [WebCite Cache ID 6pMI5hRwE]
2. US Census Bureau. Census. 1990. Census of housing / Detailed housing characteristics URL: https://www.census.gov/hhes/www/housing/census/historic/phone.html [accessed 2017-03-30] [WebCite Cache ID 6pMHd7Ack]
3. Blumberg S, Luke J, Division of Health Interview Statistics, National Center for Health Statistics. CDC. 2010. Wireless substitution: early release of estimates from the National Health Interview Survey, January-June 2010 URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201012.pdf [accessed 2017-04-10] [WebCite Cache ID 6pc3J0Y4o]

XSL•FO
RenderX

4. GeoPoll. Geopoll. Social Sector Work URL: https://research.geopoll.com/pages/social-sector.html [accessed 2017-03-30] [WebCite Cache ID 6pMIDzstX]
5. Leo B, Morello R, Mellon J, Peixoto T, Davenport S. Cgdev. 2015. Do Mobile Phone Surveys Work in Poor Countries? Center for Global Development Working Paper URL: http://cgdev.org.488elwb02.blackmesh.com/sites/default/files/CGD-Working-Paper-398-Mobile-Phones.pdf [accessed 2017-03-30] [WebCite Cache ID 6mf03yUeu]
6. Kempf AM, Remington PL. New challenges for telephone survey research in the twenty-first century. Annu Rev Public Health 2007 Apr;28:113-126. [doi: 10.1146/annurev.publhealth.28.021406.144059] [Medline: 17094769]
7. Vicente P, Reis E, Santos M. Using mobile phones for survey research: a comparison with fixed phones. Int J Market Res 2009;51(5):613-633. [doi: 10.2501/S1470785309200852]
8. Moura EC, Claro RM, Bernal R, Ribeiro J, Malta DC, Morais Neto O. A feasibility study of cell phone and landline phone interviews for monitoring of risk and protection factors for chronic diseases in Brazil. Cad Saúde Pública 2011 Feb;27(2):277-286. [doi: 10.1590/S0102-311X2011000200009]
9. Lepkowski J. Telephone sampling methods in the United States. In: Telephone Survey Methodology. New York: Wiley; 1988.
10. AAPOR. Deerfield, IL: AAPOR Cell Phone Task Force; 2010. New considerations for survey researchers when planningconducting RDD telephone surveys in the U.S. with respondents reached via cell phone numbers URL: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/2010AAPORCellPhoneTFReport.pdf [accessed 2016-11-14] [WebCite Cache ID 6m1LYoSDR]
11. Tran MC, Labrique AB, Mehra S, Ali H, Shaikh S, Mitra M, et al. Analyzing the mobile "digital divide": changing determinants of household phone ownership over time in rural bangladesh. JMIR Mhealth Uhealth 2015 Feb 25;3(1):e24 [FREE Full text] [doi: 10.2196/mhealth.3663] [Medline: 25720457]
12. Kish L. Weighting for Unequal Pi. J Off Stat 1992;8(2):183-200.
13. Poorman E, Gazmararian J, Elon L, Parker R. Is health literacy related to health behaviors and cell phone usage patterns among the text4baby target population? Arch Public Health 2014 May;72(1):13 [FREE Full text] [doi: 10.1186/2049-3258-72-13] [Medline: 24872883]
14. Groves RM. Weighting adjustments for random digit dialed surveys. In: Telephone Survey Methodology. New York: Wiley; 1988.
15. Pew Research Center. 2015. Advances in Telephone Survey Sampling: Balancing efficiency and coverage using several new approaches URL: http://www.pewresearch.org/files/2015/11/2015-11-18_telephone-sampling [accessed 2016-11-14] [WebCite Cache ID 6m1LctRQL]
16. Mahfoud Z, Ghandour L, Ghandour B, Mokdad AH, Sibai AM. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? Field methods 2014 Jul 01;27(1):39-54. [doi: 10.1177/1525822X14540084]

Abbreviations

AASM: automated active strata monitoring
CV: Coefficient of variation
F2F: face-to-face
LMICs: low- and middle-income countries
MPS: mobile phone surveys
RDD: rapid digital dialing
MNO: mobile network operator
DHS: demographic and health surveys

Edited by E Rosskam; submitted 15.01.17; peer-reviewed by Q Li, L Moulton; comments to author 02.02.17; revised version received 27.02.17; accepted 08.03.17; published 05.05.17.

Please cite as:
Labrique A, Blynn E, Ahmed S, Gibson D, Pariyo G, Hyder AA
Health Surveys Using Mobile Phones in Developing Countries: Automated Active Strata Monitoring and Other Statistical Considerations for Improving Precision and Reducing Biases
J Med Internet Res 2017;19(5):e121
URL: http://www.jmir.org/2017/5/e121/
doi:10.2196/jmir.7329
PMID:28476726

XSL•FO
RenderX

©Alain Labrique, Emily Blynn, Saifuddin Ahmed, Dustin Gibson, George Pariyo, Adnan A Hyder. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Development of an Interactive Voice Response Survey for Noncommunicable Disease Risk Factor Estimation: Technical Assessment and Cognitive Testing

Dustin G Gibson[1*], MS, PhD; Brooke A Farrenkopf[1*], BSPH; Amanda Pereira[1], MSPH; Alain B Labrique[1], PhD; George William Pariyo[1], MBChB, PhD

Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD, United States

[*]these authors contributed equally

Corresponding Author:
Dustin G Gibson, MS, PhD
Johns Hopkins Bloomberg School of Public Health
Department of International Health
615 N Wolfe St
Baltimore, MD, 21231
United States
Phone: 1 4432878763
Fax: 1 4106141419
Email: dgibso28@jhu.edu

Abstract

Background: The rise in mobile phone ownership in low- and middle-income countries (LMICs) presents an opportunity to transform existing data collection and surveillance methods. Administering surveys via interactive voice response (IVR) technology—a mobile phone survey (MPS) method—has potential to expand the current surveillance coverage and data collection, but formative work to contextualize the survey for LMIC deployment is needed.

Objective: The primary objectives of this study were to (1) cognitively test and identify challenging questions in a noncommunicable disease (NCD) risk factor questionnaire administered via an IVR platform and (2) assess the usability of the IVR platform.

Methods: We conducted two rounds of pilot testing the IVR survey in Baltimore, MD. Participants were included in the study if they identified as being from an LMIC. The first round included individual interviews to cognitively test the participant's understanding of the questions. In the second round, participants unique from those in round 1 were placed in focus groups and were asked to comment on the usability of the IVR platform.

Results: A total of 12 participants from LMICs were cognitively tested in round 1 to assess their understanding and comprehension of questions in an IVR-administered survey. Overall, the participants found that the majority of the questions were easy to understand and did not have difficulty recording most answers. The most frequent recommendation was to use country-specific examples and units of measurement. In round 2, a separate set of 12 participants assessed the usability of the IVR platform. Overall, participants felt that the length of the survey was appropriate (average: 18 min and 31 s), but the majority reported fatigue in answering questions that had a similar question structure. Almost all participants commented that they thought an IVR survey would lead to more honest, accurate responses than face-to-face questionnaires, especially for sensitive topics.

Conclusions: Overall, the participants indicated a clear comprehension of the IVR-administered questionnaire and that the IVR platform was user-friendly. Formative research and cognitive testing of the questionnaire is needed for further adaptation before deploying in an LMIC.

(J Med Internet Res 2017;19(5):e112) doi:10.2196/jmir.7340

KEYWORDS

interactive voice response; noncommunicable disease; survey methodology; public health surveillance; cellular phone; risk factors

XSL•FO
RenderX

Introduction

The increasing rise in mobile phone ownership and access in low- and middle-income countries (LMICs)—from 1 billion in 2000 to 6 billion in 2012—introduces the opportunity to transform the current paradigm of surveillance activities and to potentially improve the efficiency and timeliness of data collection and reporting [1]. One such opportunity, mobile phone surveys (MPS), offers several potential advantages over traditional household-based surveys. These advantages include real-time data entry to enable timely data analysis, survey delivery that is less demanding on financial and human resources, and anonymity of responses [2]. With the growing noncommunicable disease (NCD) burden [3], there is a subsequent greater need to collect and utilize data to guide public health programs and curb the global NCD epidemic [4].

Interactive voice response (IVR) technology is one of the several options for conducting an MPS. IVR utilizes a prerecorded questionnaire that is administered over the phone [5]. Participants select responses via touchtone keypad or voice recognition software. Responses are immediately submitted to either Web-based databases or internal servers to enable timely data synthesis and analysis [6]. IVR counters a key challenge of short message service (SMS) surveys: the requirement of literate populations.

In adapting a household-administered questionnaire to an IVR survey, cognitive testing of the IVR questionnaire and assessing its usability becomes imperative given IVR's limitations; especially due to its self-administered nature where respondents are not afforded an opportunity to ask any clarifying questions. In survey development, cognitive testing is frequently applied in order to identify questions that respondents have difficulty comprehending and to assure that the questions adequately capture information as intended [7,8]. Results from cognitive testing can guide question wording and formatting, leading to greater understandability and accuracy in survey responses [9].

The two objectives of this pilot study were to (1) cognitively test and identify challenging questions in an NCD risk factor questionnaire administered via IVR and (2) assess the usability of the IVR platform and identify future challenges for its implementation in LMICs. The findings from this research will be used to revise the questionnaire and IVR platform before conducting a similar series of formative activities in each LMIC where the IVR survey will be deployed.

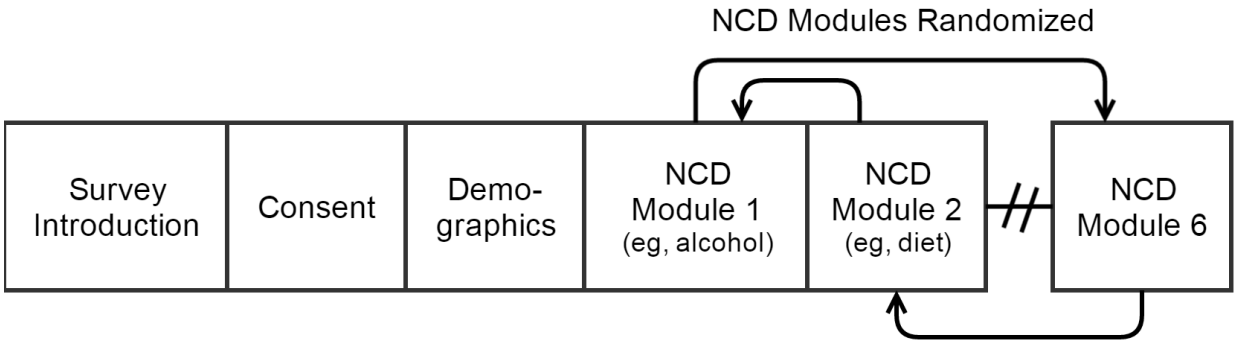
Methods

Questionnaire Development

As part of the Bloomberg Philanthropies Data for Health Initiative (BD4HI) [10], experts in survey methodology, NCDs, and mobile health convened in June 2015 to develop an NCD risk factor questionnaire that could be adapted to an MPS and used to collect population-representative estimates from LMICs [11]. Questions were selected from standardized household surveys such as WHO STEPwise Surveillance, Tobacco Questions for Surveys, and the Behavioral Risk Factor Surveillance System [12-14]. Questions that mapped to indicators in the Global Monitoring Framework for NCDs and that covered the 4 main risk factors for NCDs (physical activity, alcohol consumption, tobacco use, and diet) were preferred [15]. Questions were selected for an IVR survey independent of their perceived suitability to the IVR modality. This produced a *beta* version of the questionnaire that was adapted to VOTO mobile's IVR platform—a Ghana-based organization that works to develop MPS systems [16]—and used to cognitive test the IVR survey and to assess its usability (see Multimedia Appendices 1 and 2).

The IVR survey included a brief introduction, and was followed by a question asking for assent to participate, demographic questions, and NCD modules (Figure 1). Modules were a series of topically similar questions such as alcohol consumption, tobacco use, and dietary intake (Table 1). The IVR platform was programmed to randomize the delivery order of the NCD modules. For each question, the IVR survey was programmed so that respondents could repeat the question by pressing the asterisk, "star key," on the mobile phone.

Figure 1. Interactive voice response (IVR) survey design. NCD: noncommunicable disease.



XSL•FO
RenderX

Table 1. Number and source of questions included in each module by round of pilot testing.

Module	Round 1 n questions	Round 2 n questions	Source survey
Introduction or consent	1	1	-
Demographics	5	6	-
Tobacco	2	2	TQS
Alcohol	3	3	STEPS
Diet	10	10	STEPS
Diabetes and blood pressure medication	4	4	STEPS
Physical activity	12	-	IPAQ
Physical activity	-	20	GPAQ
Lifestyle	6	-	STEPS
Total N questions	43	46	

Participants

Participants were eligible for inclusion in the pilot study if they were proficient in English and identified as being a native of any country within Africa, Asia, Latin America, or South America. Participants were excluded if they were aged under 21 years or had a hearing impairment. The study received ethical approval from the JHSPH Institutional Review Board (IRB). Study participants provided oral consent and were compensated for their participation with a US $10 gift-card to a local coffee shop.

Data Collection and Analysis

We conducted two rounds of pilot testing, with different participants in each round, to gain clearer understanding of the participants' thought processes (round 1) and to assess the usability of the IVR survey (round 2). There were minor differences in the questions and modules between the two rounds (Table 1). Participants listened to the IVR survey on a Samsung Convoy 3 mobile phone provided by the study team. This simple phone was selected to serve as a proxy for the types of phones commonly used in LMICs. Participants in each round of pilot testing were instructed to answer the IVR survey honestly and to think about how this survey could be improved if administered in their country of origin.

Round 1: Cognitive Testing

In order to minimize recall bias, participants were administered the modules one at a time. After each module, the IVR survey was paused and participants were cognitively tested through active probing to assess their understanding of each question and to identify specific areas of misunderstanding. Representative examples of questions asked of participants in Round 1 are found in Textbox 1.

NCD questions and modules were scored for their comprehensibility, which was considered "high" if >75% of the participants did not express concern over the introduction or question content overall throughout the module; "medium" if 51-75% of participants expressed no concern; and "low" if <50% of participants found no difficulty.

Round 2: Assessment of IVR Platform Usability

During round 2, groups of 2-3 participants listened to the IVR survey in its entirety and were then asked about the IVR usability and any overall concerns with the survey (see Textbox 2). Participants were encouraged to vocalize any comments about the survey's wording, length, understandability, and ways to improve survey performance if administered nationally in an LMIC setting**.** Comments on the survey length, the survey's introduction, and features they liked or did not like about the IVR platform were compiled. Following the focus groups, participants were individually administered the GPAQ physical activity questions through IVR and were cognitively tested to assess their understanding using methods similar to those employed in round 1. No other NCD modules were cognitively tested during round 2.

Textbox 1. Examples of questions asked during pilot testing in round 1.

Round 1: Cognitive testing

- Was this question clear?
- Were there any words or phrases that were difficult to understand?
- What does the word, "XXX" mean to you?
- How confident were you in your answer?
- Do you foresee any challenges asking this question in an LMIC?

XSL•FO
RenderX

Textbox 2. Examples of questions asked during pilot testing in round 2.

Round 2: Usability of interactive voice response (IVR)

- What factors would cause you to be more or less likely to participate in a similar mobile phone administered health or NCD surveys?
- Comment on your reaction and experiences during the IVR survey or how you would react if you got such a survey in the future?
- How do you expect people in your country or community would react if they received such a survey on their mobile phones?
- The modules were randomized—do you think there are any issues to consider in randomizing questions in your home country?

Results

From January to February 2016 in Baltimore, MD, a total of 24 participants were pilot tested, with 12 participants in each of the 2 rounds.

Cognitive Testing: Round 1

In Round 1 (n=12), the median participant's age and years of education completed were 27 years old (IQR: 25-30 years; range: 22-56 years) and 15 years (IQR: 13-18 years; range: 9-20 years), respectively (Table 2). Cognitive testing through individual interviews with the participants identified that the examples that were used in the question to help respondents in answering (eg, "I would now like to ask you about smoking tobacco, including cigarettes, cigars, and pipes.") as a common area of concern across the NCD modules. All 12 participants (100%) commented that the provided examples for diet, physical activity, tobacco, and alcohol should be specific to the country where the survey is being conducted.

Table 2. Demographic characteristics of participants by round.

Demographic characteristics	Round 1 n=12	Round 2 n=12	Total n=24
Sex, n (%)			
Male	4 (33)	4 (33)	8 (33)
Female	8 (67)	8 (67)	16 (67)
Region of birth, n (%)			
Africa	2 (17)	8 (67)	10 (42)
South Asia	5 (42)	1 (8)	6 (25)
Central and East Asia	2 (17)	1 (8)	3 (13)
Latin America	2 (17)	1 (8)	3 (13)
South America	1 (8)	1 (8)	2 (8)
Age in years, median (IQR)	27 (25-30)	28.5 (26-33)	27 (25-31)
Education in years, median (IQR)	15 (13-18)	18 (14-19)	16 (13-18)

In addition to this survey-wide comment, several questions were problematic for participants. These questions and their respective challenges are listed below and summarized in Table 3.

Urban or Rural Setting

> *(In your home country) do you live in a rural or urban area? If you live in a rural area, press 1. If you live in an urban area, press 3.*

This proved to be one of the more challenging questions, with 11 (91.7%) participants raising concern that other people taking this survey may have difficulty distinguishing between and defining urban and rural. Two participants (16.7%) suggested including "peri-urban" as an available response.

Education

> *Not including preschool, how many years of school and full time study have you completed? Please enter the number of years.*

All participants had issues with this question, with many reporting that they felt rushed to calculate a response. Four participants (33.3%) recommended removing the word "preschool" and three participants (25%) recommended converting the responses to be categorical (eg, primary school, secondary school, and so on) such that it matches the country's education system.

Alcohol

> *One drink is equivalent to a 12 ounce beer, a five ounce glass of wine, or a drink with one shot of liquor...*

Three participants (25%) did not understand at least one of the terms of measurement used and stressed that in-country research would be necessary to provide the accurate measurement and country-specific examples.

Salt Intake

> *...I would like you to think about all the sources of salt, including ordinary table salt, unrefined salt such*

XSL•FO
RenderX

as sea salt, iodized salt, salty stock cubes and powders, and salty sauces such as soya sauce or fish sauce.

Approximately 92% (11/12) of participants found this confusing, stating that they were unfamiliar with some of the examples used (eg, participants said soya sauce was not used in their country). One-third of participants (n=4) said there were too many examples listed and recommended against using similar sounding examples or sentence structure.

Vegetable Consumption

A serving of vegetables is about a cup of green leafy vegetables or salad or half a cup of cooked or chopped vegetables. How many of these servings of vegetables do you eat on one of those days?

Three participants (25%) expressed confusion with this question. Participants commented that the use of "leafy greens" directed them to include only green vegetables and exclude other vegetables, like carrots.

Physical Activity

How much time do you spend doing vigorous-intensity activities at work on a typical day? I will ask you to enter hours followed by minutes. Please enter between 16 & 0 hours now?

After participants reported hours, the question was followed by

Now enter between 59 & 0 minutes.

Half of the participants (n=6) thought that the way time was measured was confusing; by first asking about hours and then minutes. Participants suggested that the question be simplified by only asking about hours.

Table 3. Summary of respondents' assessment of comprehensibility of interactive voice response (IVR) questionnaire.

Module	Level of comprehensibility[a] (high, medium, low)	Remarks
Demographics	Medium	Participants had low comprehensibility with the questions regarding education and rural or urban settings, but high comprehensibility with questions on age and sex.
Tobacco	High	No challenges other than providing country-specific examples of tobacco.
Alcohol	Medium	Participants had difficulty with units given to measure their alcohol consumption.
Diet	Medium	Participants had low comprehensibility with the salt questions, but high and medium comprehension with the fruit and vegetable questions, respectively.
Blood pressure and diabetes	High	No challenges identified.
Physical activity (IPAQ)	Medium	Participants had difficulty differentiating between levels of activity (moderate vs vigorous). Question structure was repetitive leading to reporting fatigue.
Physical activity (GPAQ)	Low	Participants had difficulty differentiating between levels of activity and with the question structure to estimate their time spent doing physical activity. Question structure was repetitive leading to reporting fatigue.
Lifestyle	High	No challenges identified.

[a]Comprehensibility was considered "high" if >75% of the participants in round 1 of testing did not express concern over the introduction or question content overall throughout the module; "medium" if 51-75% of participants expressed no concern; and "low" if <50% of participants found no difficulty.

Usability of the IVR Platform: Round 2

Round 2 participants were similar in demographics to those of Round 1; the median participant's age and years of education completed were 28.5 years old (IQR: 26-33 years; range: 22-35 years) and 18 years (IQR: 14-19 years; range: 4-20 years), respectively (Table 2). Due to skip patterns programmed into the 46-question survey, the average number of questions answered was 36. Participants spent an average of 18 min and 31 s answering the survey (data not shown).

Survey Length

Overall, the majority of participants thought that the length of the survey was appropriate, with many participants commenting on the thoroughness of the survey. When asked how long they would be willing to spend completing a survey via mobile phone, participants estimated a range of 10-20 min. Participants were more critical of the length of specific questions and modules than the overall survey. During certain modules, response fatigue was related to confusion over unfamiliar examples, repetitive question structure, and difficulty in understanding the narrator.

Survey Introduction

The survey began with an introduction that contained keypress instructions, the expected survey duration, and that the survey was "sponsored by the Ministry of Health." The majority of participants thought that the introduction and instructions were sufficient to complete the IVR survey. Some participants commented that indicating that the survey was sponsored by the government may affect response rate either positively or negatively depending on the country context. For instance, participants thought that respondents might be more hesitant to partake in a government-sponsored survey during an election period.

Survey Features

Participants appreciated the consistent use of key press options for questions that had two answers, such as the Yes or No questions (eg, Press1 for Yes, Press 3 for No). Similarly,

XSL•FO
RenderX

participants appreciated being allowed to repeat the question by pressing the asterisk (star) key.

Several participants also commented that the time between providing a response and the narration of the subsequent question was too short. Some participants did not hear the first few words of the next question as they were bringing the phone back to their ear after entering their response.

Nearly all participants experienced confusion with understanding the accent of the narrator selected to record the survey (Ghanaian). Many participants commented that the narrator was monotone and spoke too quickly. Participants recommended that the narrator use inflection to highlight key points of the question.

Question Structure

Participants found lengthy questions and modules with similar sentence structure to be repetitive, causing them to lose focus. This was a key issue in the physical activity module, where three similar sets of questions asked about three different levels of physical activity: (1) vigorous physical activity, (2) moderate physical activity, and (3) walking. After answering the vigorous physical activity questions, some participants did not realize that they were being asked new questions about moderate physical activity. Nearly all participants had issues distinguishing between the levels of activity (eg, vigorous vs moderate), and some admitted to realizing that they unintentionally double-counted their activity.

Perceived Barriers and Solutions to IVR

When asked about potential barriers to deploying IVR surveys in their birth country, participants mentioned that respondents might be unable to move to a quiet location at the time of the call and may refuse to answer an incoming call with an unfamiliar phone number. Participants also thought that future IVR surveys would benefit from including an incentive as some people might not complete the survey if they were randomly dialed. However, they also stated that requiring a user to input personal information (such as bank account information) via phone would be problematic even with an incentive and may lead to not participating in the survey at all. When told about an airtime incentive that would not request any additional or personal information, all participants agreed that this was a preferable option.

Nearly all participants preferred a survey conducted over the phone rather than in person. Participants appreciated the anonymity of the IVR survey and liked that it felt less personal than a face-to-face interview, establishing a foundation to respond honestly. Participants expressed concern over topics that may be perceived as sensitive or controversial in their home country, including alcohol intake and medical diagnoses. Some participants expressed that respondents might be encouraged to select the answer that supports the more socially accepted, "healthy" option.

Discussion

Principal Findings

The IVR platform presents an opportunity to transform current modes of data collection for large-scale household surveys and to potentially improve the quality and utilization of the data [17]. Findings from the two rounds of pilot testing provide guidance to tailor the IVR platform for increased usability among future respondents and to identify challenging questions that need further refinement to ensure that the intended data are being collected correctly.

The common request to have greater in-country influence in question development and structure suggests the need for similar in-country cognitive testing. Conducting key informant interviews and focus group discussions would enable country-specific survey adaptation before national administration. Formative research should focus on selection of local terms to use as examples (eg, types of physical activity, terms of measurement). For some surveys, using different sets of modules with subnational, region-specific terms could enhance understandability and accuracy of responses. Local country-specific adaptation with relevant country examples will be important during future developments of IVR surveys, even if the generic survey is also meant for cross-country comparisons of key indicators. For instance, in low literacy settings, the concept of time down to minutes may be difficult to elicit and one may use half an hour or quarter of an hour blocks of time to describe duration of activities being investigated [18].

Formative research should further focus on selecting narrators that are native to countries and regions, where dialects may vary throughout the country. Guidance to have the narrator speak slowly and intonate specific words, strategically selected by the questionnaire development team to emphasize key details, should be integral during the audio recording phase [19]. Greater intonation and question structure variety can also make an IVR survey more personal, while still attaining the benefits of anonymity [20]. Research to determine topics perceived as sensitive or personal would aid in determining questions where participants place greater importance on confidentiality. A reminder of confidentiality before modules with sensitive or personal content could aid in participant retention and truthfulness (eg, "As a reminder, this survey is confidential") [21]. Consideration should also be given to the age and gender of the narrator. In Ghana, researchers found that IVR surveys narrated by a female led to a higher response rate [22].

One of the most promising findings from the testing was the overall acceptance of the survey length (average: 18 min and 31 s) and the appreciation of instruction thoroughness. This suggests that surveys of similar length may be a feasible option to complement existing surveillance methods, but they require further empirical testing. Participants' appreciation of the thoroughness of the question prompts and module introductions, even at the expense of increasing the length of the survey, show that it may not always be necessary to compromise detail to minimize overall survey length as long as overall time duration is kept reasonable. Participants did not like the feeling of being rushed to answer the question, which supports lengthening the

XSL•FO
RenderX

time provided to select a response. In future formative work, it will be important to test out various durations of pauses in-between questions so that participants will not *miss out* on hearing part of the question.

