

Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 20 (2018), Issue 6 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic Review of Reviews (e10202) Paolo Zanaboni, Patrice Ngangue, Gisele Mbemba, Thomas Schopf, Trine Bergmo, Marie-Pierre Gagnon.	6
The Social Effects of Exergames on Older Adults: Systematic Review and Metric Analysis (e10486) Jinhui Li, Mojisola Erdt, Luxi Chen, Yuanyuan Cao, Shan-Qi Lee, Yin-Leng Theng.	19
Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety (e199) Theresa Fleming, Lynda Bavin, Mathijs Lucassen, Karolina Stasiak, Sarah Hopkins, Sally Merry.	65
Online Lifestyle Modification Intervention: Survey of Primary Care Providers' Attitudes and Views (e167) Reem Hanna, Gary Fischer, Molly Conroy, Cindy Bryce, Rachel Hess, Kathleen McTigue.	76
Internet-Based Interventions Aimed at Supporting Family Caregivers of People With Dementia: Systematic Review (e216) Jenny Hopwood, Nina Walker, Lorraine McDonagh, Greta Rait, Kate Walters, Stephen Iliffe, Jamie Ross, Nathan Davies.	84
Effect Evaluation of a Web-Based Coaching Intervention to Support Implementation of Sex Education Among Secondary School Teachers: Randomized Controlled Trial (e96) Lisette Schutte, Fraukje Mevissen, Suzanne Meijer, Theo Paulussen, Pepijn van Empelen, Gerjo Kok.	96
Severely Burdened Individuals Do Not Need to Be Excluded From Internet-Based and Mobile-Based Stress Management: Effect Modifiers of Treatment Outcomes From Three Randomized Controlled Trials (e211) Kiona Weisel, Dirk Lehr, Elena Heber, Anna-Carlotta Zarski, Matthias Berking, Heleen Riper, David Ebert.	108
The Effectiveness of Web-Based Tailored Smoking Cessation Interventions on the Quitting Process (Project Quit): Secondary Analysis of a Randomized Controlled Trial (e213) Bibhas Chakraborty, Raju Maiti, Victor Strecher.	124
An Online Hand Exercise Intervention for Adults With Rheumatoid Arthritis (mySARAH): Design, Development, and Usability Testing (e10457) Cynthia Srikesavan, Esther Williamson, Tim Cranston, John Hunter, Jo Adams, Sarah Lamb.	134
Web-Based Intervention for Family Carers of Persons with Dementia and Multiple Chronic Conditions (My Tools 4 Care): Pragmatic Randomized Controlled Trial (e10484) Wendy Duggleby, Jenny Ploeg, Carrie McAiney, Shelley Peacock, Kathryn Fisher, Sunita Ghosh, Maureen Markle-Reid, Jennifer Swindle, Allison Williams, Jean Triscott, Dorothy Forbes, Kathya Jovel Ruiz.	147

Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study (e210)	
Akane Sano, Sara Taylor, Andrew McHill, Andrew Phillips, Laura Barger, Elizabeth Klerman, Rosalind Picard.	161
Multicomponent mHealth Intervention for Large, Sustained Change in Multiple Diet and Activity Risk Behaviors: The Make Better Choices 2 Randomized Controlled Trial (e10528)	
Bonnie Spring, Christine Pellegrini, H McFadden, Angela Pfammatter, Tammy Stump, Juned Siddique, Abby King, Donald Hedeker.	183
What People “Like”: Analysis of Social Media Strategies Used by Food Industry Brands, Lifestyle Brands, and Health Promotion Organizations on Facebook and Instagram (e10227)	
Karen Klassen, Emily Borleis, Linda Brennan, Mike Reid, Tracy McCaffrey, Megan Lim.	197
Social Drinking on Social Media: Content Analysis of the Social Aspects of Alcohol-Related Posts on Facebook and Instagram (e226)	
Hanneke Hendriks, Bas Van den Putte, Winifred Gebhardt, Megan Moreno.	206
Young People’s Experiences of Viewing the Fitspiration Social Media Trend: Qualitative Study (e219)	
Stephanie Easton, Katherine Morton, Zara Tappy, Daniella Francis, Laura Dennison.	218
Recruiting Hard-to-Reach Populations for Survey Research: Using Facebook and Instagram Advertisements and In-Person Intercept in LGBT Bars and Nightclubs to Recruit LGBT Young Adults (e197)	
Jamie Guillory, Kristine Wiant, Matthew Farrelly, Leah Fiacco, Ishrat Alam, Leah Hoffman, Erik Crankshaw, Janine Delahanty, Tesfa Alexander.	22
Connecting With Your Dentist on Facebook: Patients’ and Dentists’ Attitudes Towards Social Media Usage in Dentistry (e10109)	
Nilesh Parmar, Lin Dong, Andreas Eisingerich.	243
Increasing the Availability of Psychological Treatments: A Multinational Study of a Scalable Method for Training Therapists (e10386)	
Marianne O’Connor, Katy Morgan, Suzanne Bailey-Straepler, Christopher Fairburn, Zafra Cooper.	255
Reproductive Health and Medication Concerns for Patients With Inflammatory Bowel Disease: Thematic and Quantitative Analysis Using Social Listening (e206)	
Michelle Keller, Sasan Mosadeghi, Erica Cohen, James Kwan, Brennan Spiegel.	266
eHealth as the Next-Generation Perinatal Care: An Overview of the Literature (e202)	
Josephus van den Heuvel, T Groenhof, Jan Veerbeek, Wouter van Solinge, A Lely, Arie Franx, Mireille Bekker.	276
Relevance of a Mobile Internet Platform for Capturing Inter- and Intrasubject Variabilities in Circadian Coordination During Daily Routine: Pilot Study (e204)	
Sandra Komarzynski, Qi Huang, Pasquale Innominato, Monique Maurice, Alexandre Arbaud, Jacques Beau, Mohamed Bouchahda, Ayhan Ulusakarya, Nicolas Beaumatin, Gabriële Breda, Bärbel Finkenstädt, Francis Lévi.	291
Telemedicine Use Among Caregivers of Cancer Patients: Systematic Review (e223)	
Chiara Marzorati, Chiara Renzi, Samuel Russell-Edu, Gabriella Pravettoni.	309
A Teledermatology Scale-Up Framework and Roadmap for Sustainable Scaling: Evidence-Based Development (e224)	
Laticha Walters, Richard Scott, Maurice Mars.	324
Home-Based Rehabilitation With Telemonitoring Guidance for Patients With Coronary Artery Disease (Short-Term Results of the TRiCH Study): Randomized Controlled Trial (e225)	
Andrea Avila, Jomme Claes, Kaatje Goetschalckx, Roselien Buys, May Azzawi, Luc Vanhees, Véronique Cornelissen.	334