When discussing potential barriers to IVR surveys, considerations on the participant's readiness to complete the survey should be made. Some participants were concerned that future respondents would not be available to complete the survey upon receiving the call. This presents the opportunity to inform the respondents in advance, via SMS message for example, that they will receive a call from an unknown phone number to administer the survey. Introducing a technology to select a window of time that is best to receive the call could also improve response rate.

With some participants raising concern over the public opinion of government sources and willingness to partake in its surveys, formative research should evaluate the perception and possibilities of text to include in the survey introduction section early on at connection. The introduction should convey the message that this is an officially approved or sanctioned survey being implemented by a nationally recognized public health agency or research institution. The mention of such a neutral agency or public institution by name and that the survey would serve a public health common good through information for better planning could improve participant willingness to continue with the survey. Response rate could also be improved by involving a well-known, popular, and respected figure, such as an athlete, to narrate the survey. To further improve response rate, future research should examine the effectiveness of overall and demographic-specific incentives to improve generalizability of the survey's study population.

Limitations

This study has several limitations. First, our sample was primarily composed of students and fellows at a single university, who have knowledge on public health topics and surveillance methods. Therefore, the participants in this study may have had less difficulty in understanding the questionnaire as compared with other members in an LMIC community. Similarly, many participants were not asked follow-up questions about risk factors or health conditions, therefore possibly resulting in a shorter survey than what might be observed if sent to a sample representative of the community. Second, our study sample was relatively young in age and therefore may be more adept at using a mobile phone. The questionnaire will need to be cognitively tested within each country and among a wide range of demographic groups before its deployment. Third, we did not collect information on how long the participant lived in their native country. This has potential implications on our findings as we asked participants to frame several of their responses through the lens of an LMIC survey participant.

Conclusions

Overall, participants did not have difficulty with understanding the questions and recording their responses. Most participants appreciated the anonymity of the IVR survey, stating that in comparison with face-to-face interviews, it encourages honest and accurate responses. Participants also felt that the length of the survey was appropriate and expressed a preference to have instructions explained thoroughly.

Incorporating the recommendations common among participants and conducting formative research will help develop an NCD survey that can be administered via IVR, particularly around the selection of country-specific examples and narrators to improve understandability. This shows that IVR may be an appropriate vehicle to administer timely, resource-efficient risk factor surveillance among populations in LMIC settings.

Acknowledgments

This study was funded by Bloomberg Philanthropies. This funding agency had no role in the preparation of this manuscript. We would like to thank colleagues from the Centers for Disease Control and Prevention and the World Health Organization for their contributions in the selection of questions to include in the survey.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire used for cognitive testing (round 1).

[PDF File (Adobe PDF File), 142KB - jmir_v19i5e112_app1.pdf]

Multimedia Appendix 2

Questionnaire used for usability assessment (round 2).

[PDF File (Adobe PDF File), 164KB - jmir_v19i5e112_app2.pdf]

References

XSL•FO
RenderX

1. Information and Communications for Development: Maximizing Mobile. Washington, DC: World Bank Publications; 2012.
2. World Health Organization. mHealth: New Horizons for Health through Mobile Technologies: Second Global Survey on eHealth (Global Observatory for Ehealth). Geneva, Switzerland: World Health Organization; 2011.
3. World Health Organization. Global action plan for the preventioncontrol of noncommunicable diseases 2013-2020. Geneva, Switzerland: World Health Organization; 2013.
4. World Health Organization. WHO. 2017. Monitoring and surveillance of noncommunicable diseases URL: http://www.who.int/ncds/surveillance/en/ [accessed 2017-03-22] [WebCite Cache ID 6p9vzXT7A]
5. Lam MY, Lee H, Bright R, Korzenik JR, Sands BE. Validation of interactive voice response system administration of the Short Inflammatory Bowel Disease Questionnaire. Inflamm Bowel Dis 2009 Apr;15(4):599-607. [doi: 10.1002/ibd.20803] [Medline: 19023897]
6. 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]
7. Finger JD, Tafforeau J, Gisle L, Oja L, Ziese T, Thelen J, et al. Development of the European Health Interview Survey - Physical Activity Questionnaire (EHIS-PAQ) to monitor physical activity in the European Union. Arch Public Health 2015;73:59 [FREE Full text] [doi: 10.1186/s13690-015-0110-z] [Medline: 26634120]
8. Levine RE, Fowler FJ, Brown JA. Role of cognitive testing in the development of the CAHPS Hospital Survey. Health Serv Res 2005 Dec;40(6 Pt 2):2037-2056 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00472.x] [Medline: 16316437]
9. Alaimo K, Olson CM, Frongillo EA. Importance of cognitive testing for survey items: an example from food security questionnaires. J Nutr Educ 1999 Sep;31(5):269-275. [doi: 10.1016/S0022-3182(99)70463-2]
10. Bloomberg. 2016. Data for Health URL: https://www.bloomberg.org/program/public-health/data-health/ [accessed 2017-03-22] [WebCite Cache ID 6p9vOvHqz]
11. Hyder A, Wosu A, Gibson D, Labrique A, Ali J, Pariyo G. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246]
12. WHO. STEPS: A framework for surveillance: WHO STEPwise approach to NCD surveillance. Geneva, Switzerland: World Health Organization; 2003.
13. Global Adult Tobacco Survey Collaborative Group. Tobacco Questions for Surveys: A Subset of Key Questions from the Global Adult Tobacco Survey (GATS), 2nd edition. Atlanta, GA: Centers for Disease Control and Prevention; 2011.
14. Centers for Disease Control and Prevention. CDC. 2016. Behavioral Risk Factor Surveillance System URL: https://www.cdc.gov/brfss/ [accessed 2017-03-11] [WebCite Cache ID 6osspBlgF]
15. WHO. 2017. NCD Global Monitoring Framework: ensuring progress on noncommunicable diseases in countries URL: http://www.who.int/nmh/global_monitoring_framework/en/ [accessed 2017-03-22] [WebCite Cache ID 6p9yHOzVl]
16. votomobile. 2017. VOTO URL: https://www.votomobile.org/ [accessed 2017-03-22] [WebCite Cache ID 6p9yUW8qu]
17. Dillon B. Using mobile phones to collect panel data in developing countries. J Int Dev 2011 Feb 09;24(4):518-527. [doi: 10.1002/jid.1771]
18. Masuda YJ, Fortmann L, Gugerty MK, Smith-Nilson M, Cook J. Pictorial approaches for measuring time use in rural Ethiopia. Soc Indic Res 2014 Jan;115(1):467-482 [FREE Full text] [doi: 10.1007/s11205-012-9995-x] [Medline: 25620832]
19. Conrad FG, Schober MF. Clarifying question meaning in a household telephone survey. Public Opin Q 2000;64(1):1-28. [Medline: 10810073]
20. Tourangeau R, Couper MP, Steiger DM. Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. Comput Human Behav 2003 Jan;19(1):1-24. [doi: 10.1016/S0747-5632(02)00032-8]
21. Singer E, Von TD, Miller E. Confidentiality assurances and response: a quantitative review of the experimental literature. Public Opin Q 1995;59(1):66-77. [doi: 10.1086/269458]
22. Pegus C. Crowdsourcing citizen feedback on district development in Ghana using interactive voice response surveys. Making All Voices Count Practice Paper. Brighton: Institute of Development Studies; 2016. URL: https://opendocs.ids.ac.uk/opendocs/handle/123456789/12716 [WebCite Cache ID 6p9yrvDp0]

Abbreviations

BRFSS: Behavioral Risk Factor Surveillance System
CDC: Centers for Disease Control and Prevention
GPAQ: Global Physical Activity Questionnaire
IPAQ: International Physical Activity Questionnaire
IQR: Interquartile range
IVR: interactive voice response
JHSPH: Johns Hopkins Bloomberg School of Public Health
LMIC: low- and middle-income country
MPS: mobile phone survey
NCD: noncommunicable disease

XSL•FO
RenderX

SMS: short message service
WHO: World Health Organization
TQS: Tobacco Questions for Surveys
STEPS: The WHO STEPwise approach to Surveillance

Edited by E Rosskam; submitted 17.01.17; peer-reviewed by N Kaonga, M Motevalli-Oliner; comments to author 02.02.17; revised version received 17.02.17; accepted 19.02.17; published 05.05.17.

Please cite as:
Gibson DG, Farrenkopf BA, Pereira A, Labrique AB, Pariyo GW
The Development of an Interactive Voice Response Survey for Noncommunicable Disease Risk Factor Estimation: Technical Assessment and Cognitive Testing
J Med Internet Res 2017;19(5):e112
URL: http://www.jmir.org/2017/5/e112/
doi:10.2196/jmir.7340
PMID:28476724

©Dustin G Gibson, Brooke A Farrenkopf, Amanda Pereira, Alain B Labrique, George William Pariyo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review

Dustin G Gibson[1], MS, PhD; Amanda Pereira[1], MSPH; Brooke A Farrenkopf[1], MSPH; Alain B Labrique[1], MHS, MS, PhD; George W Pariyo[1], MBChB, PhD; Adnan A Hyder[1,2], MPH, MD, PhD

[1]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Dustin G Gibson, MS, PhD
Department of International Health
Johns Hopkins Bloomberg School of Public Health
615 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 443 287 8763
Fax: 1 410 614 1419
Email: dgibso28@jhu.edu

Abstract

Background: National and subnational level surveys are important for monitoring disease burden, prioritizing resource allocation, and evaluating public health policies. As mobile phone access and ownership become more common globally, mobile phone surveys (MPSs) offer an opportunity to supplement traditional public health household surveys.

Objective: The objective of this study was to systematically review the current landscape of MPSs to collect population-level estimates in low- and middle-income countries (LMICs).

Methods: Primary and gray literature from 7 online databases were systematically searched for studies that deployed MPSs to collect population-level estimates. Titles and abstracts were screened on primary inclusion and exclusion criteria by two research assistants. Articles that met primary screening requirements were read in full and screened for secondary eligibility criteria. Articles included in review were grouped into the following three categories by their survey modality: (1) interactive voice response (IVR), (2) short message service (SMS), and (3) human operator or computer-assisted telephone interviews (CATI). Data were abstracted by two research assistants. The conduct and reporting of the review conformed to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

Results: A total of 6625 articles were identified through the literature review. Overall, 11 articles were identified that contained 19 MPS (CATI, IVR, or SMS) surveys to collect population-level estimates across a range of topics. MPSs were used in Latin America (n=8), the Middle East (n=1), South Asia (n=2), and sub-Saharan Africa (n=8). Nine articles presented results for 10 CATI surveys (10/19, 53%). Two articles discussed the findings of 6 IVR surveys (6/19, 32%). Three SMS surveys were identified from 2 articles (3/19, 16%). Approximately 63% (12/19) of MPS were delivered to mobile phone numbers collected from previously administered household surveys. The majority of MPS (11/19, 58%) were panel surveys where a cohort of participants, who often were provided a mobile phone upon a face-to-face enrollment, were surveyed multiple times.

Conclusions: Very few reports of population-level MPS were identified. Of the MPS that were identified, the majority of surveys were conducted using CATI. Due to the limited number of identified IVR and SMS surveys, the relative advantages and disadvantages among the three survey modalities cannot be adequately assessed. The majority of MPS were sent to mobile phone numbers that were collected from a previously administered household survey. There is limited evidence on whether a random digit dialing (RDD) approach or a simple random sample of mobile network provided list of numbers can produce a population representative survey.

(J Med Internet Res 2017;19(5):e139) doi:10.2196/jmir.7428

XSL•FO
RenderX

KEYWORDS

survey methodology; cellular phone; interactive voice response; short messages service; computer-assisted telephone interview; mobile phone surveys

Introduction

National and subnational surveys are important for monitoring disease burden, prioritizing resource allocation, and evaluating public health policies [1]. In low- and middle-income countries (LMICs), such surveys typically rely on face-to-face interviews conducted at the respondent's household. Household surveys are conducted infrequently, typically due to high costs in personnel and transportation associated with household survey implementation and the face-to-face nature of data collection [2-5]. In addition, household surveys require considerable amounts of time for data collection, data management, and data analysis which impedes the speed at which data become publically available. A more frequent surveillance of population health would allow for a more timely evaluation of implemented public health policies and response to public health emergencies.

To address the high costs and time requirements associated with household surveys, higher income countries have developed and employed telephone surveys to collect population-level estimates of health and demographics [6-8]. As mobile phone ownership and access become more common globally, with 94 subscriptions per 100 inhabitants in developing countries [9], opportunities exist to leverage mobile-health technologies and communication channels to revolutionize the current methods of data collection in LMIC. Rather than conducting household surveys, respondents can now be interviewed over their own personal mobile phone through the use of short message service (SMS), interactive voice response (IVR), and computer-assisted telephone interviews (CATI) survey modalities; collectively called mobile phone surveys (MPS).

SMS surveys utilize text messages to send survey questions to participants' mobile phones. Data are then collected from participants via SMS responses to these questions. Inherent in this survey modality is the requirement of a literate population, which may be challenging in some LMICs. IVR surveys counter the challenges in SMS surveys by using automated, prerecorded questions. With IVR surveys, respondents interact with a preprogrammed database which contains both questions and a series of preset answers which are linked to a specific numeric key, or numeric response on a touch-tone phone keypad (eg, "Press 1 for Yes"). CATI surveys most closely mimic a household survey by employing human interviewers or call centers. Interviewers follow a script provided by a software program to survey participants.

The purpose of this review was to document the current landscape of MPS being used for population-level data collection in LMICs, with a focus on IVR-, SMS-, and CATI-collected data and to identify key survey metrics, such as response and completion rates for each of the MPS modalities. Such a review is currently not available in the literature, and this comprises an important assessment of current knowledge for future research [10].

Methods

We conducted a systematic search of the literature according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [11] in March-April of 2015 to find articles that used MPS for population-level data collection in LMICs. There were no restrictions on year that records were published. Using a predefined search strategy that included a range of terms for mobile phone, interactive voice response, text message, survey, questionnaire, and data collection, we searched the primary and gray literature using PubMed, Embase, Scopus, Global Health, Web of Science, Cochrane Reviews, and Proquest Digital Dissertations databases for article titles, key words, and abstracts pertaining to MPS. Search terms were uniquely created for each database to capitalize on the database's classification of articles (see Multimedia Appendix 1).

Records that matched the search criteria or were preidentified as relevant articles before the literature review (n=6) were imported into RefWorks. The 6 preidentified articles were obtained through our knowledge of the World Bank's initiative to promote MPS in LMICs. After applying an automatic filter for duplicates, the remaining abstracts and titles were manually filtered for additional study duplication and clearly irrelevant topics, such as nonhuman studies. Two research assistants conducted a primary screening of each record's titles and abstracts. Primary inclusion criteria included the following: (1) SMS used to collect data from a respondent, (2) IVR used to collect data from a respondent, (3) CATI or call centers used to collect data from a respondent, (4) a combination of the above, (5) surveys where a respondent provided answers using their mobile phone. Primary exclusion criteria included the following: (1) mobile phone used by a human enumerator to conduct an in-person survey, (2) use of mobile phone for other activities but not data collection, or (3) no indication that mobile phones were used for data collection. Records were included for secondary screening if there was any uncertainty as to whether the article included an MPS.

Articles that met primary screening criteria underwent a secondary full-text review and were screened for the following secondary inclusion and exclusion criteria. Secondary inclusion criteria included (1) SMS, IVR, or CATI was used for data collection and (2) MPS was intended to be population-representative. Secondary exclusion criteria included (1) study was not conducted in a LMIC as defined by the World Bank [12], (2) respondents were interviewed in-person by an enumerator using a mobile phone, and (3) surveys were sent to respondent's landline telephone number. Records were included independent of the survey's content (ie, health, agriculture, economics, and so on). Articles that were written in languages other than English were not reviewed.

Two research assistants extracted data from all included articles. Disagreements in data extraction were resolved by consensus between the two research assistants. Data were entered into an

XSL•FO
RenderX

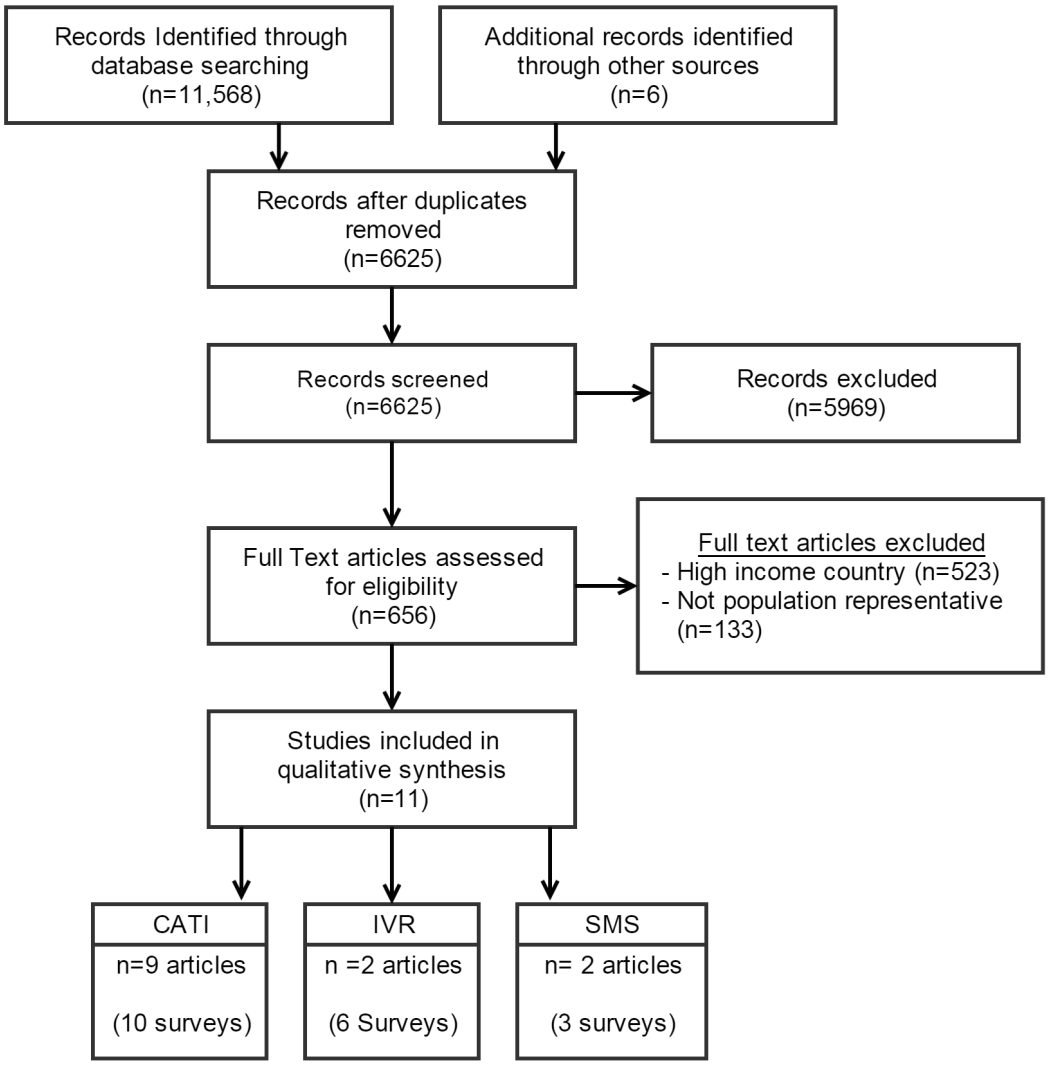
Excel (Microsoft) worksheet and grouped by survey modality to populate tables. For panel surveys where the same respondent answered a series of surveys over time, the response, completion, and refusal rates from the first round of surveys were abstracted and presented in the identified manuscripts.

Results

Overview

The literature search identified 11,568 records. After removing duplicates, 6625 records underwent a primary screening of titles and abstracts (Figure 1). Full-text articles (n=656) were then screened using secondary inclusion and exclusion criteria. Overall, we identified 11 articles that employed 19 MPS (CATI, IVR, or SMS) surveys to collect population estimates across a range of topics. Nine articles presented results for 10 CATI surveys (10/19, 53%) [13-21]. Two articles discussed the findings of 6 IVR surveys (6/19, 32%) [13,22]. Three SMS surveys were identified from 2 articles (3/19, 16%) [13,23].

Figure 1. Flow diagram of the study. CATI: computer-assisted telephone interview; IVR: interactive voice response; SMS: short message service. One article included surveys for CATI (n=2), IVR (n=2), and SMS (n=2).



CATI Surveys

The majority of MPS implemented in LMICs were conducted by human interviewers, typically stationed at call centers equipped with CATI software (Table 1). The locations and questionnaire topics were diverse; CATI surveys were conducted in Bangladesh, Brazil, Honduras, Lebanon, Liberia, Mali, Peru, South Sudan, and Tanzania and covered topics on health and socioeconomics.

Of the 10 identified CATI surveys, 60% (6/10) were implemented through the World Bank or as part of the Listening to Africa (L2A) and Listening to Latin America and the Caribbean (L2LAC) Initiatives [13,14,16-18]. In these initiatives, a population-representative sample of households was drawn and a baseline household visit was made; survey staff interviewed the selected household member in-person and provided training on how to answer future mobile phone panel surveys (Panel MPS). Panel MPS were typically sent monthly

XSL•FO
RenderX

to collect information on general welfare questions, such as household assets, food security, and employment [24]. The literature review identified one Panel MPS in Tanzania that was not affiliated with the World Bank and L2A or L2LAC [15]. The number of survey rounds or waves ranged from 2 to 33 with a typical interval of 3-6 weeks between each wave. In 71% (5/7) of the Panel MPS, mobile phones were provided to all participants [14-16] or only provided to those who did not already own one [13].

Three studies employed a cross-sectional CATI survey, rather than a panel MPS [19-21]. Mobile phones were not provided to any of these participants. The Lebanese survey sampled participants who had provided a phone number during the nationwide Nutrition and Noncommunicable Disease Risk Factor Survey [20]. The median time between the household and CATI survey was 1.8 months. In Bangladesh and Brazil, participants for noncommunicable diseases (NCD) risk factor surveys were sampled from a list of subscribers provided by a mobile network operator (MNO) [19] or through random digit dialing (RDD), respectively [21]. Since 2006, Brazil's Ministry of Health has conducted annual telephone surveys for risk and protective factors of NCD. Articles that presented VIGITEL surveys where the sampling frame contained only landline telephone numbers were excluded as the purpose of this review was to document MPS [25-29].

Overall, the response rates and completion rates for CATI surveys were highly variable, ranging from 30% to 98% and from 35% to 100%, respectively, although completion rates were only presented in 30% (3/10) of surveys. It is likely that for studies that did not report, the completion rate may near 100% as one study commented that it is the interviewer's job to make sure all questions are answered [13]. In the three studies that reported refusal rate, estimates ranged from 2% to 8% [17,20,21]. For Panel MPS, typically, panel attrition was highest at the first CATI following the household baseline survey, with attrition and nonresponse rates plateauing over the duration of the panel.

Varying airtime incentive amounts, tied to survey completion, were randomized in 40% (4/10) of CATI surveys [13,14,18], all of which were Panel MPS, to evaluate their effect on survey response and completion rates. In two surveys that did not contain a control arm (ie, no incentive), there was either no discernible effect between the low and high incentive amount on response rates [18], or the higher incentive arm had lower response rate as compared with the lower incentive arm [14]. In Peru and Honduras, panelists were randomized to one of the following three arms: (1) no incentive, (2) US $1 airtime, and (3) US $5 airtime [13]. In Honduras, both incentive arms significantly improved survey response throughout the panel, as compared with the control arm. In Peru, results were not disaggregated by survey modality (CATI, IVR, and SMS). The authors reported no appreciable difference in the first survey's response rate by the study arm; with similar gains in the two incentive arms at minimizing panel attrition over the duration of the study. Of note, the study's authors indicate that the incentive arm contained the majority of people who were provided a study-sponsored mobile phone. An additional three panel surveys provided a fixed US $1-2 airtime incentive to all panelists [15-17]. Incentives were not used in the three cross-sectional surveys [19-21].

XSL•FO
RenderX

Table 1. Computer-assisted telephone interviews (CATI) or human operator-administered surveys (n=10 surveys, 9 articles).

Author	Country (sample size)	Survey type	Sampling frame	Phone given	Response %[a] (completion %)	Average time to complete (# questions)
Ballivian et al [13]	Peru (n=384)	Panel (n=6 waves)	Household collected	If not owned	51% 100%	(10)
	Honduras (n=600)	Panel (n=2 waves)	Household collected	If not owned	88%	(10)
Demombynes et al [14]	South Sudan (n=1007)	Panel (n=4 waves)	Household collected	Yes	69%	15-20 min (16-26)
Dillon [15]	Tanzania (n=195)	Panel (n=14 waves)	Household collected	Yes	98% overall	≈27 min
Etang-Ndip et al [16]	Mali (n=501)	Panel (n=6 waves)	Household collected	Yes	99%	20-25 min
Himelein [17]	Liberia (n=2137)	Panel (n=2 waves)	Household collected	No	30%	≈15 min
Hoogeven et al [18]	Tanzania (n=458)	Panel (n=33 waves)	Household collected	No	75% overall	19 min
Islam et al [19]	Bangladesh (n=3378)	Cross-sectional	Mobile network operator	No	61%	
Mahfoud et al [20]	Lebanon (n=771)	Cross-sectional	Household collected	No	(82%)	8 min
Moura et al [21]	Brazil (n=1207)	Cross-sectional	Random digit dialing	No	(≈35%)	5 min

[a]For panel surveys, the response, completion, and refusal rates listed are for the first round of MPS unless otherwise indicated.

Table 2. IVR-administered surveys (n=6 surveys, 2 articles).

Author	Country (sample size)	Survey type	Sampling frame	Phone given	Response %[a] (completion %)	Average time to complete (# (questions)
Ballivian et al [13]	Peru (n=383)	Panel (n=6 waves)	Household collected	If not owned	20% 75%	(10 Q)
	Honduras (n=600)	Panel (n=2 waves)	Household collected	If not owned	40%	(10 Q)
Leo et al [22]	Afghanistan (n=2123)	Cross-sectional	Random digit dialing	No	31% (30%)	4-5 min (10 Q)
	Ethiopia (n=2258)	Cross-sectional	Random digit dialing	No	19% (23%)	4-5 min (10 Q)
	Mozambique (n=2229)	Cross-sectional	Random digit dialing	No	9% (38%)	4-5 min (10 Q)
	Zimbabwe (n=2192)	Cross-sectional	Random digit dialing	No	8% (51%)	4-5 min (10 Q)

[a]For panel surveys, the response, completion, and refusal rates listed are for the first round of MPS unless otherwise indicated.

IVR Surveys

MPS that employed IVR were less frequently covered in existing literature (Table 2). Our literature review identified two articles that described the findings of 6 IVR surveys to collect population-level estimates [13,22]. One article employed a standardized methodology across 4 countries—Afghanistan, Ethiopia, Mozambique, and Zimbabwe—to collect demographic and standard of living information [22]. Participants were selected through RDD with a demographic quota system and were not provided a mobile phone. The remaining 2 IVR surveys were conducted in Honduras and Peru as part of the L2LAC initiative, where participants had previously completed a baseline household survey and were provided a mobile phone

XSL•FO
RenderX

as needed [13]. All 6 surveys utilized a 10-question survey; 4 surveys reported that respondents, on average, interacted with the survey for 2-3 min and for those who completed the 10-question survey, it took between 4 and 5 min [22]. Response rates were typically higher for IVR surveys that were sent to mobile phone numbers collected from a previous household survey (20% and 40%) than from those using an RDD approach (8%, 9%, 19%, and 31%).A wide range (23-75%) of survey completion rates was observed.

The effect of airtime incentives to improve survey response and completion rates [22] and panel attrition rates [13] was evaluated across the 6 surveys and produced mixed results. One article randomized RDD participants to a control arm, 4-min airtime incentive transfer, and a raffle for a 2-h airtime incentive; where participants in the two airtime arms were eligible for the incentive if the survey was completed [22]. In Zimbabwe, the transfer and raffle incentives significantly improved the proportion of participants who completed the survey; while in Mozambique, only the raffle incentive was found to be significant. A similar evaluation was conducted in Afghanistan and Ethiopia, but the authors commented that there were problems with the randomization and allocation of study arm. In Honduras, those who were randomized to either US $1 or US $5 of airtime incentive showed higher response rates than those who did not receive an incentive [13]. As described previously, response rates were not disaggregated by survey modality in Peru.

SMS Surveys

Although data collection via SMS surveys is relatively common in LMICs, very few studies aimed to collect data on a representative sample of a population (Table 3). One study sampled 982,708 phone numbers from a network of 18 million prepaid mobile phone subscribers in Mexico to participate in a surveillance program regarding influenza-like illness [23]. Mobile phone subscribers were sent a text message from the Ministry of Health, inviting them to participate in a 6-question survey. The surveillance program resulted in a 5.8% response rate. The mean age of respondents was 25 years and nearly 90% of surveys were completed within 24 h of the initial contact. No incentives were provided.

As part of the previously described L2LAC, SMS surveys were also deployed in Peru and Honduras to collect population representative estimates [13]. The response rates for the first round of SMS surveys were 30% and 45% in Peru and Honduras, respectively. Approximately 80% of participants completed the ten question survey in Peru. In Honduras, providing either US $1 or US $5 of airtime significantly improved response rate, as compared with those who did not receive any airtime incentive.

Comparison of Survey Metrics Across Different MPS Modalities

Only two surveys compared key survey metrics such as response and completion rates across MPS. The response rate for the first round of Panel MPS was highest for CATI (Honduras, 88%; Peru, 51%), followed by SMS (45%; 30%) and IVR (40%; 20%) [13]. In Peru, CATI showed a 100% completion rate; with completion rates of 80% and 75% in SMS and IVR surveys, respectively. In the same set of surveys, the reliability of the respondent's answer was assessed through a test-retest procedure. Cronbach alpha coefficient for CATI, IVR, and SMS were .69, .86, and .74, respectively, indicating that IVR resulted in the most reliable measurements. Of note, such survey metrics have not been compared across survey modalities using sampling frames other than household collected phone numbers (eg, RDD or MNO provided).

Excluded Studies

Several large SMS surveys were identified but were excluded because they did not seek to attain representativeness. Two SMS surveys recruited participants through social media platforms. [30,31] Demographic information regarding the users of these platforms was not included thus making it difficult to assess the representativeness of respondents. Two studies attempted to achieve a subnational sample through opt-in recruitment, potentially introducing selection bias [32,33]. Numerous studies used SMS and IVR surveys as a data collection tool within a research study [34-54] or as a surveillance instrument for health care workers [55-59] and were excluded from the review.

Discussion

Principal Findings

Our literature review identified very few reports of MPS being used to collect population-level estimates. CATI surveys (n=10), most frequently relying on a household baseline survey to collect mobile phone numbers and implemented by the World Bank, were the most common type of MPS reported. When there was a household collection of mobile phone numbers, frequently, the implementing team conducted panel surveys (repeated MPS to the same respondent over time).

XSL•FO
RenderX

Table 3. SMS-administered surveys (n=3 surveys, 2 articles).

Author	Country (sample size)	Survey type	Sampling frame	Phone provided	Response %[a] (completion %)	Average time to complete (# questions)
Ballivian et al [13]	Peru (n=677)	Panel (n=6 waves)	Household collected	If not owned	30% 80%	(10 Q)
	Honduras (n=600)	Panel (n=7 waves)	Household collected	If not owned	45%	(10 Q)
Lajous et al [23]	Mexico (n=982,708)	Cross-sectional	Mobile network operator	No	6%	(6 Q)

[a]For panel surveys, the response, completion, and refusal rates listed are for the first round of MPS unless otherwise indicated.

The selection of the MPS modality has important downstream impacts on costs, survey metrics, and data quality, with each modality having its strengths and weaknesses [13,14] (Table 4). Evidence from one study that compared costs across the three modalities found that SMS and IVR surveys are less expensive than CATI surveys [13]. The primary cost of IVR and SMS surveys are airtime needed to deliver the survey, with additional costs for initial programming and monitoring survey delivery. CATI surveys, in addition to the cost of airtime and programming, also require personnel—human interviewers and supervisors—to conduct the survey, making their delivery more costly than IVR or SMS surveys [13]. The higher costs of CATI surveys are partially offset by the advantage of having a human to conduct the survey. This offers an opportunity for personalized responses to clarify any confusion a respondent may have, potentially resulting in higher quality data and lower levels of survey attrition. This benefit is supported from surveys in Peru and Honduras where response and completion rates were highest for CATI, as compared with IVR and SMS surveys. Additional studies that use a standardized approach to examine the effect of survey modality on survey response, completion, and refusal rates are needed.

Table 4. Strengths and weaknesses of mobile phone surveys (MPS) by modality (adapted from Demombynes (2013) and Ballivian (2013)).

Strengths	Weaknesses
Computer-assisted telephone interview (CATI)	
Respondent's familiarity with a phone call interaction	Resource intensive (operators, supervisors, training)
Operators can clarify questions	Inter-rater reliability concerns
Ability to build rapport with respondents	Potential for interviewer bias
Does not require respondents to be literate	Respondents may be less truthful for sensitive questions
	Requires sustained network signal
Interactive voice Response (IVR)	
Mimics a phone call	Requires sustained network signal
Does not require respondents to be literate	Respondents may not be familiar with "robot" calls
Automated surveys allows for quick data collection	Potential for respondent to be distracted while answering the survey
Minimizes interviewer bias	Poor audio quality of some phones
Less expensive than CATI due to its automation	
Short Message Service (SMS)	
Respondents answer at their convenience	May not reach illiterate respondents
Automated surveys allows for quick data collection	Requires network signal, possibility of lost messages
Minimizes interviewer bias	Question length limited by character count
Less expensive than CATI due to its automation	Inbox can become full

The majority of identified studies relied on household-collected mobile phone numbers as the sampling frame [13-18,20]. Like the MPS modality, the choice of sampling frame has implications on cost, key survey metrics, and potential representatives of a MPS. In an RDD approach there will be a significant proportion of randomly generated telephone numbers that do not exist or are not registered [60]; this represents an added cost, particularly for CATI surveys and their reliance on human operators, as more telephone calls need to be made in order to achieve the survey's sample size. Similarly, and dependent on the equation used for calculation [61], the response and completion rates may appear to be artificially lower in an RDD sample as compared with sampling frames, such as

XSL•FO
RenderX

household collected or MNO-provided ones, which ensure that the mobile phone numbers collected are active.

The use of incentives to improve response and completion (ie, cooperation) rates in telephone and postal surveys in high-income countries is well-documented [62]. Similarly, incentivizing participants through the provision of free airtime has the potential to increase the response and completion rates and the demographic representativeness of MPS, yet the findings from the few randomized trials provides inconclusive evidence on whether these interventions are effective [13,14,18,22]. Additional research studies on the use of airtime incentives and other mechanisms to improve survey performance and outcomes are needed.

Access to a mobile phone and mobile network coverage are implicit factors in a MPS's ability to generate population-representative estimates. Moreover, the "digital divide" phenomenon, where mobile phone ownership is associated with socioeconomic status, may also pose challenges with obtaining representative estimates—although evidence suggests this divide is shrinking [63]. To increase the likelihood of a survey's representativeness, household sampling methodologies can be applied to obtain a sample of household-collected mobile phone numbers. However, this requires an initial investment of human and financial resources to collect the phone numbers and is more appropriate for cohort studies or panel surveys where the initial investment will be recouped with each subsequent survey. An RDD sampling frame is more suitable for cross-sectional surveys and is the standard sampling approach for telephone surveys [60]. Our review identified very few MPS that employed RDD, but the evidence suggests that it is feasible to obtain a representative sample and that it is dependent on the saturation levels of mobile phone ownership and, to a lesser extent, linguistic fractionalization [22]. Still, options exist for obtaining population-representative results using RDD [64].

Limitations

The literature review and its inclusion and exclusion criteria identified very few articles that employed SMS (n=2 articles) or IVR surveys (n=2) to collect population representative estimates. There are three potential reasons for the infrequent use of IVR and SMS surveys. First, the search terms used in our literature review did not identify all relevant articles. Second, the pilot testing results from the L2A and L2LAC initiatives may have artificially driven the overrepresentativeness of CATI surveys identified in our literature review. Before the implementation of the initiatives, a series of pilot tests identified that CATI surveys yielded higher completion rates than IVR and SMS surveys; leading the World Bank to adopt the CATI modality as its preferred survey. Thirdly, the MPS field is in its infancy phase and there may truly be very few reports of attempts at population-representative surveys using IVR and SMS. An additional limitation is the presentation of the response and completion rates. The majority of the manuscripts did not present the equations used to calculate these rates, as recommended by the American Association for Public Opinion Research [61].

Conclusions

In conclusion, the state of MPS to collect population level estimates of health and other indicators remains nascent. Additional research that directly compares the costs, key survey metrics such as contact, response, completion, and refusal rates, and demographic representativeness across the different survey modalities is needed [10]. Still, if MPS are found to produce valid and reliable data, their use has the potential to compliment traditional household surveys and benefit existing surveillance efforts by leveraging their lower costs to allow for a more frequent monitoring of the population's health.

Acknowledgments

The studies described as part of the research agenda are funded by the Bloomberg Philanthropies. This funding agency had no role in the preparation of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of search terms used.

[PDF File (Adobe PDF File), 10KB - jmir_v19i5e139_app1.pdf]

References

1. Thacker S, Berkelman R. Public health surveillance in the United States. Epidemiol Rev 1988;10:164-190. [Medline: 3066626]
2. Lukwago L, Nanyunja M, Ndayimirije N, Wamala J, Malimbo M, Mbabazi W, et al. The implementation of Integrated Disease Surveillance and Response in Uganda: a review of progress and challenges between 2001 and 2007. Health Policy Plan 2013 Jan;28(1):30-40 [FREE Full text] [doi: 10.1093/heapol/czs022] [Medline: 22669899]

XSL•FO
RenderX

3. Somda ZC, Perry HN, Messonnier NR, Djingarey MH, Ki SO, Meltzer MI. Modeling the cost-effectiveness of the integrated disease surveillance and response (IDSR) system: meningitis in Burkina Faso. PLoS One 2010 Sep 28;5(9):e13044 [FREE Full text] [doi: 10.1371/journal.pone.0013044] [Medline: 20927386]
4. Somda ZC, Meltzer MI, Perry HN, Messonnier NE, Abdulmumini U, Mebrahtu G, et al. Cost analysis of an integrated disease surveillance and response system: case of Burkina Faso, Eritrea, and Mali. Cost Eff Resour Alloc 2009 Jan 08;7:1 [FREE Full text] [doi: 10.1186/1478-7547-7-1] [Medline: 19133149]
5. Toscano CM, Vijayaraghavan M, Salazar-Bolaños HM, Bolaños-Acuña HM, Ruiz-González AI, Barrantes-Solis T, iVPD Working Team. Cost analysis of an integrated vaccine-preventable disease surveillance system in Costa Rica. Vaccine 2013;31(Suppl 3):C88-C93 [FREE Full text] [doi: 10.1016/j.vaccine.2013.05.036] [Medline: 23777698]
6. Liu B, Brotherton JM, Shellard D, Donovan B, Saville M, Kaldor JM. Mobile phones are a viable option for surveying young Australian women: a comparison of two telephone survey methods. BMC Med Res Methodol 2011;11:159 [FREE Full text] [doi: 10.1186/1471-2288-11-159] [Medline: 22114932]
7. Iachan R, Pierannunzi C, Healey K, Greenlund KJ, Town M. National weighting of data from the Behavioral Risk Factor Surveillance System (BRFSS). BMC Med Res Methodol 2016;16(1):155 [FREE Full text] [doi: 10.1186/s12874-016-0255-7] [Medline: 27842500]
8. Brick JM, Edwards WS, Lee S. Sampling telehone numbers and adults, interview length, and weighting in the California Health Interview Survey cell phone pilot study. Public Opin Q 2007;71(5):793-813. [doi: 10.1093/poq/nfm052]
9. International Telecommunication Union. The World in 2016: ICT Facts and Figures URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2016-12-12] [WebCite Cache ID 6mgkxF0WI]
10. Hyder A, Wosu A, Gibson D, Labrique A, Ali J, Pariyo G. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246]
11. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264-9, W64. [Medline: 19622511]
12. The World Bank. datahelpdesk.worldbank. World Bank Country and Lending Groups URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2017-04-08] [WebCite Cache ID 6pZjlupkO]
13. Ballivian A, Azevedo J, Durbin W, Rios J, Godoy J, Borisova C, et al. catalog.ihsn. Washington DC: The World Bank; 2013. Listening to LAC: Using Mobile Phones for High Frequency Data Collection: Final Report URL: http://catalog.ihsn.org/index.php/catalog/4775/related_materials [accessed 2017-04-08] [WebCite Cache ID 6pZkHmMxa]
14. Demombynes G, Gubbins P, Romeo A. openknowledge.worldbank. Washington, DC: World Bank; 2013. Challenges and Opportunities of Mobile Phone-Based Data Collection Evidence from South Sudan URL: https://openknowledge.worldbank.org/handle/10986/12169 [accessed 2017-04-08] [WebCite Cache ID 6pZkZbibF]
15. Dillon B. Using mobile phones to collect panel data in developing countries. J Int Dev 2011;24(4):518-527. [doi: 10.1002/jid.1771]
16. Etang-Ndip A, Hoogeveen JG, Lendorfer J. openknowledge.worldbank. Washington, DC: World Bank; 2015. Socioeconomic impact of the crisis in North Mali on displaced people URL: https://openknowledge.worldbank.org/handle/10986/21868 [accessed 2017-04-08] [WebCite Cache ID 6pZkoxY0V]
17. Himelein K. openknowledge.worldbank. Washington, DC: World Bank; 2015. The socio-economic impacts of Ebola in Liberia: results from a high frequency cell phone survey URL: https://openknowledge.worldbank.org/handle/10986/21624 [accessed 2017-04-08] [WebCite Cache ID 6pZl3bmci]
18. Hoogeveen J, Croke K, Dabalen A, Demombynes G, Giugale M. Collecting high frequency panel data in Africa using mobile phone interviews. Can J Dev Stud 2014;35(1):186-207. [doi: 10.1080/02255189.2014.876390]
19. Islam K, Rahman M, Sharif A, Al KA, Alam Z, Rahman M. Introducing mobile phone for interview in surveillance system in Bangladesh: validation of the method. ASTMH 2013;89(5 Suppl 1):78-152. [doi: 10.4269/ajtmh.2013.89.78]
20. Mahfoud Z, Ghandour L, Ghandour B, Mokdad AH, Sibai AM. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? Field methods 2014;27(1):39-54. [doi: 10.1177/1525822X14540084]
21. Moura EC, Claro RM, Bernal R, Ribeiro J, Malta DC, Morais Neto O. A feasibility study of cell phone and landline phone interviews for monitoring of risk and protection factors for chronic diseases in Brazil. Cad Saude Publica 2011;27(2):277-286 [FREE Full text] [Medline: 21359464]
22. Leo B, Morello R, Mellon J, Peixoto T, Davenport S. cgdev. 2015. Do Mobile Surveys Work in Poor Countries URL: https://www.cgdev.org/publication/do-mobile-phone-surveys-work-poor-countries-working-paper-398 [accessed 2017-04-08] [WebCite Cache ID 6pZmoclrN]
23. Lajous M, Danon L, López-Ridaura R, Astley CM, Miller JC, Dowell SF, et al. Mobile messaging as surveillance tool during pandemic (H1N1) 2009, Mexico. Emerg Infect Dis 2010;16(9):1488-1489 [FREE Full text] [doi: 10.3201/eid1609.100671] [Medline: 20735942]
24. The World Bank. Worldbank. Listening to Africa 2016 URL: http://www.worldbank.org/en/programs/listening-to-africa [accessed 2017-04-08] [WebCite Cache ID 6pZn8yVpL]

XSL•FO
RenderX

25. Andrade SS, Malta DC, Iser BM, Sampaio PC, Moura LD. Prevalence of self-reported arterial hypertension in Brazilian capitals in 2011 and analysis of its trends in the period between 2006 and 2011. Rev Bras Epidemiol 2014;17(Suppl 1):215-226. [doi: 10.1590/1809-4503201400050017] [Medline: 25054265]
26. Ferreira AD, César CC, Malta DC, Andrade AC, Ramos CG, Proietti FA, et al. Validity of data collected by telephone survey: a comparison of VIGITEL 2008 and 'Saúde em Beagá' survey. Rev Bras Epidemiol 2011;14(Suppl 1):16-30 [FREE Full text] [Medline: 22002139]
27. Francisco PM, Barros MB, Segri NJ, Alves MC, Cesar CL, Malta DC. Comparison of estimates for the self-reported chronic conditions among household survey and telephone survey--Campinas (SP), Brazil. Rev Bras Epidemiol 2011;14(Suppl 1):5-15 [FREE Full text] [Medline: 22002138]
28. Iser BP, Malta DC, Duncan BB, de Moura L, Vigo A, Schmidt MI. Prevalence, correlates, and description of self-reported diabetes in brazilian capitals - results from a telephone survey. PLoS One 2014;9(9):e108044 [FREE Full text] [doi: 10.1371/journal.pone.0108044] [Medline: 25255096]
29. Malta DC, Andrade SC, Claro RM, Bernal RT, Monteiro CA. Trends in prevalence of overweight and obesity in adults in 26 Brazilian state capitals and the Federal District from 2006 to 2012. Rev Bras Epidemiol 2014;17(Suppl 1):267-276 [FREE Full text] [Medline: 25054269]
30. Murgor M. A mobile phone-based survey on knowledge of cervical cancer HPV vaccination in Kenya. Asia Pac J Clin Oncol 2014;10(S9):19-20. [doi: 10.1111/ajco.12332]
31. Weimann E, Stuttaford M. Consumers' perspectives on national health insurance in South Africa: using a mobile health approach. JMIR Mhealth Uhealth 2014 Oct 28;2(4):e49 [FREE Full text] [doi: 10.2196/mhealth.3533] [Medline: 25351980]
32. de Lepper AM, Eijkemans MJ, van Beijma H, Loggers J, Tuijn C, Oskam L. Response patterns to interactive SMS health education quizzes at two sites in Uganda: a cohort study. Trop Med Int Health 2013;18(4):516-521 [FREE Full text] [doi: 10.1111/tmi.12059] [Medline: 23297650]
33. Vahdat HL, L'Engle KL, Plourde KF, Magaria L, Olawo A. There are some questions you may not ask in a clinic: providing contraception information to young people in Kenya using SMS. Int J Gynaecol Obstet 2013;123(Suppl 1):e2-e6 [FREE Full text] [doi: 10.1016/j.ijgo.2013.07.009] [Medline: 24012514]
34. Andreatta P, Debpuur D, Danquah A, Perosky J. Using cell phones to collect postpartum hemorrhage outcome data in rural Ghana. Int J Gynaecol Obstet 2011;113(2):148-151. [doi: 10.1016/j.ijgo.2010.11.020] [Medline: 21420086]
35. Curran K, Mugo NR, Kurth A, Ngure K, Heffron R, Donnell D, et al. Daily short message service surveys to measure sexual behavior and pre-exposure prophylaxis use among Kenyan men and women. AIDS Behav 2013;17(9):2977-2985 [FREE Full text] [doi: 10.1007/s10461-013-0510-4] [Medline: 23695519]
36. de Tolly KM, Constant D. Integrating mobile phones into medical abortion provision: intervention development, use, and lessons learned from a randomized controlled trial. JMIR Mhealth Uhealth 2014;2(1):e5 [FREE Full text] [doi: 10.2196/mhealth.3165] [Medline: 25098569]
37. Du X, Wang W, Helena van Velthoven M, Chen L, Scherpbier RW, Zhang Y, et al. mHealth Series: text messaging data collection of infant and young child feeding practice in rural China - a feasibility study. J Glob Health 2013;3(2):020403 [FREE Full text] [doi: 10.7189/jogh.03.020403] [Medline: 24363921]
38. Haberer J, Kiwanuka J, Nansera D, Wilson IB, Bangsberg DR. Challenges in using mobile phones for collection of antiretroviral therapy adherence data in a resource-limited setting. AIDS Behav 2010;14(6):1294-1301 [FREE Full text] [doi: 10.1007/s10461-010-9720-1] [Medline: 20532605]
39. Haberer J. Real-time HIV antiretroviral therapy adherence monitoring in a resource-limited setting. Ann Behav Med 2011;41(1 Supplement):S143.
40. Irani M, Abdoli S, Bijan I, Parvizy S, Fatemi N, Amini M. Strategies to overcome type 1 diabetes-related social stigma in the Iranian society. Iran J Nurs Midwifery Res 2014;19(5):456-463 [FREE Full text] [Medline: 25400672]
41. Kew S. Text messaging: an innovative method of data collection in medical research. BMC Res Notes 2010;3:342 [FREE Full text] [doi: 10.1186/1756-0500-3-342] [Medline: 21172018]
42. Kwon HS, Cho JH, Kim HS, Lee JH, Song BR, Oh JA, et al. Development of web-based diabetic patient management system using short message service (SMS). Diabetes Res Clin Pract 2004;66(Suppl 1):S133-S137. [doi: 10.1016/j.diabres.2003.10.028] [Medline: 15563964]
43. Li Y, Wang W, van Velthoven MH, Chen L, Car J, Rudan I, et al. Text messaging data collection for monitoring an infant feeding intervention program in rural China: feasibility study. J Med Internet Res 2013 Dec 04;15(12):e269 [FREE Full text] [doi: 10.2196/jmir.2906] [Medline: 24305514]
44. Li Y, Wang W, Wu Q, van Velthoven MH, Chen L, Du X, et al. Increasing the response rate of text messaging data collection: a delayed randomized controlled trial. J Am Med Inform Assoc 2015;22(1):51-64 [FREE Full text] [doi: 10.1136/amiajnl-2014-002845] [Medline: 25332355]
45. Mukadi P, Gillet P, Lukuka A, Mbatshi J, Otshudiema J, Muyembe J, et al. External quality assessment of reading and interpretation of malaria rapid diagnostic tests among 1849 end-users in the Democratic Republic of the Congo through Short Message Service (SMS). PLoS One 2013;8(8):e71442 [FREE Full text] [doi: 10.1371/journal.pone.0071442] [Medline: 23967211]