Effects of the Digital Transformation: Qualitative Study on the Disturbances and Limitations of Using Video Visits in Outpatient Care (e221)	
Linda Sturesson, Kristina Groth.	347
Applications of Space Technologies to Global Health: Scoping Review (e230)	
Damien Dietrich, Ralitz Dekova, Stephan Davy, Guillaume Fahrni, Antoine Geissbühler.	360
Patient Acceptance of Remote Scribing Powered by Google Glass in Outpatient Dermatology: Cross-Sectional Study (e10762)	
Sandra Odenheimer, Deepika Goyal, Veena Jones, Ruth Rosenblum, Lam Ho, Albert Chan.	395
Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study (e231)	
Timothy Guetterman, Tammy Chang, Melissa DeJonckheere, Tanmay Basu, Elizabeth Scruggs, VG Vydiswaran.	403
Web-Based Versus Usual Care and Other Formats of Decision Aids to Support Prostate Cancer Screening Decisions: Systematic Review and Meta-Analysis (e228)	
Sofia Baptista, Elvira Teles Sampaio, Bruno Heleno, Luís Azevedo, Carlos Martins.	416
Impact of a Web-Based Electronic Health Record on Behavioral Health Service Delivery for Children and Adolescents: Randomized Controlled Trial (e10197)	
Eric Bruns, Alyssa Hook, Elizabeth Parker, Isabella Esposito, April Sather, Ryan Parigoris, Aaron Lyon, Kelly Hyde.	430
Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System (e208)	
Seuli Bose-Brill, Michelle Feeney, Laura Prater, Laura Miles, Angela Corbett, Stephen Koesters.	449
Pokémon GO and Physical Activity in Asia: Multilevel Study (e217)	
Ben Ma, Sai Ng, Tim Schwanen, John Zacharias, Mudi Zhou, Ichiro Kawachi, Guibo Sun.	457
Sharing and Empathy in Digital Spaces: Qualitative Study of Online Health Forums for Breast Cancer and Motor Neuron Disease (Amyotrophic Lateral Sclerosis) (e222)	
Sarah Hargreaves, Peter Bath, Suzanne Duffin, Julie Ellis.	469
Detecting Suicidal Ideation on Forums: Proof-of-Concept Study (e215)	
Ahmet Alada , Serra Muderrisoglu, Naz Akbas, Oguzhan Zahmacioglu, Haluk Bingol.	490
Assessing Statewide All-Cause Future One-Year Mortality: Prospective Study With Implications for Quality of Life, Resource Utilization, and Medical Futility (e10311)	
Yanting Guo, Gang Zheng, Tianyun Fu, Shiyang Hao, Chengyin Ye, Le Zheng, Modi Liu, Minjie Xia, Bo Jin, Chunqing Zhu, Oliver Wang, Qian Wu, Devore Culver, Shaun Alfreds, Frank Stearns, Laura Kanov, Ajay Bhatia, Karl Sylvester, Eric Widen, Doff McElhinney, Xuefeng Ling.	501
Interactive Visual Displays for Interpreting the Results of Clinical Trials: Formative Evaluation With Case Vignettes (e10507)	
Jiantao Bian, Charlene Weir, Prasad Unni, Damian Borbolla, Thomas Reese, Yik-Ki Wan, Guilherme Del Fiol.	525
Misleading Claims About Tobacco Products in YouTube Videos: Experimental Effects of Misinformation on Unhealthy Attitudes (e229)	
Dolores Albarracin, Daniel Romer, Christopher Jones, Kathleen Hall Jamieson, Patrick Jamieson.	540
Dissemination of a Web-Based Tool for Supporting Health Insurance Plan Decisions (Show Me Health Plans): Cross-Sectional Observational Study (e209)	
Jingsong Zhao, Nageen Mir, Nicole Ackermann, Kimberly Kaphingst, Mary Politi.	550
A Deep Learning Method to Automatically Identify Reports of Scientifically Rigorous Clinical Research from the Biomedical Literature: Comparative Analytic Study (e10281)	
Guilherme Del Fiol, Matthew Michelson, Alfonso Iorio, Chris Cotoi, R Haynes.	565

Content, Quality, and Assessment Tools of Physician-Rating Websites in 12 Countries: Quantitative Analysis (e212)	
Fabia Rothenfluh, Peter Schulz.	577
Cost-Effectiveness of Facilitated Access to a Self-Management Website, Compared to Usual Care, for Patients With Type 2 Diabetes (HeLP-Diabetes): Randomized Controlled Trial (e201)	
Jinshuo Li, Steve Parrott, Michael Sweeting, Andrew Farmer, Jamie Ross, Charlotte Dack, Kingshuk Pal, Lucy Yardley, Maria Barnard, Mohammed Hudda, Ghadah Alkhaldi, Elizabeth Murray.	591
Evaluation of a Diabetes Self-Management Program: Claims Analysis on Comorbid Illnesses, Health Care Utilization, and Cost (e207)	
Ralph Turner, Qinli Ma, Kate Lorig, Jay Greenberg, Andrea DeVries.	606
Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions (e10148)	
Robert Morris, Kareem Kouddous, Rohan Kshirsagar, Stephen Schueller.	621
Applying the Principles for Digital Development: Case Study of a Smartphone App to Support Collaborative Care for Rural Patients With Posttraumatic Stress Disorder or Bipolar Disorder (e10048)	
Amy Bauer, Sarah Hodsdon, Jared Bechtel, John Fortney.	631
User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews (e10120)	
Katarzyna Stawarz, Chris Preist, Debbie Tallon, Nicola Wiles, David Coyle.	646
Discovery of and Interest in Health Apps Among Those With Mental Health Needs: Survey and Focus Group Study (e10141)	
Stephen Schueller, Martha Neary, Kristen O'Loughlin, Elizabeth Adkins.	662
Detecting Recovery Problems Just in Time: Application of Automated Linguistic Analysis and Supervised Machine Learning to an Online Substance Abuse Forum (e10136)	
Rachel Kornfield, Prathusha Sarma, Dhavan Shah, Fiona McTavish, Gina Landucci, Klaren Pe-Romashko, David Gustafson.	672
Comparing Approaches to Mobile Depression Assessment for Measurement-Based Care: Prospective Study (e10001)	
Amy Bauer, Scott Baldwin, Joaquin Anguera, Patricia Areán, David Atkins.	689
The Burden of a Remote Trial in a Nursing Home Setting: Qualitative Study (e220)	
Susie Donnelly, Brenda Reginatto, Oisín Kearns, Marie Mc Carthy, Bill Byrom, Willie Muehlhausen, Brian Caulfield.	701