XSL•FO
RenderX

46. Munro ML, Lori JR, Boyd CJ, Andreatta P. Knowledge and skill retention of a mobile phone data collection protocol in rural Liberia. J Midwifery Womens Health 2014;59(2):176-183 [FREE Full text] [doi: 10.1111/jmwh.12155] [Medline: 24655593]
47. Nejmeh B, Dean T. The charms application suite: A community-based mobile data collection and alerting environment for HIV/AIDS orphan and vulnerable children in Zambia. In: Service-Learning in the Computer and Information Sciences: Practical Applications in Engineering Education. Hoboken, NJ: John Wiley & Sons; 2012.
48. Rotheram-Borus M, Richter L, Van Rooyen H, van Heerden A, Tomlinson M, Stein A, et al. Project Masihambisane: a cluster randomised controlled trial with peer mentors to improve outcomes for pregnant mothers living with HIV. Trials 2011;12:2 [FREE Full text] [doi: 10.1186/1745-6215-12-2] [Medline: 21205302]
49. Scott W, Weina PJ. Texting away malaria: a new alternative to directly observed therapy. Mil Med 2013;178(2):e255-e259. [doi: 10.7205/MILMED-D-12-00257] [Medline: 23764334]
50. Sidney K, Antony J, Rodrigues R, Arumugam K, Krishnamurthy S, D'souza G, et al. Supporting patient adherence to antiretrovirals using mobile phone reminders: patient responses from South India. AIDS Care 2012;24(5):612-617. [doi: 10.1080/09540121.2011.630357] [Medline: 22150088]
51. van Heerden AC, Norris S, Tollman S, Stein A, Richter L. Field lessons from the delivery of questionnaires to young adults using mobile phones. Soc Sci Comput Rev 2013;32(1):105-112. [doi: 10.1177/0894439313504537]
52. Yoonessi A, Ekhtiari H. Text messages as a tool for assessing public concern about drug problems. Int J Drug Policy 2013;24(6):624-627. [doi: 10.1016/j.drugpo.2013.06.002] [Medline: 23849717]
53. Zirong H, Junqing W, Coffey P, Kilbourne-Brook M, Yufeng Z, Wang C, et al. Performance of the woman's condom among couples in Shanghai, China. Eur J Contracept Reprod Health Care 2012;17(3):212-218. [doi: 10.3109/13625187.2012.663016] [Medline: 22497293]
54. Zolfaghari M, Mousavifar SA, Pedram S, Haghani H. The impact of nurse short message services and telephone follow-ups on diabetic adherence: which one is more effective? J Clin Nurs 2012;21(13-14):1922-1931. [doi: 10.1111/j.1365-2702.2011.03951.x] [Medline: 22239205]
55. Armstrong K, Liu F, Seymour A, Mazhani L, Littman-Quinn R, Fontelo P, et al. Evaluation of txt2MEDLINE and development of short messaging service-optimized, clinical practice guidelines in Botswana. Telemed J E Health 2012;18(1):14-17. [doi: 10.1089/tmj.2011.0014] [Medline: 22150636]
56. Githinji S, Kigen S, Memusi D, Nyandigisi A, Wamari A, Muturi A, et al. Using mobile phone text messaging for malaria surveillance in rural Kenya. Malar J 2014;13:107 [FREE Full text] [doi: 10.1186/1475-2875-13-107] [Medline: 24642130]
57. Rajatonirina S, Heraud J, Randrianasolo L, Orelle A, Razanajatovo N, Raoelina Y, et al. Short message service sentinel surveillance of influenza-like illness in Madagascar, 2008-2012. Bull World Health Organ 2012;90(5):385-389 [FREE Full text] [doi: 10.2471/BLT.11.097816] [Medline: 22589573]
58. Randrianasolo L, Raoelina Y, Ratsitorahina M, Ravolomanana L, Andriamandimby S, Heraud J, et al. Sentinel surveillance system for early outbreak detection in Madagascar. BMC Public Health 2010;10:31 [FREE Full text] [doi: 10.1186/1471-2458-10-31] [Medline: 20092624]
59. Schuttner L, Sindano N, Theis M, Zue C, Joseph J, Chilengi R, et al. A mobile phone-based, community health worker program for referral, follow-up, and service outreach in rural Zambia: outcomes and overview. Telemed J E Health 2014;20(8):721-728 [FREE Full text] [doi: 10.1089/tmj.2013.0240] [Medline: 24926815]
60. Potthoff R. Telephone sampling in epidemiologic research: to reap the benefits, avoid the pitfalls. Am J Epidemiol 1994;139(10):967-978. [Medline: 8178785]
61. The American Asssociation for Public Opinion Research. Standard definitions: Final dispositions of case codes and outcome rates for surveys. 9th edition. Oakbrook Terrace, IL: AAPOR; 2016.
62. Singer E, Ye C. The use and effects of incentives in surveys. Ann Am Acad Pol Soc Sci 2012;645(1):112-141. [doi: 10.1177/0002716212458082]
63. Tran MC, Labrique AB, Mehra S, Ali H, Shaikh S, Mitra M, et al. Analyzing the mobile “digital divide”: changing determinants of household phone ownership over time in rural bangladesh. JMIR Mhealth Uhealth 2015;3(1):e24 [FREE Full text] [doi: 10.2196/mhealth.3663] [Medline: 25720457]
64. Labrique AB, Blynn E, Ahmed S, Gibson DG, Pariyo GW, Hyder AA. Health surveys using mobile phones in developing countries: automated active strata monitoring and other statistical considerations for improving precision and reducing biases. J Med Internet Res 2017;19(5):e121 [FREE Full text] [doi: 10.2196/jmir.7329]

Abbreviations

CATI: computer-assisted telephone interviews
IVR: interactive voice response
L2A: Listening to Africa
L2LAC: Listening to Latin America and the Caribbean
LMIC: low- and middle-income countries
MNO: mobile network operator

XSL•FO
RenderX

MPS: mobile phone surveys
NCD: noncommunicable diseases
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RDD: random digit dialing
SMS: short message service

Edited by E Rosskam; submitted 30.01.17; peer-reviewed by C Hoe, P Ndebele; comments to author 16.02.17; revised version received 11.03.17; accepted 11.03.17; published 05.05.17.

Please cite as:
Gibson DG, Pereira A, Farrenkopf BA, Labrique AB, Pariyo GW, Hyder AA
Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review
J Med Internet Res 2017;19(5):e139
URL: http://www.jmir.org/2017/5/e139/
doi:10.2196/jmir.7428
PMID:28476725

©Dustin G Gibson, Amanda Pereira, Brooke A Farrenkopf, Alain B Labrique, George W Pariyo, Adnan A Hyder. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Building the Evidence Base for Remote Data Collection in Low- and Middle-Income Countries: Comparing Reliability and Accuracy Across Survey Modalities

Abigail R Greenleaf[1], MPH; Dustin G Gibson[2], MS, PhD; Christelle Khattar[2]; Alain B Labrique[2], MHS, MS, PhD; George W Pariyo[2], MBChB, PhD

[1]Johns Hopkins Bloomberg School of Public Health, Department of Population, Family and Reproductive Health, Baltimore, MD, United States
[2]Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD, United States

Corresponding Author:
Abigail R Greenleaf, MPH
Johns Hopkins Bloomberg School of Public Health
Department of Population, Family and Reproductive Health
615 N Wolfe Street
Baltimore, MD, 21205
United States
Phone: 1 410 955 3543
Fax: 1 410 955 2303
Email: agreenleaf@jhu.edu

Abstract

Background: Given the growing interest in mobile data collection due to the proliferation of mobile phone ownership and network coverage in low- and middle-income countries (LMICs), we synthesized the evidence comparing estimates of health outcomes from multiple modes of data collection. In particular, we reviewed studies that compared a mode of remote data collection with at least one other mode of data collection to identify mode effects and areas for further research.

Objective: The study systematically reviewed and summarized the findings from articles and reports that compare a mode of remote data collection to at least one other mode. The aim of this synthesis was to assess the reliability and accuracy of results.

Methods: Seven online databases were systematically searched for primary and grey literature pertaining to remote data collection in LMICs. Remote data collection included interactive voice response (IVR), computer-assisted telephone interviews (CATI), short message service (SMS), self-administered questionnaires (SAQ), and Web surveys. Two authors of this study reviewed the abstracts to identify articles which met the primary inclusion criteria. These criteria required that the survey collected the data from the respondent via mobile phone or landline. Articles that met the primary screening criteria were read in full and were screened using secondary inclusion criteria. The four secondary inclusion criteria were that two or more modes of data collection were compared, at least one mode of data collection in the study was a mobile phone survey, the study had to be conducted in a LMIC, and finally, the study should include a health component.

Results: Of the 11,568 articles screened, 10 articles were included in this study. Seven distinct modes of remote data collection were identified: CATI, SMS (singular sitting and modular design), IVR, SAQ, and Web surveys (mobile phone and personal computer). CATI was the most frequent remote mode (n=5 articles). Of the three in-person modes (face-to-face [FTF], in-person SAQ, and in-person IVR), FTF was the most common (n=11) mode. The 10 articles made 25 mode comparisons, of which 12 comparisons were from a single article. Six of the 10 articles included sensitive questions.

Conclusions: This literature review summarizes the existing research about remote data collection in LMICs. Due to both heterogeneity of outcomes and the limited number of comparisons, this literature review is best positioned to present the current evidence and knowledge gaps rather than attempt to draw conclusions. In order to advance the field of remote data collection, studies that employ standardized sampling methodologies and study designs are necessary to evaluate the potential for differences by survey modality.

(J Med Internet Res 2017;19(5):e140) doi:10.2196/jmir.7331

XSL•FO
RenderX

KEYWORDS

mHealth; developing countries; Africa South of the Sahara; cell phones; health surveys; reproducibility of results; surveys and questionnaires; text messaging; interviews as topic; humans; research design; data collection methods

Introduction

In low- and middle-income countries (LMICs), where vital registration, surveillance, and health record systems are underdeveloped [1], improved modes of data collection are needed [2]. Public health practitioners could benefit from more timely estimates and indicators to better plan programs, design interventions, and assess progress. The financial and human resource burden of a large survey as well as the need for more frequent data collection, particularly as mandated by the Sustainable Development Goals [3], all justify an improved system to monitor health indicators in LMICs.

The rapid increase of mobile phone ownership in LMICs offers a platform for low-cost, frequent data collection. Urbanization, increased mobile phone network coverage, and the low cost of purchasing a mobile phone have contributed to increased mobile phone ownership in LMICs [4]. According to the International Telecommunications Union, in 2015 the number of mobile subscriptions worldwide was 98.66 per 100 people [5]. Increased mobile phone ownership presents the opportunity to survey respondents remotely, whether via short message service (SMS), computer-assisted telephone interview (CATI), interactive voice response (IVR), or Web surveys. The advantages and disadvantages of these interview modalities are discussed by Gibson et al [6]. As remote data collection becomes more common in LMICs, the reliability and accuracy of data collected should be compared with established methods, including the reference-standard household survey.

There is a well-established body of literature on mobile phones as survey instruments in high-income countries [7-11], but there is a dearth of rigorous research that compares the quality, reliability, and accuracy of remote data collection modes in LMICs. Data collection mode can influence social desirability bias, can impact response rates, or can change the cognitive process for answer retrieval [9]. Although misreporting of sensitive behaviors has long been of interest to survey methodologists, a superior interviewing tool is yet to be identified for use in LMICs [12]. Cognitive models illustrate how mode of data collection affects information retrieval, judgments about the appropriate responses, and answer choices [11]. Notably, response rates are traditionally lower for remote data collection compared with face-to-face (FTF) data collection [10].

The purpose of this literature review was to identify and synthesize the available literature from LMICs that compare a mode of remote health data collection with at least one other data collection mode. We also discuss reliability and construct validity across measures. By synthesizing the research that compares a mode of remote data collection to another mode, we identify the strengths and limitations of remote modes in LMICs as well as areas for future research.

Methods

This literature review utilized the search terms and primary inclusion and exclusion criteria from a previously conducted literature review in March and April 2015 [6]. We adapted search terms for mobile phone, IVR, text message, survey, questionnaire, and data collection to each database's classification system to query seven databases of peer-reviewed and grey literature. The abstracts and titles were screened against a set of primary inclusion and exclusion criteria [6]. The primary inclusion criteria required that the research was collected from the respondent by SMS, IVR, CATI, or via mobile phone.

Once all articles that met the primary inclusion and exclusion criteria were identified, two of the authors (AG and CK) independently reviewed the articles using the secondary inclusion and exclusion criteria, as listed in Textbox 1.

Summary briefs, inaccessible full texts, and a number of manuscripts that used a remote form of data collection but did not compare the method against a standard were excluded. The references of articles included in this review were searched to find relevant publications that were not identified by the literature search. Surveys included in this review were not required to be nationally representative.

Textbox 1. Secondary inclusion and exclusion criteria.

Secondary inclusion criteria

- Study conducted in LMIC as defined by the World Bank
- Two or more modes of data collection are compared in the study
- At least one of the data collection modes is remote
- The survey includes questions about health

Secondary exclusion criteria

- Studies without a human component to the research
- Studies that collect only adherence information or that strictly examine the use of reminders for health-seeking behaviors and outcomes
- Studies that compare modes of facility-based surveillance data collection

XSL•FO
RenderX

Table 1. Categories of data collection included in this literature review.

Survey administration	Remote	In-person
Interviewer administered	CATI[a]	FTF[b]
Self-administered	IVR[c]	IVR[d]
	SMS[e,f]	SAQ[g]
	Postal SAQ	
	Web MP[h]	
	Web PC[i]	

[a]CATI: computer-assisted telephone interview.

[b]FTF: face-to-face.

[c]IVR: interactive voice response.

[d]IVR is traditionally administered remotely but can be administered in-person by the interviewer by handing a phone to the respondent that plays an IVR survey.

[e]SMS: short message service.

[f]Two forms of SMS surveys were included in this study: single sitting (survey completed at one time) and modular (an SMS sent each day until survey is completed).

[g]SAQ: self-administered questionnaire.

[i]Web PC: Web on personal computer.

[h]Web MP: Web on mobile phone.

A standardized data collection tool containing inclusion and exclusion criteria was completed by two reviewers for each screened article. Variables in the extraction form included study design, study location, data collection mode, response rates, sensitive questions, study limitations, cost, and findings. Once compiled, the two reviewers discussed any differences in their respective reviews to make final inclusion or exclusion decisions. The two reviewers relied on a third person to clarify any inclusion disagreements.

Included articles were grouped by location of respondent in relation to interviewer (remote or in-person) and by the person administering the questionnaire (self-administered or interviewer administered; see Table 1). If the participant was in a different geographic location from the interviewer while administering the survey (no FTF interaction), data collection was classified as remote. In-person data collection was defined as FTF interviewer-respondent interaction. Self-administered was defined as surveys where respondents answer without questions from the interviewer. Interviewer administered was defined as the interviewer speaking with the respondent to elicit responses. FTF surveys were defined as an in-person, interviewer-administered survey. CATI was the only form of remote interviewer-administered survey included in this literature review. IVR, SMS, Web surveys, and a self-administered questionnaire (SAQ) sent back via post were defined as remote, self-administered surveys. There were two examples of an in-person self-administered survey: SAQ and IVR administered on-site.

We reported inter-method reliability for articles that compared the same respondents across two or more modes. For articles that did not compare the same population across the two modes, we considered external construct validity, the degree to which a measure satisfies theoretical predictions about a measurement. Included articles were reviewed and marked for sensitive questions, defined as asking about a subject that is private or taboo [13]. In assessing the results, we considered not only statistical significance, but effect size, direction of difference, and potential confounding factors [11].

Results

Overview

The parent systematic literature search identified 11,568 records, which after removing the duplicates and adding 6 articles identified by the authors, was reduced to 6625 records (see Figure 1). The primary inclusion and exclusion criteria further decreased the number of articles to 145. After removing 126 articles that did not include a comparison mode and 9 surveillance articles, we conducted full-text abstraction on 10 articles that compared two or more modes of data collection in a LMIC, with at least one form being remote data collection (see Table 2). All but one of the articles were published between 2011 and 2015. The 10 articles collected data in 7 countries, across 4 regions (Asia, Latin America, Europe, and the Middle East); notably none took place in Sub-Saharan Africa (SSA). One article reported data from 2 countries, Honduras and Peru [14].

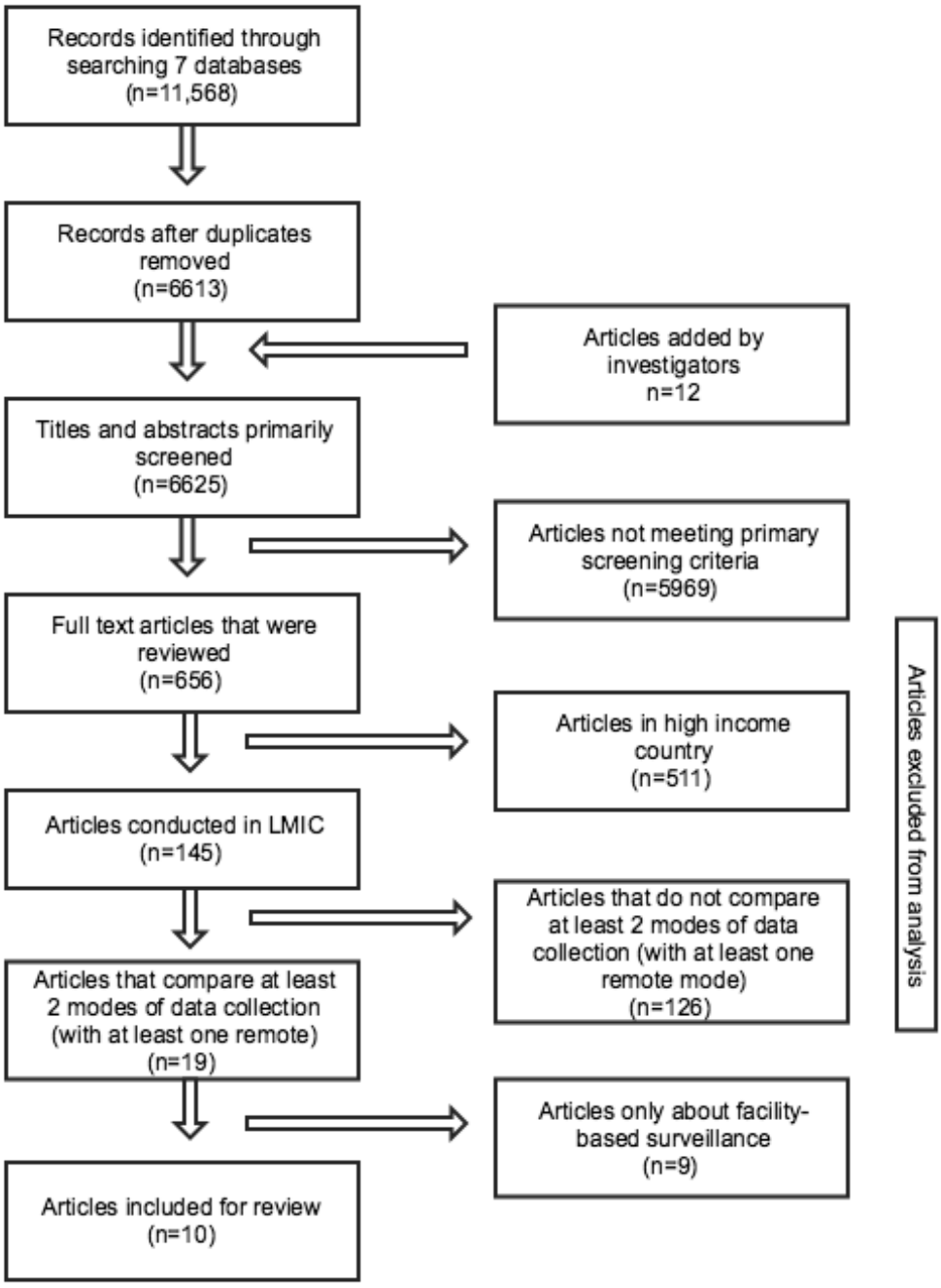
Ten distinct modes of data collection were used in the 10 studies (see Table 3). The three in-person modes were FTF [14-21], SAQ, and in-person IVR [18]. The 7 types of remote data collection included two types of phone calls, IVR (remote) [14,22] and CATI [14,16,17,20,23,24]; and three modes that required respondents to type their responses into a mobile phone or computer, including SMS (all but one were singular-design) [14,15,19,21,24] and two types of Web surveys [25,26], (one administered on a personal computer (PC) and the other taken via a mobile browser on a smartphone). The most frequent mode of data collection was FTF (5/10 studies) and second most

XSL•FO
RenderX

frequent was SMS (4/10 studies). The 10 articles made 25 comparisons, of which 12 were from a World Bank study in Latin America [14]. The most common comparison was FTF and CATI [14,16,17,20] (compared 5 times in 4 articles) and the second most common comparison was FTF to SMS [14,15,19] (compared 4 times in 3 articles).

The majority (8/10) of the articles collected cross-sectional data and did not provide respondents with a mobile phone; only the studies in Nepal and those in Honduras and Peru provided respondents with mobile phones [14,24]. Respondents were sampled in a variety of ways. Four of the identified articles were population-based studies, all of which enrolled participants FTF [14,16,17,20]. Five studies compared the same population across methods of data collection [14,15,19,20,25]. Finally, 6 of the 10 articles included sensitive questions [14,18,20,24,25,26].

Figure 1. Flowchart of articles identified and included in review.



XSL•FO
RenderX

Table 2. Types of data collection which are compared and the number of included studies.

Method #1	Method #2						
	CATI[a]	IVR[b] remote	SMS[c]: singular	SMS: modular	Web MP[d]	IVR in-person	SAQ[e] remote
CATI			1	1			
FTF[f]	5	2	4				
SAQ in-person						1	1
IVR remote	2						
Web PC[g]					2		
SAQ remote						1	
SMS: singular	2	2		1			

[a]CATI: computer-assisted telephone interview.

[f]FTF: face-to-face.

[b]IVR: interactive voice response.

[e]SAQ: self-administered questionnaire.

[c]SMS: short message service.

[g]Web PC: Web on personal computer.

[d]Web MP: Web on mobile phone.

Table 3. Comparison of in-person interviewer administered compared with remote interviewer administered.

Author (year)	Country (sample type)[a]	Data collection #1 (sample size)	Data collection #2 (sample size)
Ballivian [14] (2013)	Honduras (dependent)	FTF[b] (1500)	CATI[c] (600)
Ballivian [14] (2013)	Peru (dependent)	FTF (1500)	CATI (384)
Ferreira [16] (2011)	Brazil (independent)	FTF (4048)	CATI (440)
Francisco [17] (2011)	Brazil (independent)	FTF (2636)	CATI (2015)
Mahfoud [20] (2014)	Lebanon (dependent)	FTF (2836)	CATI (771)

[a]When participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

[b]FTF: face-to-face.

[c]CATI: computer-assisted telephone interview.

Comparison of Modes of Data Collection

In-Person Interviewer Administered Compared With Remote Interviewer Administered

We identified 5 comparisons of FTF interviews with CATI surveys in 4 articles [14,16,17,20]. One article included comparisons of FTF and CATI in both Peru and Honduras [14]. Of these 5 comparisons, 2 compared responses in independent samples [16,17] and 3 used the same population across the two modalities [14,20]. All comparisons generally showed concordance of results between modes.

An independent probability sample of respondents in Brazil who were over 18, had a landline phone, and were interviewed via CATI, produced estimates similar to an independent sample of respondents who also have a landline and who answered a household survey (FTF) [16]. The respondents contacted via CATI had statistically significant different estimates for 5 (number of household residents, mean age, schooling, smoking, health insurance) of the 18 measures compared with the FTF respondents with a landline. CATI respondents when compared with all FTF respondents (regardless of landline ownership) differed on 8 of the 18 variables, but after applying poststratification weights, only three estimates were biased.

The second study in Brazil compared FTF respondents with CATI landline respondents, sampled in the same manner as the aforementioned article. Two of the four estimates (diabetes, asthma, bronchitis, or emphysema) were the same between the two samples. The other two estimates (hypertension and

XSL•FO
RenderX

osteoporosis) had a higher reported prevalence among the CATI respondents [17]. Nonetheless, authors from both the studies concluded that the telephone survey was a rapid alternative to FTF surveys to provide global prevalence estimates.

After a FTF survey in Lebanon, half of the respondents were called on their mobile phones and asked an abridged version of the FTF questionnaire [20]. Overall, there was high concordance (kappa) between the CATI and FTF surveys. Kappa was above .8 for measures including age, health insurance, diabetes, current cigarette-smoking (highest agreement and kappa: agreement=95.6%, κ=.91), and ever cigarette-smoking (second highest agreement=93.5%, κ=.87). Kappa was between .6 and .8 for questions about current water-pipe smoking and past-year alcohol consumption. Reports of past-year alcohol consumption was slightly higher via CATI compared with FTF [20]. The authors concluded that estimates from the modes are reasonably comparable when data were stratified by age, gender, and education and that the difference in past-year alcohol consumption may be caused by social desirability bias [20].

The World Bank's study in Honduras and Peru aimed to validate a survey across four modalities: FTF, IVR, CATI, and SMS [14]. The study enrolled households who answered a 7-question FTF survey on household assets and poverty into a panel survey. The survey included questions about water and sanitation. Participants in Peru were sampled using the National Statistics office sampling frame, and households below the poverty line were oversampled. Honduras also used probabilistic sampling but used the Gallup World Poll Sampling Frame and did not oversample households below the poverty line. Regardless of mode, attrition was highest among less educated, less affluent, older, rural participants [14]. In both the countries, CATI estimates were very similar to the estimates collected FTF. Compared with the FTF survey, discordant responses from the panel in Honduras ranged from −2.1% to 0% for CATI (unreported for Peru). Furthermore, CATI had the lowest discordance with the FTF survey compared with SMS and IVR [14].

In-Person Interviewer Administered Compared With Remote Self-Administered

Three articles made 6 comparisons of in-person interviewer administered (all FTF) and either IVR or SMS (see Table 4). Two articles from China, both about infant feeding practices, used the test-retest method to compare FTF and SMS surveys [15,19]. One of the articles began with a FTF interview then followed up with a SMS survey [15], and the other article interviewed participants in the opposite order [19]. Both articles administered the second survey after a short time period (less than 24 h and less than 3 days). The study that sent 10 text messages to participants then followed up with FTF surveys had moderate to good agreement and 62.4% of questions had the same answers for both surveys [19]. All but one kappa and inter-class correlation were between .56 and .76; (the outlier kappa=.23 was for a question about the usefulness of a feeding calendar). The last question, which was a multiple-choice categorical question, about the source of feeding knowledge, had the highest agreement (85% of the 33 responses were the same across methods) and a kappa value of .76.

Data agreement in the other Chinese feeding study was inconsistent [15]. The highest agreement was a kappa of .86 for the first question on the survey which was about breastfeeding the day before. The other 4 questions had moderate to poor agreement. Data agreement was worst for dietary recall, with a kappa ranging from .02 to .36 for the 7 food categories. The authors proposed that certain terms were difficult for mothers to understand (eg, iron-fortified food, solid or semi-solid food) and that during the FTF survey, the interviewers could explain these concepts, a feat that SMS cannot achieve due to the limited characters in a text and constraints on the number of texts a respondent is willing to receive. Du et al also explored the length of time between the two surveys (3 hours compared to 8 hours) but did not find a statistically significant difference in reported measure by the length of time between surveys.

The World Bank study in Honduras and Peru compared FTF with two modes of remote self-administered: IVR and SMS. Compared with the FTF survey, discordant responses from the panel in Honduras ranged from −14.6% to 12.7% for IVR, and from −15.6% to 15.3% for SMS. Discordant responses were similar for IVR and SMS on a per question basis. The IVR and SMS responses were statistically significantly different from the FTF estimates. To assess the reliability, the same respondents were asked a question second time, within 10 weeks of the first administration of the questions. The total reliability coefficient in Honduras for SMS and FTF were quite similar (.74 and .77, respectively). IVR had the highest reliability coefficient (.86); but because the IVR results were most discordant with the other modes, the authors concluded that IVR was not a suitable mode for this survey. However, authors were satisfied with the reliability of SMS surveys in their study's context.

In-Person Self-Administered Compared With Remote Self-Administered

To assess human immunodeficiency virus (HIV)–related risk behaviors among Hong Kong migrant men who were aged 18-60 years, authors systematically sampled 2416 migrants at a customs check point in Hong Kong. Authors compared three modes of data collection (see Table 5). First, all participants completed a FTF demographic survey and then were randomized to one of the three modes (in-person IVR, SAQ returned on-site in a self-sealing envelope, and SAQ to be completed off-site and returned via post) [18]. This article also included a comparison of two in-person self-administered questionnaires (in-person IVR compared with in-person SAQ; see Table 6). The authors found differential reporting of sensitive behaviors by mode. The low response rate (only 36% of men randomized to complete the postal SAQ returned the questionnaire) limits analysis of results. Item nonresponse rates and frequency of self-reported, sensitive sexual behaviors were statistically significantly different across the three methods for all reported questions. The IVR estimates were more similar to the remote (postal) SAQ than to the in-person SAQ. The postal mode reported socially desirable answers more frequently than the other two modes (both in-person and self-administered) [18]. This study's authors suppose that subjective psychological responses, such as the perception of confidentiality, explain the lower report of undesirable behaviors in the self-administered in-person article survey compared with the other two modes.

XSL•FO
RenderX

Table 4. In-person interviewer administered compared with remote self-administered.

Author (year)	Country (sample type)[a]	Data collection #1 (sample size)	Data collection #2 (sample size)
Ballivian [14] (2013)	Honduras (dependent)	FTF[b] (1500)	SMS[c] (900)
Ballivian [14] (2013)	Honduras (dependent)	FTF (1500)	IVR[d] (600)
Ballivian [14] (2013)	Peru (dependent)	FTF (1500)	SMS (677)
Ballivian [14] (2013)	Peru (dependent)	FTF (1500)	IVR (383)
Du [15] (2013)	China (dependent)	FTF (591)	SMS (591)
Li [19] (2013)	China (dependent)	FTF (177)	SMS (99)

[a]When participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

[b]FTF: face-to-face.

[c]SMS: short message service.

[d]IVR: interactive voice response.

Table 5. In-person self-administered compared with remote self-administered.

Author (year)	Country (sample type)[a]	Data collection #1 (n)	Data collection #2 (n)
Lau [18] (2000)	Hong Kong (dependent)	In-person IVR[b] (1254)	Postal SAQ[c] (556)
		In-person SAQ (606)	Postal SAQ (556)

[a]When participants were the same across modes, we classified the sample as dependent. Different participants across modes were labeled as an independent sample.

[b]IVR: interactive voice response.

[c]SAQ: self-administered questionnaire.

Table 6. In-person self-administered compared with a second mode of in-person self-administered.

Author (year)	Country (sample type)[a]	Data collection #1 (sample size)	Data collection #2 (sample size)
Lau [18] (2000)	Hong Kong (dependent)	In-person SAQ[b] (606)	In-person IVR (1254)

[a]When participants were the same across modes, we classified the sample as dependent. Different participants across modes were labeled as an independent sample.

[b]SAQ: self-administered questionnaire.

Remote Self-Administered Compared With a Second Mode of Remote Self-Administered

Four articles make five comparisons of remote self-administered modes with a second mode of remote self-administered (see Table 7). Two articles compare Russian respondents' answers on a Web survey on a mobile phone with a Web survey on a PC. The survey that compared a different population across the two modes did not find any difference in report of sensitive behavior indicators [26]. In the second study that was a cross-over experiment, 2 of the 5 sensitive questions, namely, alcohol consumption and income, were statistically significantly different [25]. The PC-based Web survey reported higher levels of alcohol consumption and higher income. The respondents reported higher trust in data confidentiality on the Web data collected on PC compared with the Web data collected on

XSL•FO
RenderX

mobile phone. The authors also tested for interaction between gender and survey mode but did not find any statistically significant gender differences.

The World Bank study also compared SMS and IVR. The study found a higher attrition and lower survey-completion rate among IVR and SMS respondents compared with the other two modes [14]. Furthermore, panelists responding via self-administered mode were more likely to leave questions unanswered compared with CATI. Finally, SMS then IVR were estimated to be the least expensive options for data collection, compared with CATI and FTF.

In addition to the work in Peru, Honduras, and Russia, 1 article compared two modes of remote data collection. This survey was nested in a panel study in Nepal and compared CATI, and two types of SMS surveys. During single-sitting SMS interviews participants completed the survey at one time, and during module-design text interviews participants answered one question per day [24]. There were very few differences in results when comparing the two SMS modes. The modular survey did have higher nonresponse rate than the single sitting SMS survey, but the respondents in the modular design group found the survey to be significantly easier to complete than persons in the other two groups [24].

Remote Self-Administered Compared With Remote Interviewer-Administered

The research in Nepal compared the two aforementioned modes of SMS to CATI (see Table 8). They found that both text message modes increased the probability of disclosing sensitive information (eg, age of drinking onset, ever-smoking marijuana) compared with CATI, but mode did not impact the report of factual survey items (eg, marital status, age) [24]. The authors note that they are not sure whether sensitive behavior is reported more frequently via text due to decreased time pressure or increased privacy [24].

The final comparisons from the World Bank study is CATI compared with IVR and SMS. SMS, although the least expensive of the three modes, has twice the attrition rates as CATI [14]. Another important consideration about SMS from the World Bank study is that personal Internet access was reported more frequently via SMS. The World Bank study's authors hypothesize that this could be caused by younger informants who are more likely to respond to an SMS survey compared with other ages. Reliability co-efficients for SMS range from .57 (Do you consider yourself poor?) to .87 (Do you currently have a television at home?) [14]. The Cronbach alpha for IVR is higher (at .86) than SMS, and item-level reliability has a smaller range from .79 (In the last 30 days have you access the Internet or not?) to .93 (Do you currently have a television at home?) [14].

Table 7. Remote self-administered compared with a second mode of remote self-administered.

Author (year)	Country (sample type)[a]	Data collection #1 (sample size)	Data collection #2 (sample size)
Mavletova [26] (2013)	Russia (independent)	Web MP[b] (481)	Web PC[c] (532)
Mavletova and Couper [25] (2013)	Russia (dependent)	Web MP (884)	Web PC (884)
Ballivian [14] (2013)	Peru (dependent)	IVR[d] (383)	SMS[e] (677)
Ballivian [14] (2013)	Honduras (dependent)	IVR (600)	SMS (900)
West [24] (2015)	Nepal (independent)	SMS: singular (150)	SMS: modular (150)

[a]When participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

[b]Web PC: Web on personal computer.

[c]Web MP: Web on mobile phone.

[d]IVR: interactive voice response.

[e]SMS: short message service.

XSL•FO
RenderX

Table 8. Remote self-administered compared with remote interviewer-administered.

Author (year)	Country (sample type)[a]	Data collection #1 (sample size)	Data collection #2 (sample size)
West [24] (2015)	Nepal (independent)	CATI[b] (150)	SMS[c]modular (150)
West [24] (2015)	Nepal (independent)	CATI (150)	SMS singular (150)
Ballivian [14] (2013)	Peru (dependent)	CATI (384)	SMS (677)
Ballivian [14] (2013)	Honduras (dependent)	CATI (600)	SMS (900)
Ballivian [14] (2013)	Peru (dependent)	CATI (384)	IVR[d] (383)
Ballivian [14] (2013)	Honduras (dependent)	CATI (600)	IVR (600)

[a]When participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

[b]CATI: computer-assisted telephone interview.

[c]SMS: short message service.

[d]IVR: interactive voice response.

Discussion

Principal Findings

This literature review synthesizes research that compares two or more modes of data collection in LMICs, with a special focus on remote data collection. We identified 10 articles that covered a range of modes and a variety of sampling methods. Three articles collected data in East Asia, 3 in Central and South America, 2 in Russia, 1 in South Asia, and 1 in Lebanon. No articles were identified that compared in-person self-administered and remote interviewer-administered surveys (eg, CATI). In-person self-administered includes a SAQ or a computer-assisted self-interview, with or without audio. Due to both heterogeneity of outcomes among the studies and the limited number of studies, this literature review is best positioned to present the current evidence rather than attempt to draw conclusions.

The most comprehensive study was conducted by the World Bank, where researchers made six comparisons across four modes of data collection (FTF vs IVR, FTF vs CATI, FTF vs SMS, IVR vs CATI, IVR vs SMS, and CATI vs SMS) in each of two countries [14]. This is the only study in the literature review to compare IVR versus CATI and IVR versus SMS. Comparing IVR and CATI is a particularly useful comparison because if the same sampling method is used for CATI and IVR, the impact of administration of interview can be better isolated and assessed. The finding that CATI and FTF estimates produced the best criterion validity when compared with the other modes in the study is consistent with findings from other studies included in this review.

Half of the studies enrolled participants FTF. Enrolling participants FTF mitigates one of the main benefit of remote data collection—reduced data collection cost. Only two of the studies used random digit dialing (RDD) and both were limited to landlines [16,17]. No articles in this review were conducted in SSA, but we expect an increasing amount of evidence will be emerging from the area.

A minority of articles explicitly compared the profile of respondents between the two modes of data collection. By comparing sample demographics to the target population, we will better understand the respondent bias a mode may introduce. Ferreira et al found that groups with higher telephone ownership or coverage were more likely to report better health [16] and the World Bank study identified young people as more likely to respond to a SMS survey [14]. Other key information, including cost, length of the questionnaire, and the reliability of measures were not reported consistently but would provide important implementation information. It is imperative to use American Association for Public Opinion Research Reporting Guidelines so that survey metrics are comparable [27].

The impact of mode on reporting sensitive behaviors in LMIC is discrepant [12]. A meta-analysis that compared 15 data sets (which included no forms of remote data collection), mostly comparing FTF and audio computer-assisted self-interview found that non-FTF methods did not consistently produce a significant increase in the reporting of 4 sensitive questions [12]. In this literature review, only 3 articles offered a straight-forward comparison of nondesirable behaviors. In 2 of the articles, remote data collection elicited higher report of nondesirable behaviors compared with in-person data collection [18,20]. The article that compared CATI, single-sitting and modular-design SMS found that the SMS respondents reported more socially undesirable behaviors compared to CATI [24].

XSL•FO
RenderX

Considering that all but one article in this literature review had been published in the past 5 years, we anticipate an increase in publications comparing modes of data collection in the coming years. As evidence continues to emerge, research designed to isolate the cause of differences in measures between modes should be a priority. For example, researchers should ask direct questions around the impact of increased privacy, greater anonymity, or greater convenience of remote data collection. Only 3 of the articles in this literature review included such questions [24,25,26]. By eliciting participant's opinions about the different modes, discrepancies can be better explained. Country context, such as literacy levels, mobile phone ownership, and network coverage are particularly important to note when considering remote data collection in LMICs. Therefore researchers should include aforementioned information in publications so that conditions can be considered. Furthermore, reporting factual, sensitive, or perceptual questions, as well listing the type of question (such as multiple choice, numeric, text) all provides pertinent information for decision making.

Although studying mode effect is an important aspect of remote data collection research, sampling is equally pertinent. RDD functions without a sampling frame in many countries, which means after identifying all mobile network operator prefixes in a country, numbers are randomly generated [28]. It is unknown whether RDD can consistently produce nationally representative estimates of a health outcome or which mode is best suited for RDD. Six of the 10 studies in this literature reviewed enrolled participants in-person. The advantage of FTF enrollment is that the research team has a reference standard against which the remote data collection tool is measured. When enrolling participants for FTF, asking how many mobile phone numbers each participant has and their estimated network coverage helps to estimate how representative the sample will be. Finding out the preferred language of survey while enrolling a participant will allow the first follow-up contact to be in the respondent's language of choice, negating the need for a language question and likely increasing the response rate. It is unknown whether RDD is more likely to enroll respondents who are hard to reach in FTF surveys, but this hypothesized advantage should be assessed. RDD and remote data collection generally have the advantage of faster collection of data than a FTF survey, thus making this approach particularly useful during a crisis.

To isolate a superior method of data collection, studies that compare more than two modes of remote data collection are preferred. Only three studies in this review compared more than two modes [14,18,24]. Specifically, future research should follow mHealth guidelines [29], incorporating a factorial design where possible. As a minimum, it would be advantageous for authors to identify which questions are sensitive in their context so that mode effects for sensitive questions can be compared, even if the subject matter is different.

Limitations

We note three main limitations to this literature review. First, the small number of studies (n=10) that compared two modes of data collection or more (n=25 comparisons), made it difficult to draw conclusions. The included research used a wide variety of data collection modes and sampling techniques and covered a plethora of topics and populations, thus negating the ability to make conclusions. Furthermore, the nonlinear relationship of effects can make pattern identification a challenge [11]. Second, owing to the inherent limitations of searching for grey literature, our search strategy could have missed important articles. A third limitation pertains to the inconsistent reporting of key survey metrics as well as lack of a formal statistical test to analyze the variation between the results of each article. Regardless of these limitations, this literature review contributes to efforts to characterize current evidence on the effect of remote data collection mode on data quality in LMICs.

Conclusions

Due to the nascent state of remote data collection in LMICs, several research areas merit further investigation. The advantages of remote data collection are presented, but a superior mode for a population has yet to be established due to a dearth of evidence. We encourage randomized control trials with multiple arms to identify a mode appropriate for the context. Ultimately, researchers must balance the desire for more efficient, cost-effective data collection methods with study aims and the limitations of a novel mode of data collection.

Acknowledgments

The studies described as part of the research agenda are funded by the Bloomberg Philanthropies. This funding agency had no role in the preparation of this manuscript. We would like to thank Victoria Gammino for a careful review of the manuscript.

Conflicts of Interest

None declared.

References

1. Okonjo-Iweala N, Osafo-Kwaako P. Improving health statistics in Africa. Lancet 2007 Nov 03;370(9598):1527-1528. [doi: 10.1016/S0140-6736(07)61644-4] [Medline: 17980722]
2. Hyder A, Wosu A, Gibson D, Labrique A, Ali J, Pariyo G. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246]
3. UN. Goal 3 Ensure healthy lives and promote well-being for all at all ages 2016 URL: http://www.un.org/sustainabledevelopment/health/ [accessed 2017-04-19] [WebCite Cache ID 6pqltShQG]

XSL•FO
RenderX

4. Poushter J, Oates R. Pew Research Center. 2015. Cell phones in Africa: communication lifeline URL: http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/ [accessed 2017-04-19] [WebCite Cache ID 6pqnC5jhS]
5. World Bank. Mobile cellular subscriptions (per 100 people) URL: http://data.worldbank.org/indicator/IT.CEL.SETS.P2 [accessed 2017-04-19] [WebCite Cache ID 6pqnLVdpq]
6. Gibson D, Pereira A, Farrenkopf B, Pariyo G, Labrique A, Hyder A. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. J Med Internet Res 2017;19(5):e139 [FREE Full text] [doi: 10.2196/jmir.7428]
7. Corkrey R, Parkinson L. Interactive voice response: review of studies 1989-2000. Behav Res Methods Instrum Comput 2002 Aug;34(3):342-353. [Medline: 12395550]
8. Kempf A, Remington P. New challenges for telephone survey research in the twenty-first century. Annu Rev Public Health 2007;28:113-126. [doi: 10.1146/annurev.publhealth.28.021406.144059] [Medline: 17094769]
9. Cannell C, Miller P, Oksenberg L. Research on interviewing techniques. Sociol. Methodol 1981;12:389-437. [doi: 10.2307/270748]
10. Groves R. Three eras of survey research. Public Opin Q 2011 Dec 15;75(5):861-871. [doi: 10.1093/poq/nfr057]
11. Jackle A, Roberts C, Lynn P. Assessing the effect of data collection mode on measurement. Int Stat Rev 2010;78(1):3-20. [doi: 10.1111/j.1751-5823.2010.00102.x]
12. Phillips A, Gomez G, Boily M, Garnett G. A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries. Int J Epidemiol 2010;39(6):1541-1555. [doi: 10.1093/ije/dyq114]
13. Tourangeau R, Smith T. Asking sensitive questions: the impact of data collection mode, question format, and question context. Public Opin Q 1996;60(2):275-304. [doi: 10.1086/297751]
14. Ballivian A, Azevedo J, Durbin W, Rios J, Godoy J, Borisova C. Using mobile phones for high-frequency data collection. In: Toninelli D, Pinter R, de Pedraza P, editors. Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies. London: Ubiquity Press; 2015:21-39.
15. Du X, Wang W, van Velthoven MH, Chen L, Scherpbier RW, Zhang Y, et al. mHealth series: text messaging data collection of infant and young child feeding practice in rural China - a feasibility study. J Glob Health 2013 Dec;3(2):020403 [FREE Full text] [doi: 10.7189/jogh.03.020403] [Medline: 24363921]
16. Ferreira A, César CC, Malta DC, Souza Andrade AC, Ramos CG, Proietti FA, et al. Validade de estimativas obtidas por inquérito telefônico: comparação entre VIGITEL 2008 e inquérito Saúde em Beagá. Rev bras epidemiol 2011 Sep;14(1):16-30. [doi: 10.1590/S1415-790X2011000500003]
17. Francisco P, Azevedo Barros MB, Segri N, Alves MC, Cesar C, Malta D. Comparação de estimativas para o auto-relato de condições crônicas entre inquérito domiciliar e telefônico - Campinas (SP), Brasil. Rev bras epidemiol 2011 Sep;14(3):5-15. [doi: 10.1590/S1415-790X2011000500002]
18. Lau JTF, Thomas J, Liu JLY. Mobile phone and interactive computer interviewing to measure HIV-related risk behaviors: the impacts of data collection methods on research results. AIDS 2000 Jun 16;14(9):1277-1280.
19. Li Y, Wang W, van Velthoven MH, Chen L, Car J, Rudan I, et al. Text messaging data collection for monitoring an infant feeding intervention program in rural China: feasibility study. J Med Internet Res 2013 Dec 04;15(12):e269 [FREE Full text] [doi: 10.2196/jmir.2906] [Medline: 24305514]
20. Mahfoud Z, Ghandour L, Ghandour B, Mokdad A, Sibai A. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? Field Methods 2014 Jul 01;27(1):39-54. [doi: 10.1177/1525822X14540084]
21. van Velthoven MH, Li Y, Wang W, Du X, Wu Q, Che N. mHealth project in Zhao County, rural China - Description of objectives, field site and methods. J Glob Health 2013;3(2):020401. [doi: 10.7189/jogh.03.020401]
22. Abbott SA, Friedland BA, Sarna A, Katzen LL, Rawiel U, Srikrishnan AK, et al. An evaluation of methods to improve the reporting of adherence in a placebo gel trial in Andhra Pradesh, India. AIDS Behav 2013 Jul;17(6):2222-2236 [FREE Full text] [doi: 10.1007/s10461-012-0402-z] [Medline: 23299877]
23. Moura E, Claro RM, Bernal R, Ribeiro J, Malta D, Morais Neto O. A feasibility study of cell phone and landline phone interviews for monitoring of risk and protection factors for chronic diseases in Brazil. Cad Saúde Pública 2011 Feb;27(2):277-286. [doi: 10.1590/S0102-311X2011000200009]
24. West BT, Ghimire D, Axinn WG. Evaluating a modular design approach to collecting survey data using text messages. Surv Res Methods 2015;9(2):111-123 [FREE Full text] [Medline: 26322137]
25. Mavletova A, Couper M. Sensitive topics in PC web and mobile web surveys: is there a difference? Surv Res Meth 2013;7(3):191-205. [doi: 10.18148/srm/2013.v7i3.5458]
26. Mavletova A. Data quality in PC and mobile web surveys. Soc Sci Comput Rev 2013;31(6):19.
27. The American Association for Public Opinion Research.: Washington, DC; 2016. Standard definitions: final dispositions of case codes and outcome rates for surveys URL: http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2017-04-19] [WebCite Cache ID 6pr64foLk]
28. Gibson D, Pariyo G, Wosu A, Greenleaf A, Ali J, Ahmed S, et al. Evaluation of mechanisms to improve performance of mobile phone surveys: a research protocol. JMIR Res Protoc 2017;6(5) [FREE Full text] [doi: 10.2196/resprot.7534]

XSL•FO
RenderX

29. Agarwal S, LeFevre A, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174]

Abbreviations

CATI: Computer-assisted telephone interview
FTF: face-to-face
HIV: Human Immunodeficiency Virus
IVR: interactive voice response
LMIC: low- and middle-income countries
PC: personal computer
RDD: random digit dialing
SAQ: self-administered questionnaire
SMS: short message service
SSA: Sub-Saharan Africa

Edited by E Rosskam, A Hyder; submitted 16.01.17; peer-reviewed by J de Jong, TR Soron; comments to author 28.02.17; revised version received 22.03.17; accepted 23.03.17; published 05.05.17.

Please cite as:
Greenleaf AR, Gibson DG, Khattar C, Labrique AB, Pariyo GW
Building the Evidence Base for Remote Data Collection in Low- and Middle-Income Countries: Comparing Reliability and Accuracy Across Survey Modalities
J Med Internet Res 2017;19(5):e140
URL: http://www.jmir.org/2017/5/e140/
doi:10.2196/jmir.7331
PMID:28476728

©Abigail R Greenleaf, Dustin G Gibson, Christelle Khattar, Alain B Labrique, George W Pariyo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Evaluation and Referral Assistant: A Randomized Controlled Trial of a Web-Based Screening, Brief Intervention, and Referral to Treatment System to Reduce Risky Alcohol Use Among Emergency Department Patients

Brianna L Haskins[1], MS; Rachel Davis-Martin[1], PhD; Beau Abar[2], PhD; Brigitte M Baumann[3], MSCE, MD; Tina Harralson[4], PhD; Edwin D Boudreaux[1], PhD

[1]University of Massachusetts Medical School, Worcester, MA, United States
[2]University of Rochester Medical Center, Rochester, NY, United States
[3]Cooper Medical School of Rowan University, Camden, NJ, United States
[4]Polaris Health Directions, Inc, Wayne, PA, United States

Corresponding Author:
Brianna L Haskins, MS
University of Massachusetts Medical School
55 Lake Avenue North
Worcester, MA, 01655
United States
Phone: 1 5084211400
Email: brianna.haskins@umassmed.edu

Abstract

Background: Computer technologies hold promise for implementing alcohol screening, brief intervention, and referral to treatment (SBIRT). Questions concerning the most effective and appropriate SBIRT model remain.

Objective: The aim of this study was to evaluate the impact of a computerized SBIRT system called the Health Evaluation and Referral Assistant (HERA) on risky alcohol use treatment initiation.

Methods: Alcohol users (N=319) presenting to an emergency department (ED) were considered for enrollment. Those enrolled (n=212) were randomly assigned to the HERA, to complete a patient-administered assessment using a tablet computer, or a minimal-treatment control, and were followed for 3 months. Analyses compared alcohol treatment provider contact, treatment initiation, treatment completion, and alcohol use across condition using univariate comparisons, generalized estimating equations (GEEs), and post hoc chi-square analyses.

Results: HERA participants (n=212; control=115; intervention=97) did not differ between conditions on initial contact with an alcohol treatment provider, treatment initiation, treatment completion, or change in risky alcohol use behavior. Subanalyses indicated that HERA participants, who accepted a faxed referral, were more likely to initiate contact with a treatment provider and initiate treatment for risky alcohol use, but were not more likely to continue engaging in treatment, or to complete treatment and change risky alcohol use behavior over the 3-month period following the ED visit.

Conclusions: The HERA promoted initial contact with an alcohol treatment provider and initiation of treatment for those who accepted the faxed referral, but it did not lead to reduced risky alcohol use behavior. Factors which may have limited the HERA's impact include lack of support for the intervention by clinical staff, the low intensity of the brief and stand-alone design of the intervention, and barriers related to patient follow-through, (eg, a lack of transportation or childcare, fees for services, or schedule conflicts).

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): NCT01153373; https://clinicaltrials.gov/ct2/show/NCT01153373 (Archived by WebCite at http://www.webcitation.org/6pHQEpuIF)

(J Med Internet Res 2017;19(5):e119) doi:10.2196/jmir.6812

KEYWORDS

alcohol consumption; intervention study; emergency medicine; referral and consultation

XSL•FO
RenderX

Introduction

Background

Between 2006 and 2010, excessive alcohol consumption was responsible for 88,000 deaths and an estimated 2.5 million years of potential life lost each year. Risky alcohol use is among the leading preventable causes of death in the United States [1], but it remains highly prevalent and poorly intervened among the emergency department (ED) population. Prevalence rates of risky alcohol use among ED patients exceed the national average, making EDs an ideal location for innovative alcohol intervention efforts [2-5]. Over 130 million patients visit the ED each year [6], with a large percentage of these patients having unrecognized alcohol-related treatment needs (eg, risky drinking, problem drinking, and alcohol dependence). Patients with untreated needs are more likely to be admitted to the hospital and repeatedly rely on ED services [2,7-9]. Furthermore, disadvantaged populations including minorities, immigrants, people without insurance, and low-income households comprise the underserved populations that currently rely disproportionately on EDs for primary care [10] and suffer from alcohol-related concerns at elevated rates [11-12]. ED-originated alcohol intervention efforts have the potential for considerable impact on public health by promoting change in both high-risk and hard-to-reach populations [4,10,13].