Viewpoint

Using Social Media to Target Cancer Prevention in Young Adults: Viewpoint (e203)	
Urmimala Sarkar, Gem Le, Courtney Lyles, Danielle Ramo, Eleni Linos, Kirsten Bibbins-Domingo.	30

Tutorial

Tutorial for Using Control Systems Engineering to Optimize Adaptive Mobile Health Interventions (e214)	
Eric Hekler, Daniel Rivera, Cesar Martin, Sayali Phatak, Mohammad Freigoun, Elizabeth Korinek, Predrag Klasnja, Marc Adams, Matthew Buman.	37

Short Paper

Using a Virtual Reality Social Network During Awake Craniotomy to Map Social Cognition: Prospective Trial ([e10332](#))

Florian Bernard, Jean-Michel Lemée, Ghislaine Aubin, Aram Ter Minassian, Philippe Menei. 560

Review

Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic Review of Reviews

Paolo Zanaboni¹, PhD; Patrice Ngangue², PhD; Gisele Irène Claudine Mbemba³, PhD; Thomas Roger Schopf¹, PhD; Trine Strand Bergmo¹, PhD; Marie-Pierre Gagnon^{3,4}, PhD

¹Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

²Department of Health Sciences, Université du Québec en Abitibi-Témiscamingue, Rouyn Noranda, QC, Canada

³Research Center of the CHU de Québec-Université Laval, Québec, QC, Canada

⁴Faculty of Nursing Sciences, Université Laval, Québec, QC, Canada

Corresponding Author:

Paolo Zanaboni, PhD

Norwegian Centre for E-health Research

University Hospital of North Norway

PO Box 35

Tromsø, 9038

Norway

Phone: 47 91507766

Email: paolo.zanaboni@ehealthresearch.no

Abstract

Background: Digital health can empower citizens to manage their health and address health care system problems including poor access, uncoordinated care and increasing costs. Digital health interventions are typically complex interventions. Therefore, evaluations present methodological challenges.

Objective: The objective of this study was to provide a systematic overview of the methods used to evaluate the effects of internet-based digital health interventions for citizens. Three research questions were addressed to explore methods regarding approaches (study design), effects and indicators.

Methods: We conducted a systematic review of reviews of the methods used to measure the effects of internet-based digital health interventions for citizens. The protocol was developed a priori according to Preferred Reporting Items for Systematic review and Meta-Analysis Protocols and the Cochrane Collaboration methodology for overviews of reviews. Qualitative, mixed-method, and quantitative reviews published in English or French from January 2010 to October 2016 were included. We searched for published reviews in PubMed, EMBASE, The Cochrane Database of Systematic Reviews, CINAHL and Epistemonikos. We categorized the findings based on a thematic analysis of the reviews structured around study designs, indicators, types of interventions, effects and perspectives.

Results: A total of 20 unique reviews were included. The most common digital health interventions for citizens were patient portals and patients' access to electronic health records, covered by 10/20 (50%) and 6/20 (30%) reviews, respectively. Quantitative approaches to study design included observational study (15/20 reviews, 75%), randomized controlled trial (13/20 reviews, 65%), quasi-experimental design (9/20 reviews, 45%), and pre-post studies (6/20 reviews, 30%). Qualitative studies or mixed methods were reported in 13/20 (65%) reviews. Five main categories of effects were identified: (1) health and clinical outcomes, (2) psychological and behavioral outcomes, (3) health care utilization, (4) system adoption and use, and (5) system attributes. Health and clinical outcomes were measured with both general indicators and disease-specific indicators and reported in 11/20 (55%) reviews. Patient-provider communication and patient satisfaction were the most investigated psychological and behavioral outcomes, reported in 13/20 (65%) and 12/20 (60%) reviews, respectively. Evaluation of health care utilization was included in 8/20 (40%) reviews, most of which focused on the economic effects on the health care system.