This potential has been acknowledged by the latest health care legislation and numerous health care agencies. The Affordable Care Act includes strong incentives for the integration of behavioral health and medical treatment [2,14]. Numerous studies have demonstrated the effectiveness of screening and brief intervention (SBI) programs as well as screening, brief intervention, and referral to treatment (SBIRT) programs aimed at addressing alcohol use problems and treatment needs among ED patients [2,15-18]. As a result, the US Preventive Services Task Force and the Substance Abuse and Mental Health Services Administration have recommended universal SBIRT for alcohol and other substances in general medical settings, including EDs [19,20]. The Centers for Medicare and Medicaid Services and the American Medical Association have authorized billing codes to reimburse SBIRT services for alcohol, tobacco, and illicit drug use [21], and the Centers for Disease Control and Prevention have called for increased alcohol SBI, including systems-level changes to include integration into the electronic health record system [22].

Despite the support of numerous studies and many health agencies [2,14-23], ED-originated alcohol screening rates remain low, with many hospitals only screening alcohol toxicology reports, which do not assess problem-level severity [2,24,25]. Factors likely contributing to these low rates include lack of specialized behavioral health training, competing demands on time and resources inherent to the ED setting, and a primary objective of acute medical care rather than treatment for chronic conditions [5,13,23]. The use of convenient and brief procedures requiring minimal specialized training and focusing on connecting patients with outpatient resources for continued treatment after their ED visit could maximize successful implementation of SBIRT. Behavioral intervention technology advancement (eg, computerized assessments, personalized feedback reports, faxed referrals, electronic health records), holds promise for facilitating the implementation of SBIRT in EDs and a variety of health care settings [26-33]. Such advancements have also allowed for the development of computer-assisted SBIRT models designed to diminish interruptions in clinical care and mitigate clinician burden without sacrificing effectiveness [13,23].

Objectives

Questions remain concerning the most effective and appropriate SBIRT model. The objective of this study was to assess an innovative Web-based program's ability to facilitate alcohol SBIRT. The Health Evaluation and Referral Assistant (HERA) is patient-administered on a tablet computer during the ED visit and is modeled after the face-to-face SBIRT screening approach. This study hypothesized that the HERA would improve initiation of specialized outpatient treatment for risky alcohol use and reduce risky alcohol use among ED patients at 3 months postvisit as compared with a minimal intervention control condition.

Methods

Previous Reporting

A complete description of the HERA development and randomized controlled trial (RCT) methods were previously published [26,34]. Although the HERA assesses and refers patients to treatment for multiple substances, only results pertaining to alcohol are reported and discussed in this paper. A previous publication reported results for tobacco use [23], and a subsequent paper will address the results pertaining to illicit drug use. This clinical trial was registered with ClinicalTrials.gov as the Dynamic Assessment and Referral System.

Health Evaluation and Referral Assistant (HERA)

Assessment

The HERA is a self-administered patient assessment completed on a tablet computer during the ED visit. The assessment was designed to require no computer literacy beyond the ability to read at the 8th grade reading level and respond to questions using a numeric keypad or stylus. The HERA used the Alcohol Use Disorders Identification Test (AUDIT) to assess alcohol use behaviors [35]. The version was based on the Cutting Back study [36], which adjusted the responses to the first three items of the AUDIT to reflect US alcohol content standards. This version has ultimately become what is referred to as the USAUDIT and has been adopted by the Centers for Disease Control and Prevention [37].

Readiness to change was assessed with an initial question that asked, "Would you like to change your alcohol use? No; Undecided; Yes, I would like to CUT BACK; Yes, I would like to QUIT COMPLETELY." If interested in quitting, the participant was asked, "When would you like to quit? Within the next 30 days; Within the next 6 months; More than 6 months from now." Treatment history was assessed by asking, "Have you ever been in treatment for alcohol use? No; Yes, but I AM

XSL•FO
RenderX

NOT CURRENTLY in treatment; Yes, and I AM CURRENTLY in treatment." Readiness to enter treatment was assessed for those who scored in the risky alcohol use range, were not currently in treatment, and reported interest in changing alcohol use by asking, "You have reported that you are interested in changing your alcohol use. This computer program can help you connect with a counselor or treatment program. Would you like some help with finding a counselor or treatment program? Yes; No." Withdrawal symptoms were assessed using a checklist of items: "Please check all of the withdrawal symptoms you had in the past 30 days, including today: seizures or convulsions; hallucinations (saw, heard, or felt something that was not there); confusion or disorientation; paranoid thinking; severe depression; severe loss of energy (lethargy); none of the above." The Patient Health Questionaire-2 (PHQ-2) [38] was used to screen participants for depression, and a complete psychiatric history was documented using a checklist of common psychiatric diagnoses (eg, anxiety, post-traumatic stress disorder, bipolar disorder, schizophrenia, attention deficit hyperactivity disorder, and so on). See Multimedia Appendix 1 for sample assessment screenshots.

Report Generator

The assessment data were used to automatically produce two reports at the end of the computerized assessment, which are described in detail in the aforementioned manuscripts [26,34]. The health care provider report provided a summary of the assessment and was given to the patient's treating physician for review. The patient feedback report was given to the patient and consisted of 3 sections: (1) the Face Sheet, which included an overview and tailored alcohol use treatment referral list; (2) the Patient Assessment Summary, which provided individually tailored feedback pertaining to the patient's alcohol use; and (3) the Motivation Toolkit, which provided several worksheets based on Motivational Interviewing [39] and the Transtheoretical Model [40]. See Multimedia Appendix 2 for a sample patient feedback report.

Referral Generator

The referral generator utilized a library of alcohol use treatment services maintained by Polaris Health Directions, Inc. to create individually tailored referral lists and to send dynamic referrals. Referral lists contained free and fee-for-service treatment options, and dynamic referrals were based on a "best match" facility dependent on patient characteristics, such as the individual's ZIP code, insurance provider, and preference for telephone or in-person treatment. If accepted by the patient, the dynamic referral was faxed by the HERA to a matched treatment facility, along with a brief assessment summary and the patient's contact information. The participating services had agreed to contact the patient within 48 h of receiving the referral to complete an initial evaluation and discuss treatment options.

Procedure

Patients were enrolled from 4 EDs (see Table 1) between 8 am and 7 pm, with shifts occurring every day of the week. Research assistants (RAs) approached all adult patients at their bedside during their ED visit. Patients aged 18 years and older with risky alcohol use were considered. Risky alcohol users were defined as having used alcohol above the AUDIT quantity or frequency guidelines, with or without tobacco use but with no illicit drug use in the past 12 months. This paper focuses only on alcohol users who may have been smokers but did not use illicit drugs. Exclusion criteria were severe illness or distress, cognitive insufficiency, in state custody or restraints, being held involuntarily, and language barriers. Patients who were actively involved in alcohol treatment were eligible for the study, but few agreed to participate. Patients were enrolled regardless of whether they were admitted or discharged. The study components, including baseline and intervention, were completed while patients were in the ED. Participants were randomized to either the intervention or control condition by a random number generator from the Java programming language standard library embedded within the HERA. Immediately after discharge or transfer from the ED, the enrolling RA completed a brief interview with the participant, either in person or by telephone within 48 h (postvisit interview). A trained RA not affiliated with the data collection sites contacted all participants by telephone at 1 and 3 months following the ED visit to assess alcohol treatment initiation and to reassess alcohol use. This study was approved by the Institutional Review Boards for all data collection sites, in accordance with the ethical standards of the Helsinki Declaration of 1975. All participants gave their informed consent and signed a written consent form before inclusion in the study.

Table 1. Site characteristics. This table was previously published with the reporting of the tobacco results [23].

Type	Annual volume	Location	Race or ethnicity
Academic, urban	90,733	Worcester, MA[a]	W[b] 82%, H[c] 11%, B[d] 4%
Community, urban	47,364	Worcester, MA	W 74%, H 14%, B 9%
Community, suburban	23,217	Marlboro, MA	W 80%, H 15%, B 3%, U[e] 2%
Academic, urban	59,482	Camden, NJ[f]	W 35%, H 20%, B 45%

[a]MA: Massachusetts.

[b]W: white, non-Hispanic.

[c]H: Hispanic.

[d]B: black.

[e]U: unknown.

[f]NJ: New Jersey.

XSL•FO
RenderX

Study Conditions

Intervention and control conditions were treated the same in all aspects of the study procedures; however, the groups differed on the type of referral and availability of reports. Participants in the intervention condition (HERA) (1) were offered a dynamic referral, (2) received the patient feedback report with a tailored referral list, and (3) their treating physician received the health care provider report. Participants assigned to the minimal intervention control condition (control) were given a standardized, printed list of local treatment providers instead of dynamic referrals, and health care provider reports were not made available.

Blinding

The RA who performed the outcome assessments was partially blinded. Because the HERA is heavily focused on the referral process, and not all patients received the same type of referrals, to avoid confusion, the follow-up questions were tailored to the referral type received at baseline (printed list vs dynamic referral). Despite blinding efforts, the presence of particular questions for the intervention group revealed some information about group assignment. For example, only patients who chose a dynamic referral were asked whether they had been contacted by an alcohol treatment provider.

Measures

Health Evaluation and Referral Assistant (HERA)

The HERA assessment was previously described under *Methods* or *Assessment*.

Postvisit Interview

Immediately after patients were discharged or transferred from the ED, the enrolling RA completed a brief interview to establish whether the treating clinicians provided alcohol treatment counseling, education materials, or referrals for alcohol use treatment. Chart review was not used because of unreliability associated with documentation.

Follow-Up Assessment

All participants were phoned by an RA and asked if they had initiated contact with an alcohol treatment provider or program (treatment contact); completed an initial assessment (treatment initiation); attended any additional treatment sessions beyond the initial assessment (treatment engagement); and completed treatment (treatment completion). Participation in self-help groups, like Alcoholics Anonymous, was also assessed. Additionally, the RA assessed self-reported current alcohol use using the first three items from the USAUDIT (frequency of drinking, amount on a typical day, frequency consuming four or more drinks on a single occasion). This was used to quantify use and to determine abstinence, which was defined as 0 drinks since the ED visit. Efforts to decrease use were assessed with the following questions: “In the past ‘x’ months, have you tried to reduce your alcohol use? Yes; No. In the past ‘x’ months, have you intentionally gone for more than 24 h without having a drink? Yes; No. In the past ‘x’ months, how many days have you gone without having a drink?”

Data Analyses

Baseline characteristics (eg, demographics, alcohol use) were compared across intervention conditions using chi-square test of independence and independent samples *t* test to confirm randomization success, and the potential for differential retention rates across conditions was examined using chi-square test of independence. Our primary outcomes (ie, alcohol treatment provider contact, treatment initiation, alcohol use) were then compared across conditions at 1 and 3 months using generalized estimating equation (GEE) models. Post hoc chi-square test of independence was performed following a statistically significant GEE model in order to better isolate the observed differences at each follow-up point. Chi-square analyses were also used to make comparisons across conditions for outcomes collected only at a single follow-up point (eg, ED counseling).

We then performed a series of analyses comparing participants in 3 distinct groups: (1) the control condition, (2) the intervention condition that declined a dynamic referral to providers (tailored list only), and (3) the intervention condition that accepted a dynamic referral (dynamic referral group). Because this categorization allows for preexisting differences across groups (particularly between the tailored list and dynamic referral groups), these models included theoretically relevant covariates that might impact the outcomes of interest (baseline AUDIT scores and readiness to quit). Missing data or attrition at follow-up was addressed using standard intention-to-treat principles whereby the least favorable outcome (eg, no provider contact, no treatment completion) was assigned to missing data points. Specifically, if data were missing at both follow-up points for a case, the least favorable outcome was imputed. If data from the first follow-up indicated a favorable outcome (eg, quit attempt, initiated treatment) and data was missing at the second follow-up, a favorable outcome would be imputed as we were interested in the event occurring by a given time point. If data were missing at the first follow-up and present at the second follow-up, regardless of the outcome at the second follow-up, the least favorable outcome would be imputed at the first follow-up. Given the use of these principles, the frequencies presented in each table represent observed data, whereas the percentages represent intention-to-treat estimates. All analyses were performed using Statistical Package for the Social Science 22 (IBM, 2012), with an a priori alpha level of .05.

Results

Preliminary Analysis

Of 319 alcohol users who met eligibility criteria and did not report any drug use, 212 individuals were enrolled (see Multimedia Appendix 3). A greater proportion of eligible females (75.7%, 78/103) enrolled in the study than males (62.0%, 134/216), χ^2_1=5.9 *P*=.02, and enrolled individuals were younger, on average (mean 38.1 years; SD 13.4) than nonenrolled individuals (mean 42.3 years; SD 13.9), t_{317}=2.63, *P*=.009. There were no differences in percentage of enrolled eligible patients across sites, concomitant tobacco use, and insurance status. Of the 212 participants enrolled, 115 were assigned to the control condition and 97 to the intervention condition. At baseline, there were no differences between the

XSL•FO
RenderX

two conditions on demographics (see Multimedia Appendix 4), data collection site (*P*=.06), mental health diagnoses (*P*=.19 to .83), AUDIT scores (*P*=.24), or readiness to change (*P*=.10).

Of the analyzed participants, 196 out of 212 (92.5%) completed the postvisit interview, 157 out of 212 (74.1%) completed the 1-month follow-up, and 157 out of 212 (74.1%) completed the 3-month follow-up (see Multimedia Appendix 3). There were no differences between retained individuals and those lost to follow-up on age (*P*=.99; .17, respectively), baseline AUDIT scores (*P*=.62; .34), mental health diagnoses (*P*=.27 to .81; .09 to .74), or readiness to change (*P*=.58; .21). However, at the 3-month follow-up, there were more control individuals retained (95/115; 83.0%) than experimental (62/97; 64%), χ^2_1=9.6, *P*=.002. There were also more female retained at the 3-month follow-up (64/78; 82%) than male (93/134; 69.4%), χ^2_1=4.1, *P*=.04.

Comparisons on Outcomes of Interest

Specialized Alcohol Use Treatment

There were no differences in initial contact between participants and alcohol use treatment provider across conditions (odds ratio, OR 1.04; 95% CI 0.45-2.40; see Table 2). No differences were observed on treatment initiation (*P*=.53), treatment engagement (*P*=.21), and treatment completion rates either (*P*=.31). Among the participants in the HERA, 14/97 (14%) accepted a dynamic referral.

Alcohol Use

Sustained abstinence at both follow-up periods was not statistically different across intervention and control conditions (see Table 2). Quit attempts and efforts to reduce alcohol use were more common among control participants than experimental (OR 0.44, 95% CI 0.26-0.77, *P*=.004; OR 0.66, 95% CI 0.51-0.87, *P*=.01, respectively).

XSL•FO
RenderX

Table 2. Comparisons between alcohol intervention and control conditions.[a]

Characteristics	Intervention (n=97), n (%)	Control (n=115), n (%)
ED[b] clinician (MD[c] or RN[d]) counseling[e]		
MD or RN asked about alcohol use	62 (64)	80 (69.6)
MD or RN counseled participant to quit	11 (11)	12 (10.4)
Received educational materials	2 (2)	4 (3.5)
Received an alcohol abuse referral	1 (1)	4 (3.7)
Outpatient alcohol abuse treatment		
Contact with alcohol abuse treatment provider		
GEE[f] odds ratio 1.04 (95% CI 0.45-2.40), *P*=.94		
Contact at 1 month	7 (7)	10 (8.7)
Contact at 3 months	13 (13)	13 (11.3)
Initiated treatment (evaluated by alcohol abuse treatment provider)		
GEE odds ratio 0.70 (95% CI 0.23-2.15), *P*=.53		
Treatment initiation at 1 month	3 (3)	7 (6.1)
Treatment initiation at 3 months	6 (6)	8 (7.0)
Treatment engagement at either time	3 (3)	8 (7.0)
Treatment completion	3 (3)	7 (6.1)
Alcohol use behavior		
Used alcohol (since ED visit)		
GEE odds ratio 0.80 (95% CI 0.30-2.14), *P*=.66		
Abstinent for 1st month (since visit)	8 (8)	12 (10.4)
Abstinent for 3 months (since visit)	3 (3)	4 (3.5)
At least one quit attempt at 1 month	17 (18)	37 (32.2)
At least one quit attempt at 3 months	30 (30)	58 (50.4)
Attempted to reduce use at 1 month	25 (26)	45 (39.1)
Attempted to reduce use at 3 months	33 (34)	57 (49.6)

[a]All percentages and analyses use the intention-to-treat principle of worst outcome for missing values.

[b]ED: emergency department.

[c]MD: doctor of medicine.

[d]RN: registered nurse.

[e]ED clinician behavior assessment included behaviors over and above the materials provided as part of the research study. All patients in both groups had alcohol assessed as part of the study and received a referral list. The control group received a preprinted list, whereas the intervention group received a personally tailored list, as well as a dynamic referral if desired.

[f]GEE: generalized estimating equation.

Physician Behavior

Clinician counseling, provision of educational materials, and provision of referrals, beyond those provided as part of the study protocol, were not statistically different across intervention and control conditions (see Table 2).

Exploring the Effect of Dynamic Referrals

Supplemental GEE analyses demonstrated large differences across groups on treatment contact. Using dummy codes (control condition as the reference), results indicated that experimental participants who accepted a dynamic referral contacted a provider at a much greater rate than control individuals (OR 7.14, 95% CI 2.33-20.41, *P*<.001; see Table3). Effects on treatment initiation were also significant, with higher rates of initiation among experimental participants who accepted a dynamic referral and control participants (OR 3.92, 95% CI 1.01-15.15, *P*=.05). There were no differences in treatment initiation between experimental individuals who did not accept a dynamic referral and control individuals (OR 0.26, 95% CI 0.05-1.36, *P*=.11). The difference in contact with providers between experimental participants who accepted a dynamic referral and control participants remained significant (*P*=.001) when accounting for baseline readiness to change and baseline

XSL•FO
RenderX

AUDIT scores in a post hoc GEE model. The effect of a dynamic referral on treatment initiation was no longer significant in a similar model.

There was a marginally significant effect of group membership on engagement in alcohol treatment, χ^2_2=5.8, *P*=.06 (see Table 3). Although engagement was quite infrequent across all groups, the rate of engagement for those accepting a dynamic referral was more than double the rate observed in the control condition (2/14, 14% vs 8/115, 7.0%; see Table 3). This effect was no longer significant when baseline readiness to change and AUDIT scores were included as covariates (OR control vs tailored list only=0.23, 95% CI 0.02-2.13, *P*=.20; OR control vs dynamic referral=2.14, 95% CI 0.26-17.54, *P*=.48). There were no effects of group membership on alcohol reduction (see Table 3).

Table 3. Comparisons across alcohol intervention, tailored list only; alcohol intervention, dynamic referral; and control conditions.

Characteristic	Intervention-provider list (n=83), n (%)	Intervention-dynamic referral (n=14), n (%)	Control (n=115), n (%)
Outpatient alcohol abuse treatment			
Contact with alcohol abuse treatment provider			
Contact at 1 month	3 (4)	4 (29)	10 (8.7)
Contact at 3 months	5 (6)	8 (57)	13 (11.3)
Initiated treatment (evaluated by alcohol abuse treatment provider)			
Treatment initiation at 1 month	1 (1)	2 (14)	7 (6.1)
Treatment initiation at 3 months	2 (2)	4 (29)	8 (7.0)
Treatment engagement, either time	1 (1)	2 (14)	8 (7.0)
Treatment completion	1 (1)	2 (14)	7 (6.1)
Alcohol use behavior			
Used alcohol (since ED[a] visit)			
Abstinent for first month (since visit)	8 (10)	0 (0)	12 (10.4)
Abstinent for 3 months (since visit)	3 (4)	0 (0)	4 (3.5)
At least one quit attempt at 1 month	15 (18)	2 (14)	37 (32.2)
At least one quit attempt at 3 months	27 (33)	3 (21)	58 (50.4)
Attempted to reduce use at 1 month	21 (25)	4 (29)	45 (39.1)
Attempted to reduce use at 3 months	29 (35)	4 (29)	57 (49.6)

[a]ED: emergency department.

Discussion

Principal Findings

ED-originated alcohol interventions have potential for substantial public health impact by offering widespread SBIRT for risky alcohol use within a population that is both high risk and difficult to reach [4,10,13]. However, many challenges continue to impede the adoption of interventions into routine clinical care, including competing time demands and priorities, a focus on acute care, and insufficient specialized training of providers in risky alcohol use interventions. Technology facilitated intervention models that maximize efficiency and relieve clinician burden may offer a solution.

The results of this clinical trial exploring the benefits of using a single administration, stand-alone computerized intervention were mixed. All participants scored positive for risky alcohol use, and therefore received a patient feedback report with personalized information and referrals. Those who reported not currently being in treatment and who reported some desire to change their drinking were offered a dynamic referral. Although overall no significant differences were observed between conditions for contact with a treatment provider, treatment initiation, treatment engagement, and treatment completion, a closer look at the data suggests that the dynamic referral may still hold promise for promoting treatment engagement. Subanalyses revealed that among the experimental participants, those who accepted a dynamic referral were more likely to make contact with a treatment provider and have higher rates of treatment initiation than control participants. However, these effects did not lead to continued engagement in treatment or changes in alcohol use over the 3-month period following the ED visit. Moreover, some of the trends for attempting to change, such as reporting any attempt to reduce use, favored the control condition, rather than the intervention condition, though these differences were not statistically different. Additional research is needed to probe this pattern to establish if there may be an iatrogenic impact of providing personalized information and referrals in dampening self-change.

There are several factors that may have hampered the HERA's impact on treatment and alcohol use behavior. One factor may be a lack of adoption and implementation by ED clinical staff. Although the clinical staff members who received the health care provider reports were trained to interpret the findings, they

XSL•FO
RenderX

were not specifically trained or mandated to provide counseling or additional intervention materials to patients as a result of reviewing the report. Analyses indicated that clinical staff did not provide additional counseling or intervention materials to participants in the intervention group, which could be interpreted as weak clinician adoption or support of the intervention. Although the HERA is designed to offer brief intervention and referral to treatment as a stand-alone service, a cooperative approach which includes protocols for clinician involvement in response to a positive screen on the health care provider report may prove a stronger intervention than a stand-alone automated referral. Furthermore, the sample was heterogeneous, with only a minority scoring in the severe range on the AUDIT (low to moderate risk, n=173/212 [82%], moderate to high risk, n=13/212 [6.1%], and high to very high risk, n=26/212 [12.3%]). This undoubtedly dampens the level of interest in specialized treatment.

An additional factor limiting clinical impact could be the low-intensity nature of the HERA intervention. The HERA was designed as a one-time, brief interaction due to the fast-paced ED environment filled with competing demands for time and resources. Minimizing the intervention for this purpose could have adversely affected the HERA's potential for clinical impact. The brief encounter with the HERA, while efficient and time-saving for clinicians, may not be powerful enough to support long-term changes in alcohol use behavior. Future technology-facilitated interventions may need to integrate motivational tools for behavior change, such as Web-based multimedia content or longitudinal interaction beyond the ED visit.

A final factor that may have impeded continued treatment and change in risky alcohol use behavior are barriers related to patient follow-through, including a lack of transportation or childcare, fees for services, and schedule conflicts. Although the dynamic referral was designed to connect patients with a "best match" treatment facility based on personal characteristics, the scope is limited to general characteristics such as location, insurance provider, and desire for telephone or in-person treatment. Motivated patients, who initiated contact with a nearby treatment provider compatible with their insurance carrier, may still have been unable to attend treatment due to the aforementioned circumstances [41]. Additionally, individuals were offered free alternatives to fee-for-service treatment models to help address cost barriers, but free treatment providers may not have been conveniently located or available during hours conducive to every patient's schedule.

Limitations

Several limitations exist that impact interpreting the results. First, because a minimal treatment control group was used, rather than true treatment as usual, the assessment and resource list provided to the minimal treatment control group may have had an intervention effect and artificially inflated treatment contact and behavior change in the control group. Second, the use of a modified AUDIT allowed for time-sensitive brief assessment of alcohol use, but assessed use over a shorter period than other methods, such as the Timeline Follow Back [42], and may not have allowed for a large enough assessment window to detect risky alcohol use in some individuals. Third, results may have been skewed by the nature of self-report due to the ambiguity inherent in measuring alcohol consumption by number of drinks. Although standard measurements have been outlined by the AUDIT [35], and clear examples of "one alcoholic drink" were provided during the assessment, participant understanding of the size and volume of "one alcoholic drink" varies considerably. Finally, by focusing solely on alcohol users, who used alcohol above the AUDIT quantity or frequency guidelines and who had not used illicit drug in the past 12 months, the generalizability of the results is limited. Future research should examine the efficacy of automated referral systems for alcohol treatment among all alcohol users.

The fact that very few participants accepted the dynamic referral highlights a potential limitation of the HERA model itself. Although participants who accepted a dynamic referral were more likely to contact a treatment provider and demonstrated higher rates of treatment initiation than control participants, impact will be minimal unless more patients begin accepting the referral. Future studies of similar models should aim to identify and overcome barriers to referral acceptance. A final limitation is that participants who failed to follow-through with treatment after receiving the referral were not questioned as to what factors contributed to their failure to follow-through. Costs associated with fee-for-service treatment options may have been a barrier to treatment initiation and engagement, although potentially alleviated by the inclusion of free treatment options in addition to the fee-for-service selections. Barriers to patient follow-through in systems like the HERA should be explored in future studies.

Conclusions

The HERA aims to satisfy clinical practice mandates for SBIRT for risky alcohol users in the ED setting. For those who accepted the dynamic referral, the HERA was effective at promoting contact with an alcohol treatment provider and initiating risky alcohol use treatment. Unfortunately, when employed as a stand-alone intervention, the HERA did not lead to sustained treatment engagement or changes in alcohol use during the 3 months following the initial ED visit. These results raise two questions: (1) Do stand-alone, brief, automated interventions lack the power to sufficiently motivate sustained alcohol use treatment engagement and behavior change? and (2) Is SBIRT for risky alcohol use satisfactory for all populations, particularly those unable or unwilling to pay fees associated with treatment services or underserved populations with limited access to health care, as represented in this study? This study highlights the need for developing and studying interventions that work alongside alcohol treatment linkage strategies. The prototype of the HERA was called the Dynamic Assessment and Referral System for Substance Abuse (DARSSA). The name was changed to reflect our long-term plans to expand the system to provide SBIRT for other nonsubstance problems, like depression and interpersonal violence.

XSL•FO
RenderX

Acknowledgments

This study was funded by a Small Business Technology Transfer grant from the National Institutes of Health (R42DA021455) to Polaris Health Directions, Inc.

Authors' Contributions

Brianna L. Haskins, Rachel Davis-Martin, and Tina Harralson assisted with manuscript preparation. Beau Abar completed the data analyses, assisted in data interpretation, and assisted with manuscript preparation. Brigitte M. Baumann assisted with study design, study completion, and manuscript preparation. Edwin D. Boudreaux participated in the study design, oversaw study completion, assisted with data interpretation, and oversaw manuscript preparation.

Conflicts of Interest

An agreement related to technology used in this study exists between the University of Massachusetts Medical School and Polaris Health Directions. Dr Boudreaux is an employee of the University of Massachusetts Medical School and receives consulting income from Polaris Health Directions. In addition, if the aforementioned technology should be licensed and result in licensing-related income, Dr Boudreaux would receive a share under the University's allocation policy to inventors. Dr Harralson is an employee of Polaris Health Directions. Dr Abar, Dr Baumann, Dr Davis, and Ms Haskins have no conflicts to disclose.

Multimedia Appendix 1

Sample assessment screenshots.

[PDF File (Adobe PDF File), 1MB - jmir_v19i5e119_app1.pdf]

Multimedia Appendix 2

Sample patient feedback report.

[PDF File (Adobe PDF File), 406KB - jmir_v19i5e119_app2.pdf]

Multimedia Appendix 3

Total Health evaluation and referral assistant (HERA) potential participants.

[PDF File (Adobe PDF File), 149KB - jmir_v19i5e119_app3.pdf]

Multimedia Appendix 4

Demographic characteristics of the analyzed sample.

[PDF File (Adobe PDF File), 396KB - jmir_v19i5e119_app4.pdf]

Multimedia Appendix 5

CONSORT eHealth checklist.

[PDF File (Adobe PDF File), 459KB - jmir_v19i5e119_app5.pdf]

Multimedia Appendix 6

CONSORT-EHEALTH checklist V1.6.2.

[PDF File (Adobe PDF File), 446KB - jmir_v19i5e119_app6.pdf]

References

1. Gonzales K, Roeber J, Kanny D, Tran A, Saiki C, Johnson H, Centers for Disease Control and Prevention (CDC). Alcohol-attributable deaths and years of potential life lost--11 States, 2006-2010. MMWR Morb Mortal Wkly Rep 2014 Mar 14;63(10):213-216 [FREE Full text] [Medline: 24622285]
2. Sanjuan PM, Rice SL, Witkiewitz K, Mandler RN, Crandall C, Bogenschutz MP. Alcohol, tobacco, and drug use among emergency department patients. Drug Alcohol Depend 2014 May 1;138:32-38 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.01.025] [Medline: 24594289]
3. Centers for Disease Control and Prevention (CDC). Vital signs: binge drinking prevalence, frequency, and intensity among adults - United States, 2010. MMWR Morb Mortal Wkly Rep 2012 Jan 13;61(1):14-19 [FREE Full text] [Medline: 22237031]

XSL•FO
RenderX

4. Cherpitel CJ, Ye Y. Drug use and problem drinking associated with primary care and emergency room utilization in the US general population: data from the 2005 national alcohol survey. Drug Alcohol Depend 2008 Oct 1;97(3):226-230 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.03.033] [Medline: 18499355]
5. Cunningham RM, Bernstein SL, Walton M, Broderick K, Vaca FE, Woolard R, et al. Alcohol, tobacco, and other drugs: future directions for screening and intervention in the emergency department. Acad Emerg Med 2009 Nov;16(11):1078-1088 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00552.x] [Medline: 20053226]
6. Centers for Disease Control and Prevention (CDC). CDC. National Hospital Ambulatory Medical Care Survey URL: https://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf [WebCite Cache ID 6lEUp4rpq]
7. Hankin A, Daugherty M, Bethea A, Haley L. The Emergency Department as a prevention site: a demographic analysis of substance use among ED patients. Drug Alcohol Depend 2013 Jun 1;130(1-3):230-233. [doi: 10.1016/j.drugalcdep.2012.10.027] [Medline: 23253936]
8. Rockett IR, Putnam SL, Jia H, Chang CF, Smith GS. Unmet substance abuse treatment need, health services utilization, and cost: a population-based emergency department study. Ann Emerg Med 2005 Feb;45(2):118-127. [doi: 10.1016/j.annemergmed.2004.08.003] [Medline: 15671966]
9. Rockett IR, Putnam SL, Jia H, Smith GS. Assessing substance abuse treatment need: a statewide hospital emergency department study. Ann Emerg Med 2003 Jun;41(6):802-813. [doi: 10.1067/mem.2003.189] [Medline: 12764335]
10. Lowenstein SR, Koziol-McLain J, Thompson M, Bernstein E, Greenberg K, Gerson LW, et al. Behavioral risk factors in emergency department patients: a multisite survey. Acad Emerg Med 1998 Aug;5(8):781-787 [FREE Full text] [Medline: 9715239]
11. Keyes KM, Liu XC, Cerda M. The role of race/ethnicity in alcohol-attributable injury in the United States. Epidemiol Rev 2012;34:89-102 [FREE Full text] [doi: 10.1093/epirev/mxr018] [Medline: 21930592]
12. Williams EC, Gupta S, Rubinsky AD, Jones-Webb R, Bensley KM, Young JP, et al. Racial/ethnic differences in the prevalence of clinically recognized alcohol use disorders among patients from the U.S. Veterans Health Administration. Alcohol Clin Exp Res 2016 Feb;40(2):359-366. [doi: 10.1111/acer.12950] [Medline: 26842254]
13. Murphy MK, Bijur PE, Rosenbloom D, Bernstein SL, Gallagher EJ. Feasibility of a computer-assisted alcohol SBIRT program in an urban emergency department: patient and research staff perspectives. Addict Sci Clin Pract 2013;8:2. [doi: 10.1186/1940-0640-8-2] [Medline: 23324597]
14. Substance AbuseMental Health Services Administration, Department of HealthHuman Services. SAMHSA. 2013. Innovations in addiction treatment: addiction treatment providers working with integrated primary care services URL: http://www.integration.samhsa.gov/clinical-practice/13_May_CIHS_Innovations.pdf [accessed 2016-10-13] [WebCite Cache ID 6lEX58LyU]
15. Field CA, Baird J, Saitz R, Caetano R, Monti PM. The mixed evidence for brief intervention in emergency departments, trauma care centers, and inpatient hospital settings: what should we do? Alcohol Clin Exp Res 2010 Dec;34(12):2004-2010 [FREE Full text] [doi: 10.1111/j.1530-0277.2010.01297.x] [Medline: 20860610]
16. Babor TF, McRee BG, Kassebaum PA, Grimaldi PL, Ahmed K, Bray J. Screening, brief intervention, and referral to treatment (SBIRT): toward a public health approach to the management of substance abuse. Subst Abus 2007;28(3):7-30. [doi: 10.1300/J465v28n03_03] [Medline: 18077300]
17. Zarkin GA, Bray JW, Davis KL, Babor TF, Higgins-Biddle JC. The costs of screening and brief intervention for risky alcohol use. J Stud Alcohol 2003 Nov;64(6):849-857. [Medline: 14743949]
18. Academic ED SBIRT Research Collaborative. The impact of screening, brief intervention, and referral for treatment on emergency department patients' alcohol use. Ann Emerg Med 2007 Dec;50(6):699-710, 710.e1. [doi: 10.1016/j.annemergmed.2007.06.486] [Medline: 17870206]
19. U.S. Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. Ann Intern Med 2004 Apr 06;140(7):554-556. [Medline: 15068984]
20. Substance Abuse and Mental Health Services Administration, Department of Health and Human Services. SAMHSA. 2006. Screening: adds prevention to treatment URL: http://archive.samhsa.gov/SAMHSA_News/VolumeXIV_1/JanFeb2006.pdf [accessed 2016-10-13] [WebCite Cache ID 6lEYaMkHc]
21. Substance Abuse and Mental Health Services Administration, Department of Health and Human Services. SAMHSA. Rockville, MD: Substance Abuse and Mental Health Services Administration, Department of Health and Human Services; 2012. Coding for SBI reimbursement URL: http://www.integration.samhsa.gov/sbirt/reimbursement_for_sbirt.pdf [accessed 2016-10-13] [WebCite Cache ID 6lEd2xJIC]
22. McKnight-Eily LR, Liu Y, Brewer RD, Kanny D, Lu H, Denny CH, Centers for Disease Control and Prevention (CDC). Vital signs: communication between health professionals and their patients about alcohol use--44 states and the District of Columbia, 2011. MMWR Morb Mortal Wkly Rep 2014 Jan 10;63(1):16-22 [FREE Full text] [Medline: 24402468]
23. Boudreaux ED, Abar B, Haskins B, Bauman B, Grissom G. Health evaluation and referral assistant: a randomized controlled trial to improve smoking cessation among emergency department patients. Addict Sci Clin Pract 2015 Nov 05;10:24 [FREE Full text] [doi: 10.1186/s13722-015-0045-2] [Medline: 26542471]

XSL•FO
RenderX

24. Cunningham RM, Harrison SR, McKay MP, Mello MJ, Sochor M, Shandro JR, et al. National survey of emergency department alcohol screening and intervention practices. Ann Emerg Med 2010 Jun;55(6):556-562. [doi: 10.1016/j.annemergmed.2010.03.004] [Medline: 20363530]
25. Terrell F, Zatzick DF, Jurkovich GJ, Rivara FP, Donovan DM, Dunn CW, et al. Nationwide survey of alcohol screening and brief intervention practices at US Level I trauma centers. J Am Coll Surg 2008 Nov;207(5):630-638 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2008.05.021] [Medline: 18954773]
26. Boudreaux ED, Abar B, Baumann BM, Grissom G. A randomized clinical trial of the health evaluation and referral assistant (HERA): research methods. Contemp Clin Trials 2013 Jul;35(2):87-96 [FREE Full text] [doi: 10.1016/j.cct.2013.04.010] [Medline: 23665335]
27. Etter JF, Perneger TV. Effectiveness of a computer-tailored smoking cessation program: a randomized trial. Arch Intern Med 2001 Nov 26;161(21):2596-2601. [Medline: 11718591]
28. Holtz K, Landis R, Nemes S, Hoffman J. Development of a computerized screening system to identify substance abuse in primary care. J Healthc Qual 2001;23(3):34-7, 45. [Medline: 11378975]
29. Prochaska JO, Velicer WF, Fava JL, Ruggiero L, Laforge RG, Rossi JS, et al. Counselor and stimulus control enhancements of a stage-matched expert system intervention for smokers in a managed care setting. Prev Med 2001 Jan;32(1):23-32. [doi: 10.1006/pmed.2000.0767] [Medline: 11162323]
30. Revere D, Dunbar PJ. Review of computer-generated outpatient health behavior interventions: clinical encounters “in absentia”. J Am Med Inform Assoc 2001;8(1):62-79 [FREE Full text] [Medline: 11141513]
31. Rhodes KV, Lauderdale DS, Stocking CB, Howes DS, Roizen MF, Levinson W. Better health while you wait: a controlled trial of a computer-based intervention for screening and health promotion in the emergency department. Ann Emerg Med 2001 Mar;37(3):284-291. [doi: 10.1067/mem.2001.110818] [Medline: 11223765]
32. Velicer WF, Prochaska JO, Fava JL, Laforge RG, Rossi JS. Interactive versus noninteractive interventions and dose-response relationships for stage-matched smoking cessation programs in a managed care setting. Health Psychol 1999 Jan;18(1):21-28. [Medline: 9925042]
33. Zeiler CA, Nemes S, Holtz KD, Landis RD, Hoffman J. Responses to a drug and alcohol problem assessment for primary care by ethnicity. Am J Drug Alcohol Abuse 2002;28(3):513-524. [Medline: 12211363]
34. Boudreaux ED, Bedek KL, Gilles D, Baumann BM, Hollenberg S, Lord SA, et al. The Dynamic Assessment and Referral System for Substance Abuse (DARSSA): development, functionality, and end-user satisfaction. Drug Alcohol Depend 2009 Jan 1;99(1-3):37-46 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.06.015] [Medline: 18775606]
35. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
36. Babor TE, Higgins-Biddle J, Dauser D, Higgins P, Burleson JA. Alcohol screening and brief intervention in primary care settings: implementation models and predictors. J Stud Alcohol 2005 May;66(3):361-368. [Medline: 16047525]
37. mail.umassmed.edu. 2014. Planning and implementing screening and brief intervention for risky alcohol use: a step-by-step guide for primary care practices URL: https://mail.umassmed.edu/owa/redir.aspx?C=1EYUaKJHghE5hhrMTmU1P7Kn0EIH_Q4kqeEdg-4cVH11RLU3akrUCA.&URL=http%3A%2F%2Fwww.webcitation.org%2F6nwd2ph1T
38. 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]
39. William MR, Rollnick S. Motivational Interviewing, Third Edition: Helping People Change (Applications of Motivational Interviewing). New York, NY: The Guilford Press; 2013.
40. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change: applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589]
41. Boudreaux ED, Baumann BM, Perry J, Marks D, Francies S, Camargo CA, et al. Emergency department initiated treatments for tobacco (EDITT): a pilot study. Ann Behav Med 2008 Dec;36(3):314-325. [doi: 10.1007/s12160-008-9066-3] [Medline: 19050988]
42. Sobell L, Sobell M. National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism. Alcohol consumption measures URL: https://pubs.niaaa.nih.gov/publications/assessingalcohol/measures.htm [accessed 2017-01-17] [WebCite Cache ID 6nafxKC3x]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
ED: emergency department
GEE: generalized estimating equation
HERA: Health Evaluation and Referral Assistant
OR: odds ratio
PHQ-2: Patient Health Questionaire-2

XSL•FO
RenderX

RA: research assistant
RCT: randomized controlled trial
SBI: screening and brief intervention
SBIRT: screening, brief intervention, and referral to treatment
SD: standard deviation

Edited by R Calvo; submitted 14.10.16; peer-reviewed by A Takano, E Jennings; comments to author 30.11.16; revised version received 02.02.17; accepted 06.02.17; published 01.05.17.

Please cite as:
Haskins BL, Davis-Martin R, Abar B, Baumann BM, Harralson T, Boudreaux ED
Health Evaluation and Referral Assistant: A Randomized Controlled Trial of a Web-Based Screening, Brief Intervention, and Referral to Treatment System to Reduce Risky Alcohol Use Among Emergency Department Patients
J Med Internet Res 2017;19(5):e119
URL: http://www.jmir.org/2017/5/e119/
doi:10.2196/jmir.6812
PMID:28461283

©Brianna L Haskins, Rachel Davis-Martin, Beau Abar, Brigitte M Baumann, Tina Harralson, Edwin D Boudreaux. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Computer Administered Safety Planning for Individuals at Risk for Suicide: Development and Usability Testing

Edwin D Boudreaux[1], PhD; Gregory K Brown[2], PhD; Barbara Stanley[3], PhD; Rajani S Sadasivam[1], PhD; Carlos A Camargo Jr[4], DrPH, MD; Ivan W Miller[5], PhD

[1]University of Massachusetts Medical School, Worcester, MA, United States

[2]Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Psychiatry, Columbia University College of Physicians & Surgeons and New York State Psychiatric Institute, New York City, NY, United States

[4]Massachusetts General Hospital, Boston, MA, United States

[5]Butler Hospital and the Warren Alpert Medical School of Brown University, Providence, RI, United States

Corresponding Author:
Edwin D Boudreaux, PhD
University of Massachusetts Medical School
55 Lake Avenue North
Worcester, MA, 01655
United States
Phone: 1 508 334 3817
Fax: 1 508 421 1490
Email: edwin.boudreaux@umassmed.edu

Abstract

Background: Safety planning is a brief intervention that has become an accepted practice in many clinical settings to help prevent suicide. Even though it is quick compared to other approaches, it frequently requires 20 min or more to complete, which can impede adoption. A self-administered, Web-based safety planning application could potentially reduce clinician time, help promote standardization and quality, and provide enhanced ability to share the created plan.

Objective: The aim of this study was to design, build, and test the usability of a Web-based, self-administered safety planning application.

Methods: We employed a user-centered software design strategy led by a multidisciplinary team. The application was tested for usability with a target sample of suicidal patients. Detailed observations, structured usability ratings, and Think Aloud procedures were used. Suicidal ideation intensity and perceived ability to cope were assessed pre-post engagement with the Web application.

Results: A total of 30 participants were enrolled. Usability ratings were generally strong, and all patients successfully built a safety plan. However, the completeness of the safety plan varied. The mean number of steps completed was 5.5 (SD 0.9) out of 6, with 90% (27/30) of participants completing at least 5 steps and 67% (20/30) completing all 6 steps. Some safety planning steps were viewed as inapplicable to some individuals. Some confusion in instructions led to modifications to improve understandability of each step. Ratings of suicide intensity after completion of the application were significantly lower than preratings, pre: mean 5.11 (SD 2.9) versus post: mean 4.46 (SD 3.0), t_{27}=2.49, P=.02. Ratings of ability to cope with suicidal thoughts after completion of the application were higher than preratings, with the difference approaching statistical significance, pre: mean 5.93 (SD 2.9), post: mean 6.64 (SD 2.4), t_{27}=−2.03, P=.05.

Conclusions: We have taken the first step toward identifying the components needed to maximize usability of a self-administered, Web-based safety planning application. Results support initial consideration of the application as an adjunct to clinical contact. This allows for the clinician or other personnel to provide clarification, when needed, to help the patient build the plan, and to help review and revise the draft.

(J Med Internet Res 2017;19(5):e149) doi:10.2196/jmir.6816

KEYWORDS

technology; safety; health planning; suicide; computers; telemedicine

XSL•FO
RenderX

Introduction

Background

Suicide and suicide attempts have increased over the past 10 years in the United States despite efforts by national organizations to reduce them [1-4]. Safety planning is a common suicide prevention tool designed to help an individual develop a plan for managing his or her suicidal thoughts. Although no universally accepted safety planning method exists, the Safety Planning Intervention [5,6] has gained widespread acceptance in the suicide prevention community and has been incorporated into numerous treatment guidelines and interventions [5-8]. The Safety Planning Intervention [5] is collaboratively built by a clinician with a patient and encourages individuals to engage in six sequential steps when feeling suicidal: (1) identify early warning signs, (2) employ internal coping strategies, (3) distract with social engagement or change of environment, (4) access suicide-protective social support, (5) seek help through crisis resources, and (6) restrict access to lethal means. The Safety Planning Intervention has a strong empirical foundation supporting each of its six steps [5], as well as evidence that it improves the average number of outpatient mental health visits for suicidal patients during the 6 months following the index emergency department (ED) visit, when compared with treatment as usual [9].

Although the Safety Planning Intervention is quick when compared with other suicide prevention interventions such as cognitive behavioral therapy, it commonly takes 20 min or more to administer [5,10-13]. Given that many health care settings, such as EDs and primary care settings, are characterized by multiple competing demands and suffer from serious time constraints on how much time can be spent with individual patients, the time it takes to build a collaborative safety plan remains an important barrier limiting adoption and implementation in clinical practice [10]. Moreover, it takes clinician training and practice to build a high quality, personalized safety plan [5]. Consequently, not having ready access to well-trained clinicians who know how to build a good safety plan is a barrier that further impedes widespread implementation in most health care settings, especially those not specifically devoted to providing behavioral health care. Novel approaches to safety planning that decrease staff burden by making the process less time-intensive, which support the systematic building of high-quality plans, and remain effective in preventing suicidal behavior are needed.

Safety Planning

With the ubiquity of computers and Internet access in modern society, a Web-based, self-administered safety planning application could offer tremendous potential for accomplishing these objectives and improving scalability of safety planning. Supporting this premise, several mobile phone apps designed for self-administered safety planning are currently available [14-16]. However, these safety planning apps are not designed to be used as clinical tools in health care settings; as a result, they have significant limitations, including being dependent on a patient having a mobile phone, downloading the app successfully, and completing the safety planning steps, most of which require extensive text entry, during a clinical encounter. Should these barriers be resolved, review by the clinician on the patient's phone is not practical. Moreover, there is no published data on the usability of these mobile phone safety planning apps. Thus, we designed, built, and tested the usability of a Web-based, self-administered safety planning application that could be completed on a desktop, laptop, or tablet computer. Our goal was to have the application include standardized instructions, "as needed" access to video instructions, and ready downstream access to the safety plan for Web-based editing, reprinting, and sharing with other clinicians and caregivers. We do not intend for the Web-based application to replace the clinician; rather, we expect it will make building safety plans more efficient for the clinician by reducing the total time required to build a safety plan by offloading the bulk of the orientation to the safety plan and the creation of the plan itself to the computer. The clinician will still maintain a role in reviewing the safety plan and editing it, as needed.

Before such a system can be widely promoted, the usability of the application must first be established, and the alpha version of the application adapted based on user feedback. The final system's impact on clinical workflow, such as whether clinician efficiency improved when compared with traditional safety planning, and impact on suicide-related outcomes, such as whether suicidal behavior decreases, needs to be conducted. This paper describes the development process and initial results of the usability testing (ie, the first step). This is important, because usability can be seriously impacted by a variety of factors, even with a seemingly simple task, such as completing an online form. For example, usability can be adversely impacted by context or setting (eg, busy, distracting medical settings) and patient population (eg, mixed demographics, mixed computer literacy, medically ill, psychologically distressed). Moreover, this is the first systematic test of the transition of modalities from clinician administered to self-administered, and the team needed to validate that users could create at least the initial draft of the safety plans themselves.

Methods

Overview of Application Development

We employed a user-centered, iterative software design strategy. Our development was funded through the Emergency Department-Safety Assessment and Follow-Up Evaluation (ED-SAFE) [17] and overseen by one of the ED-SAFE's Principal Investigators (EDB). The team comprised psychologists, psychiatrists, emergency physicians, nurses, informaticians, software engineers, and additional subject matter experts, including the developers of the Safety Planning Intervention [5] (GKB, BS), and included both experienced researchers and active clinicians. The entire team was intimately involved with every step of the design and testing process to ensure that the application was firmly rooted in the latest advances in safety planning and was user-friendly. It was built to be compliant with applicable health care security and privacy standards, including the Health Insurance Portability and Accountability Act (HIPAA).

XSL•FO
RenderX

The computer application used the same six steps as the Safety Planning Intervention, but the order of the steps was changed to put means restriction first, rather than last, because members of the team noted that it is important for the individual to restrict means to suicide immediately upon discharge from the hospital. Presenting means restriction first aligned with the presumed order of action and priority upon discharge. The new order, along with associated instructions and response options used for the application, is presented in Figure 1. Multimedia Appendix 1 contains some representative screenshots.

The original Safety Planning Intervention, like other versions of safety planning, is designed to be completed by clinician interview, with the patient's responses documented in text on a template paper form. We considered simply computerizing this by presenting instructions for each of the six steps followed by an open text field to allow the patient to enter his or her free-text responses. However, this was viewed by the design team to be a potential barrier because of the need for computer keyboard literacy and with potential for lengthening the time of administration. Consequently, to facilitate rapid, user-friendly completion, multiple choice options were developed for the first three steps. The response options were created by soliciting suggestions from subject matter experts, iterative review by the development team and consultants, and revision based on patient input to select the final response options and to perfect the wording. A free-text field was included for each step to allow individuals to input their own responses, if and when they desire. This combination of multiple choice and free-text balanced the need for simplicity with the goal of personally tailored planning. Each of the six traditional Safety Planning Intervention steps is presented sequentially, with easy-to-follow instructions and a "MORE" button that provides additional detail, if needed. A brief instructional video was developed to provide additional instructions to the patient, accessible as needed.

Once all the six steps have been completed, the individual reviews and "confirms" the safety plan. The printed version is formatted similar to the traditional Safety Planning Intervention and is designed to fit on a single page for ease of printing, access, and manipulation. The safety plan is stored on a secure server, and the individual can access the plan from any computer or mobile phone with Internet connection. The safety plan can be edited, printed, saved as a portable document format (PDF) and shared with others by email through a secure email service.

Setting and Participants

The usability testing was set in an urban, tertiary care hospital in Central Massachusetts. Consecutively presenting adult patients being evaluated for an acute psychiatric emergency in either the ED or inpatient psychiatric unit were screened during research assistant (RA) shifts. Inclusion criteria included 18 years of age and endorsement of active suicidal ideation in the past two weeks. Exclusion criteria included persistent severe medical illness, severe emotional distress, cognitive insufficiency (eg, dementia, psychosis, altered consciousness), incarceration, and insurmountable language barriers.

Usability Testing Procedures

Eligible patients, who agreed to participate, signed written informed consent. We used a "Think Aloud" protocol testing approach [18,19]; although the participants interacted with the application and completed their safety plan, they were asked to vocalize their thoughts, feelings, and opinions about each screen. Think Aloud informs how a user approaches the interface and his or her mental processes when utilizing the interface. Additionally, Morae (TechSmith version 3.2.1, 2010) usability software was used [20-22]. It allows for video and audio recordings of the subject being tested, including recording clicks, keystrokes, and other events. For our purpose, one reviewer reviewed the audio recordings and summarized the issues identified by the participant. Structured usability ratings and open-ended questions asking about pros, cons, and recommendations for improvement were obtained by the RA. All problems or difficulties encountered throughout the protocol through Think Aloud, direct observation, and patient interview were documented in detail and reviewed with the study team weekly. Software, instructions, and item wording were modified in response to feedback.

Because there was some concern that having patients engage in self-administered safety planning could actually have the paradoxical effect of increasing momentary intensity of suicidal ideation, participants provided ratings of suicidal ideation before and after engaging in the safety planning. Conversely, engaging with the application could have the effect of improving an individual's perceived ability to cope with suicidal thoughts, so ratings of ability to cope with suicidal thoughts were obtained. The medical records of each participant were reviewed for one month before and after the enrollment date to assess acute suicide-related health care utilization.