Conclusions: Although observational studies and surveys have provided evidence of benefits and satisfaction for patients, there is still little reliable evidence from randomized controlled trials of improved health outcomes. Future evaluations of digital health interventions for citizens should focus on specific populations or chronic conditions which are more likely to achieve clinically meaningful benefits and use high-quality approaches such as randomized controlled trials. Implementation research methods should also be considered. We identified a wide range of effects and indicators, most of which focused on patients as main end

users. Implications for providers and the health system should also be included in evaluations or monitoring of digital health interventions.

(*J Med Internet Res* 2018;20(6):e10202) doi:[10.2196/10202](https://doi.org/10.2196/10202)

KEYWORDS

review; electronic health records; patient access to records; patient portals; epidemiological methods; evaluation studies as topic

Introduction

Background

Digital health is defined as the use of digital technologies to provide practical, cost-effective, safe, and scalable interventions to improve health [1], health care services, and wellness for individuals and across populations [2]. Today, numerous types of digital health interventions are available to citizens, patients, carers and the public. They are used to address health system problems including poor access, uncoordinated care and increasingly costly health care [3]. Patient portals [4], mobile health applications [5] and patients' access to electronic health records (EHR) are common examples of digital health interventions. Interventions can also include other online platforms [6] such as medication refills, appointment scheduling, access to general medical information, or secure messaging between a patient and an institution [7]. Digital health interventions can disseminate information, aid informed decision making, and promote health. They also provide a means for information exchange and support, and manage demand for health services, lowering direct medical costs [8].

Recently, there has been an increasing public interest in digital health [9]. Digital health interventions can empower citizens to track, manage, and improve their health and quality of life while providing a more personalized health care delivery, at a lower cost and with higher efficiency and availability [10]. Digital health has also enabled unprecedented patient engagement in self-management and well-being [2]. Patients seek information from the internet to learn more about their symptoms, diagnoses, and treatments and use a range of digital health interventions to manage their illness at home and support independent living and self-care [11].

Digital health interventions have enormous potential as scalable tools to support better health and health care delivery by improving many different outcomes such as effectiveness, efficiency, accessibility, safety, and personalization [1]. Although evidence of the potential benefit of digital health for improving care delivery and patient outcomes has been described [9], numerous factors can affect patient and public engagement in using digital health interventions such as lack of motivation, busy lifestyle, poor digital literacy, complexity and usability [3]. Other difficulties include the rapid change of technology, which requires digital health interventions to evolve and be constantly updated [1].

Digital health interventions are typically complex with multiple components, and many have multiple aims. As a consequence, evaluations of digital health interventions present unique methodological challenges [1]. A variety of study designs have been implemented in practice to evaluate digital health

interventions. Much of the evidence has been generated through quantitative methods, such as pilot studies or clinical trials, although the number of qualitative studies is also increasing [3]. Randomized controlled trials (RCTs) are considered the gold standard in evaluating health care interventions [12]. While RCTs are more likely to be used to assess new treatments or medicines, their applicability to evaluate the complex, multifaceted nature of digital health interventions has been widely debated [13]. RCTs have predefined protocols and strict inclusion criteria that can often mask wider implementation issues [14]. Many challenges only emerge when technologies are scaled up and implemented in 'real-world' complex health systems [15]. When evaluating a digital health intervention, it is essential to identify the likely benefits, define the causal model describing how the intervention will achieve its intended benefits, and broaden the portfolio of evaluation methods [1].

Researchers need to support the public, patients, clinicians, and policy-makers by creating an actionable knowledge base to identify the effects of digital health. Specific frameworks for evaluating digital health interventions have been recently developed to generate evidence required for decision-making on the appropriate approach to integrate effective strategies into broader national health systems [16]. An example is the Canada Health Infoway Benefits Evaluation Framework based on dimensions of quality, system usage and net benefits, as well as specific indicators [17]. Careful monitoring and systematic evaluations of digital health interventions, however, have been few, in contrast to the proliferation of digital health pilot projects [16]. As a consequence, the current research evidence on which methods should be used