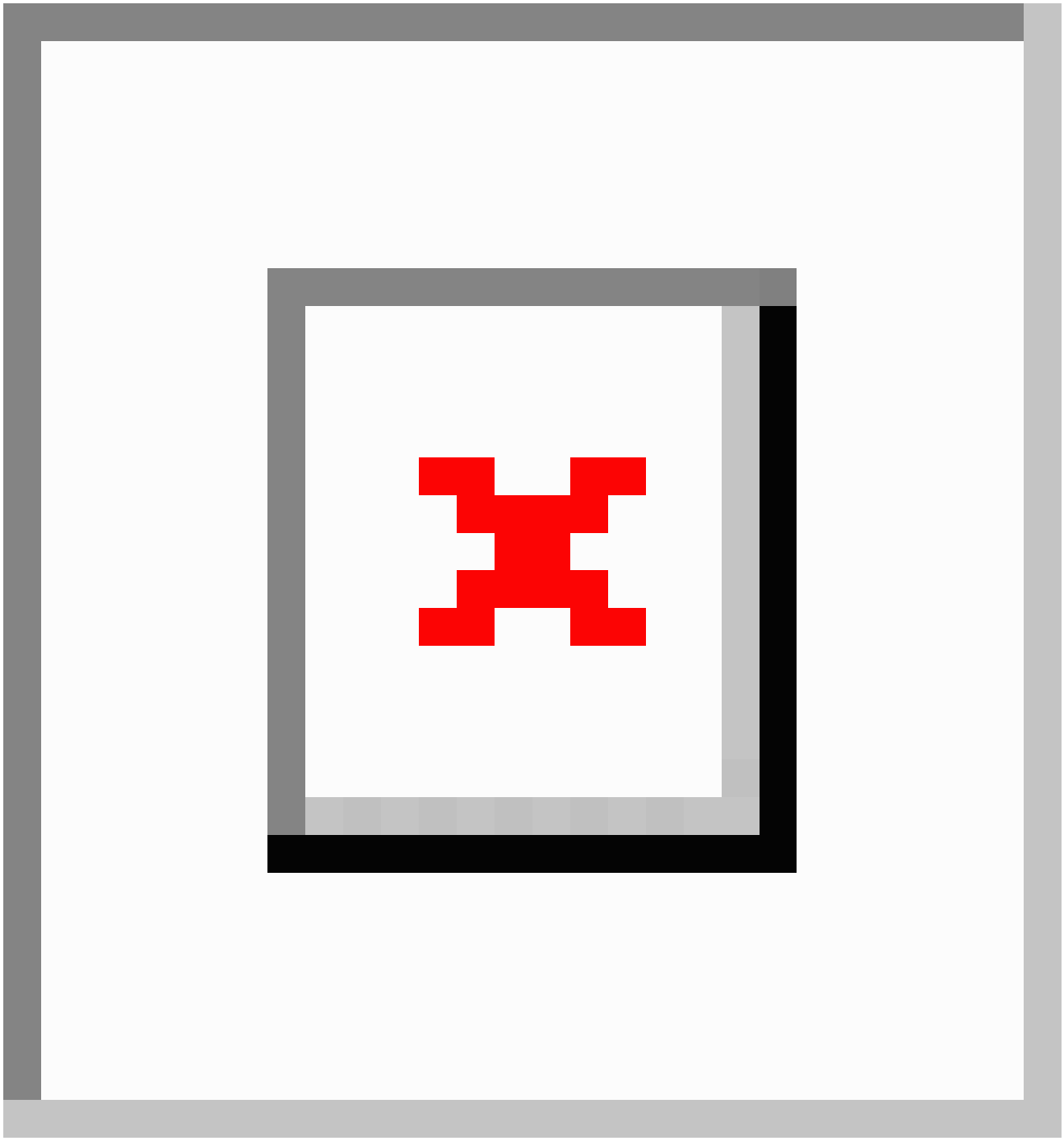
Measures

Demographics and Descriptives

Age, sex, race (white, black or African American, Asian or Pacific Islander, American Indian, Alaska Native, Aleut, Other), ethnicity (Hispanic, non-Hispanic), and insurance type (private, Medicare, Medicaid or State, other, none) were documented on all enrolled patients. In addition to demographics, we abstracted the following information from their medical record: presence of alcohol abuse (current intoxication or evidence of any problem use), intentional illegal or prescription drug misuse (current intoxication or evidence of any problem use), presence of depressed mood, disposition (discharged home, admitted to medical unit, admitted or transferred to psychiatric or substance abuse unit), and emergency physician discharge diagnosis.

XSL•FO
RenderX

Figure 1. Safety planning application steps.



Safety Planning Application

All patients were provided the laptop on the "home" page. They registered by creating a username and strong password and then read and responded to each of the six steps sequentially, as summarized in Figure 1.

Process Log

The RA completed a process log for each participant, documenting any problems noted, the solutions applied, and the outcome. Problems were categorized based on the following domains: technical failure (eg, Internet disconnection, hardware dysfunction), computer literacy (eg, mouse use, text entry), safety plan step completion (eg, trouble understanding how to respond, applicability of the step), interruptions during safety plan completion (eg, clinical care, meals, visitors), and use of proxy to help complete the plan (eg, family member or visitor).

Usability Ratings

All participants provided structured usability ratings (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree) across a variety of specific tasks, including ability to easily move between screens, ability to understand the language of the safety planning steps, confidence in being able to create a safety plan, helpfulness of the instructional video (if watched), desire to send the safety plan to someone else by email, likelihood of using the safety plan if suicidal thoughts arise in the future, and understanding of how the safety plan can help manage suicidal thoughts. These domains were identified by

XSL•FO
RenderX

the study team as tasks considered important for usability and most consistent with the ultimate goal of the application. Open-ended questions were used to assess the positive and negative impressions of the application, as well as recommendations for improvement. The ratings and interview were administered by the RA after the participant completed their safety plan.

Suicide Ratings

Two aspects of suicidal ideation were measured immediately before and after engagement with the application: suicidal ideation intensity (0=none, 10=constant) and perceived ability to cope with suicidal ideation (0=no ability to cope, 10=strong ability to cope). Although standardized, validated instruments measuring suicidality have advantages, there were two reasons we used simple 0-10 point items instead. First, we enrolled subjects in the ED setting. This setting is very time-sensitive, and patients were often uncomfortable or distressed, so we wanted the total participant burden to be as light as possible, precluding lengthier scales. Second, we were interested in detecting changes over a very short period of time—pre- and post-engagement with the software. Standardized scales are not designed to detect short-term state-changes in suicidal ideation.

Emergency Department Utilization Review

The medical records for the month before and after the index visit were reviewed. All ED visits were noted and classified based on whether or not they were related to suicidal ideation or behavior (0=no, 1=yes). Only medical records associated with the study's health care system were accessible. Although it would have been advantageous to acquire medical records from other health systems, this proved infeasible because of patient confidentiality laws.

Data Analysis

The primary outcomes reflecting usability were summarized using descriptive statistics, including proportions, means, and standard deviations (SDs). The safety planning application technical completion rate was defined as the proportion of participants who initiated the application and were able to complete navigation from start to end without critical technical failures, usability issues, or other interruptions resulting in an aborted encounter. The safety plan step completion rate was defined as the proportion of participants who provided at least one response for each of the six steps. This completion rate was also calculated for each of the six steps individually. Process problems were summarized as present or absent based on the categories described in "Measures—Process Log" section. Descriptive statistics were calculated on usability ratings for each domain described in "Measures—Usability Ratings" section. Additional analyses included paired sample *t* test comparing suicidal ideation intensity and ability to cope with suicidal ideation pre-post safety plan administration, as well as a paired sample *t* test comparing suicide-related acute health care visits in the one month before and after the index visit. Statistical Package for the Social Sciences 22 (IBM, Armonk, NY) was used for all analyses.

Results

Demographics and Descriptives

A total of 69 patients with suicidal ideation or a suicide attempt were approached by the RA; of these, 37 met all of the eligibility criteria, and 81% of those eligible (30/37) consented to participate and were enrolled. The average age of the sample was 39 years old (SD 14 years), with 47% (14/30) being male, 83% (25/30) indicating white race, and 10% (3/30) indicating Hispanic ethnicity. The majority (n=25/30; 83%) were insured by Medicaid or another state program. The clinical characteristics of the sample were as follows: presence of alcohol abuse (current intoxication or evidence of any problem use), 30% (9/30); intentional illegal or prescription drug misuse (current intoxication or evidence of any problem use), 30% (9/30); presence of depressed mood, 53% (16/30); and disposition, 10% (3/30); discharged home; admitted to medical unit, 3% (1/30); and admitted or transferred to psychiatric or substance abuse unit, 87% (26/30). The most common primary emergency physician discharge diagnosis was mood-related (eg, depression, emotional distress, emotional crisis), which was assigned to 60% (18/30). The remaining had a variety of medical and substance abuse diagnoses.

Usability

Multimedia Appendix 2 summarizes the usability statistics. All participants successfully registered, viewed all six of the safety planning build screens, and created a final, one-page safety plan PDF that could be printed, saved, or shared by secure email. However, completeness of the safety plan varied. The mean number of steps completed was 5.5 (SD 0.9), with 90% (27/30) completing at least 5 steps and 67% (20/30) completing all six steps.

It was found that 3 of the first 5 subjects experienced some type of technical failure, including interrupted Internet connection and "frozen" screen resulting from a glitch in the Morae usability software. At the time of enrollment, these problems were rectified by the RA, and they did not prevent the participants from resuming the application and completing their safety plans. Root causes were identified and addressed, resulting in none of the final 25 enrollees experiencing technical failures.

In total, 40% (12/30) of the participants reported or demonstrated at least one problem related to computer literacy, including unfamiliarity with using the mousepad and with typing on a keyboard. This did not result in any aborted safety plans, but it did require some modest technical assistance from the RA on occasion. It was found that 23% (7/30) of the participants reported or demonstrated problems with the actual process of building the safety plan that was not related to general computer literacy. The initial registration process caused some confusion because it required creation of a strong password; instructions around this process were improved, resulting in the final 20 enrollees reporting no problems with the registration process. Some participants needed clarification on how to complete individual steps. Most commonly, early in the study, patients confused Steps 4 and 5, because both involve identifying people who can help the individual in some way. Although Step 4

XSL•FO
RenderX

instructs the individual to identify people who can help *distract* from suicidal thoughts, Step 5 instructs the individual to identify people they can confide in and who can actually help them manage their suicidal thoughts and associated problems. Some of the confusion arose because the same set of people applied to both steps. This overlap is acceptable for building a safety plan, but the instructions initially were unclear, creating uncertainty within the participants. On the basis of the participant feedback, the instructions were modified to help clarify this distinction between the two steps. Instructions were improved, resulting in no difficulty identified with comprehension for the final 10 enrollees.

Completion of Step 5 (*identifying individuals who can help the individual through a suicidal crisis*) demonstrated the highest incompletion rate; this was not only because of the aforementioned confusion with Step 4, but also because some participants simply did not have anyone with whom they felt comfortable talking about their suicidal thoughts. Consequently, they left it blank. This perceived lack of personal relevance was a common cause for other steps being left blank as well. For example, Step 1 (*lethal means restriction*) was not answered by an individual who reported not having any lethal means accessible to him, and Step 6 (*professional services to access when suicidal*) was left blank by an individual who had no existing physician or clinician and did not know the address of the nearest ED, so could not populate any of the fields. The software was modified to allow patients to select "skip or does not apply."

Interruptions during the process were relatively common (n=8/30; 27%) due to testing, meals, and visitors. These interruptions did not prevent the individuals from building a safety plan, but it often required them to pause the process and resume later. Although the instruction video was available to all, only 17% (5/30) chose to watch it, with the others stating that they did not think they needed to because the text instructions were clear.

In general, the usability ratings were strong, with the averages ranging from 3.8 (*helpfulness of the video; n=5*) to 4.4 (*understanding safety plan step instructions*). It was found that 80% (24/30) of the participants agreed or strongly agreed that they would use their safety plan in the future if they experienced suicidal ideation. Open-ended comments generally supported the favorable usability.

Suicide Ratings

Ratings of suicide intensity after completion of the application were significantly lower than preratings, pre: mean 5.11 (SD 2.9) versus post: mean 4.46 (SD 3.0), t_{27}=2.49, P=.02. Ratings of ability to cope with suicidal thoughts after completion of the application were higher than preratings, with the difference approaching statistical significance, pre: mean 5.93 (SD 2.9), post: mean 6.64 (SD 2.4), t_{27}=−2.03, P=.05.

Suicide Related Acute Health Care

The total number of ED visits related to suicide decreased from an average of 1.00 (SD 0.45) visit in the month before the index visit to an average of 0.07 (SD 0.25) after the index visit, t_{29}=11.34, P<.001. In the month before the index visit, 93% (28/30) of patients had one or more suicide-related ED visits; in the month after the index visit, 7% (2/30) of participants had one or more suicide-related ED visits.

Discussion

Principal Findings

We designed and built a self-administered safety planning application for use with patients at risk for suicide and assessed its usability in 30 suicidal patients. All 30 participants were able to successfully navigate the software and produce a safety plan. Self-reported usability ratings were generally positive, even with early recruits using the initial version that had not yet benefited from adjustments resulting from user testing. However, our observations revealed important parameters surrounding the application's usability and optimal deployment, including the need for technical problem solving during early deployment, ensuring the availability of clinical or other personnel to help with usability issues if and when they arise, and review and editing of the safety plan with a clinician, with particular attention given to understanding skipped steps.

We experienced technical problems in 3 of the first 5 participants. These were navigated in real-time by the RA, allowing the 3 participants to complete the safety plan, and root cause analyses reduced technical failures to 0 for the final 25 enrollees. This experience highlighted the importance of identifying and addressing rudimentary technical issues before and during initial clinical deployment. Despite generally positive self-reported usability ratings, 40% (12/30) of participants self-reported or demonstrated upon direct observation at least one minor computer literacy issue. None of these computer literacy problems prevented the individual from building a safety plan; instead, it simply caused the build to take longer than it would have otherwise taken if the issues had not been encountered and, at times, resulted in the RA having to provide instruction, such as showing the individual how to use the mousepad.

It was found that 33% (10/30) participants skipped at least one of the six steps, which was due to two primary reasons. First, some of the instructions were not clear about exactly what the step required of the individual. Most of the confusion centered around Steps 4 and 5, as described in the "Results" section. Both Steps 4 and 5 required the participant to identify people who can help the individual in some way, with Step 4 identifying people who can help distract from suicidal thoughts and Step 5 identifying people they can confide in and who can actually help them manage their suicidal thoughts and associated problems. This led to confusion because many individuals did not understand the distinction, or pointed out that the same set of people applied to both steps. With improvements to the clarity of the instructions, these issues were resolved. Second, individual steps that were not viewed as applicable or personally relevant, were skipped. This occurred most frequently with Step 5, which many patients chose to skip because they had no one in their social network with whom they felt comfortable divulging their suicidal thoughts. Reluctance to discuss suicide is not uncommon; it may also be reflective of the tendency for

XSL•FO
RenderX

suicidal individuals to be socially isolated [23,24]. As a result, the software was modified to allow an individual to indicate if he or she desired to skip the step. It is important for clinicians to review skipped steps with the patient to determine the reason and to help problem-solve alternatives. Safety planning is primarily a process of helping patients to learn to use coping strategies during a suicidal crisis. We believe clinicians, who are competently trained in this intervention, are an essential component. However, the degree of clinical involvement required when using the safety planning Web application could be minimized and streamlined. The two models—traditional safety planning administered by a clinician using a paper-based form versus safety planning facilitated by the self-administered, Web-based application with clinician review—warrants further study in a randomized clinical trial examining not only clinical efficacy on suicide-related outcomes but also workflow efficiency, cost effectiveness, and clinician and patient satisfaction.

The application allowed multiple choice and free-text options for the first three steps. All participants completed the first three steps, with the vast majority choosing at least one of the multiple choice response options for each step. Every multiple choice response option across the three steps was selected by at least one participant. Free-text responses were entered by many as well, ranging from 23% (7/30) for Step 3 to 53% (16/30) for Step 1. This pattern supports continued use of multiple choice options, their relevance supported by their selection by patients, as well as inclusion of the free-text capabilities to allow for added personalization.

Engagement in the application did not appear to increase suicidal ideation intensity; in fact, the pattern was the opposite. Statistical decreases in suicidal ideation intensity, and increases in perceived ability to cope with suicidal ideation, were observed when comparing pre- and post-ratings. ED visits related to suicidal ideation or behavior also decreased in the month after enrollment, when compared with the number of visits in the month before enrollment. These indices were exploratory and interpretation should be conservative, because there was no control condition against which to compare performance. Moreover, decreases in ED visits after the index visit could be the result of other interventions delivered to mitigate the participants' suicide risk, such as inpatient psychiatric hospitalization, which occurred in 87% (26/30) participants.

Limitations

The safety planning application is subject to the limitations inherent in a Web-based platform; it requires a computer, Internet connectivity, some computer literacy, and general literacy. The use of multiple choice options may funnel patients into options that are not tailored to them, a scenario that the option of free-text fields may only partly counteract. The study was a usability study, not a trial, so it did not follow patients prospectively or have a control condition against which to compare the application's impact on suicidal ideation or behavior. Furthermore, although a self-administered version has scalability, it may lack the personalization and expert guidance that a clinician-administered version has. For example, the clinician-administered version aids patients in determining which strategies and individuals will be most helpful and safest for them, allowing for a collaborative evaluation of their value, a task that patients, on their own, may not be able to accomplish. Another limitation was the sample size. It was small and may not represent all suicidal patients. This is acceptable for early usability tests; further testing in a large-scale trial should also seek to better identify subpopulations for whom the intervention is most appropriate and effective. Finally, the RA attempted to preserve naturalistic, real-world administration. However, she was present to observe and document problems and answer any questions. This is an inevitable tension in early usability studies that are also designed to obtain information to improve the software through using real-time feedback.

Conclusions

A computerized, self-administered safety planning system that produces high-quality safety plans anywhere, anytime is highly innovative. Although most medical settings are usually fraught with pressing time demands for clinicians, leading to deprioritizing interventions like safety planning, patients often have considerable downtime as they are waiting for tests, clinicians, consultations, or procedures. Many settings simply lack the available interventionists to provide safety planning during the encounter. Our application has strong potential to address this problem. However, before its full impact can be tested, establishing its usability in challenging real-world settings is an essential first step. We have built the application with downstream dissemination in mind and have taken the first step toward identifying the components needed to maximize usability and foster adoption in clinical settings. It is not intended to be a stand-alone intervention. Rather, it is intended to allow the patient to begin the safety plan and allows for the clinician or other personnel to provide clarification, when needed, to help the patient build the plan, and to help review and revise the draft. A deployment where the patient can ask questions and receive technical assistance, whereas having clinicians inquire about skipped steps to clarify the reasons for the omission, is optimal. The next step is to demonstrate clinical efficacy of this model.

Acknowledgments

The authors would like to acknowledge the editorial assistance of Brianna L Haskins, MS, of the Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA, USA. This project was funded by the National Institute of Mental Health (U01MH088278, Administrative Supplement). This was a collaborative award from the NIMH, whose representatives were a member of the larger study team, though none participated in this substudy. None of the authors have financial disclosures.

XSL•FO
RenderX

Authors' Contributions

EDB advised on study design, data analysis, interpretation, and write-up. GKB, BS, CAC, and IWM were involved in the write-up. RS advised on data analysis, interpretation, and write-up.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Safety planning intervention Web application steps 1-5 screenshots.

[PDF File (Adobe PDF File), 762KB - jmir_v19i5e149_app1.pdf]

Multimedia Appendix 2

Usability statistics (N=30).

[PDF File (Adobe PDF File), 39KB - jmir_v19i5e149_app2.pdf]

References

1. Chakravarthy B, Frumin E, Lotfipour S. Increasing suicide rates among middle-age persons and interventions to manage patients with psychiatric complaints. West J Emerg Med 2014 Feb;15(1):11-13 [FREE Full text] [doi: 10.5811/westjem.2013.12.19513] [Medline: 24578763]
2. Centers for Disease Control and Prevention. Suicide now claims more US lives than vehicular crashes. JAMA 2013 Jun 19;309(23):2432. [doi: 10.1001/jama.2013.6689]
3. Curtin SC, Warner M, Hedegaard H. Increase in Suicide in the United States, 1999-2014. NCHS Data Brief 2016 Apr(241):1-8 [FREE Full text] [Medline: 27111185]
4. Ting SA, Sullivan AF, Boudreaux ED, Miller I, Camargo CA. Trends in US emergency department visits for attempted suicide and self-inflicted injury, 1993-2008. Gen Hosp Psychiatry 2012;34(5):557-565 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.03.020] [Medline: 22554432]
5. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. Cogn Behav Pract 2012 May;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
6. Stanley B, Brown GK. Mental Health. 2008. Safety plan treatment manual to reduce suicide risk: veteran version URL: http://www.mentalhealth.va.gov/docs/va_safety_planning_manual.pdf [accessed 2016-10-14] [WebCite Cache ID 6lF5M8Qtf]
7. SPRC. ZERO Suicide in health and behavioral health care URL: http://zerosuicide.sprc.org/ [accessed 2016-10-14] [WebCite Cache ID 6lF5gVa7E]
8. American Psychiatric Association. Practice guideline for the assessment and treatment of patients with suicidal behaviors. Am J Psychiatry 2003 Nov;160(11 Suppl):1-60. [Medline: 14649920]
9. Knox KL, Stanley B, Currier GW, Brenner L, Ghahramanlou-Holloway M, Brown G. An emergency department-based brief intervention for veterans at risk for suicide (SAFE VET). Am J Public Health 2012 Mar;102(Suppl 1):S33-S37. [doi: 10.2105/AJPH.2011.300501] [Medline: 22390597]
10. Chesin MS, Stanley B, Haigh EA, Chaudhury SR, Pontoski K, Knox KL, et al. Staff views of an emergency department intervention using safety planning and structured follow-up with suicidal veterans. Arch Suicide Res 2016 Apr 20:1-11. [doi: 10.1080/13811118.2016.1164642] [Medline: 27096810]
11. Stanley B, Brown GK, Currier GW, Lyons C, Chesin M, Knox KL. Brief intervention and follow-up for suicidal patients with repeat emergency department visits enhances treatment engagement. Am J Public Health 2015 Aug;105(8):1570-1572. [doi: 10.2105/AJPH.2015.302656] [Medline: 26066951]
12. Stanley B, Chaudhury SR, Chesin M, Pontoski K, Bush AM, Knox KL, et al. An emergency department intervention and follow-up to reduce suicide risk in the VA: acceptability and effectiveness. Psychiatr Serv 2016 Jun 01;67(6):680-683. [doi: 10.1176/appi.ps.201500082] [Medline: 26828397]
13. Rings J, Alexander P, Silvers V, Gutierrez P. Adapting the safety planning intervention for use in a veterans psychiatric inpatient group setting. J Ment Health Couns 2012 Apr;34(2):95-109. [doi: 10.17744/mehc.34.2.a77036631424nmq7]
14. Larsen ME, Cummins N, Boonstra TW, O'Dea B, Tighe J, Nicholas J, et al. The use of technology in suicide prevention. Conf Proc IEEE Eng Med Biol Soc 2015;2015:7316-7319. [doi: 10.1109/EMBC.2015.7320081] [Medline: 26737981]
15. Vahabzadeh A, Sahin N, Kalali A. Digital suicide prevention: can technology become a game-changer? Innov Clin Neurosci 2016;13(5-6):16-20 [FREE Full text] [Medline: 27800282]
16. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]

XSL•FO
RenderX

17. Boudreaux ED, Miller I, Goldstein AB, Sullivan AF, Allen MH, Manton AP, et al. The emergency department safety assessment and follow-up evaluation (ED-SAFE): method and design considerations. Contemp Clin Trials 2013 Sep;36(1):14-24 [FREE Full text] [doi: 10.1016/j.cct.2013.05.008] [Medline: 23707435]
18. Kushniruk AW. Analysis of complex decision-making processes in health care: cognitive approaches to health informatics. J Biomed Inform 2001 Oct;34(5):365-376 [FREE Full text] [doi: 10.1006/jbin.2001.1021] [Medline: 12123153]
19. Kushniruk AW, Patel VL. Cognitive computer-based video analysis: its application in assessing the usability of medical systems. Medinfo 1995;8 Pt 2:1566-1569. [Medline: 8591502]
20. Yen P, Bakken S. Usability testing of a web-based tool for managing open shifts on nursing units. Stud Health Technol Inform 2009;146:81-85 [FREE Full text] [Medline: 19592813]
21. Choi J, Bakken S. Heuristic evaluation of a web-based educational resource for low literacy NICU parents. Stud Health Technol Inform 2006;122:194-199. [Medline: 17102247]
22. TechSmith. 2014. Morae URL: http://www.techsmith.com/morae.html [accessed 2016-10-14] [WebCite Cache ID 6lF6yttYJ]
23. King CA, Merchant CR. Social and interpersonal factors relating to adolescent suicidality: a review of the literature. Arch Suicide Res 2008;12(3):181-196 [FREE Full text] [doi: 10.1080/13811110802101203] [Medline: 18576200]
24. Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. Psychol Rev 2010 Apr;117(2):575-600 [FREE Full text] [doi: 10.1037/a0018697] [Medline: 20438238]

Abbreviations

ED: emergency department
ED-SAFE: Emergency Department-Safety Assessment and Follow-Up Evaluation
HIPAA: Health Insurance Portability and Accountability Act
PDF: portable document format
RA: research assistant
SD: standard deviation

Edited by R Calvo; submitted 14.10.16; peer-reviewed by S Berrouiguet, G Wadley, J Thrul; comments to author 17.11.16; revised version received 20.01.17; accepted 08.02.17; published 15.05.17.

Please cite as:
Boudreaux ED, Brown GK, Stanley B, Sadasivam RS, Camargo CA Jr, Miller IW
Computer Administered Safety Planning for Individuals at Risk for Suicide: Development and Usability Testing
J Med Internet Res 2017;19(5):e149
URL: http://www.jmir.org/2017/5/e149/
doi:10.2196/jmir.6816
PMID:28506957

©Edwin D Boudreaux, Gregory K Brown, Barbara Stanley, Rajani S Sadasivam, Carlos A Camargo Jr, Ivan W Miller. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX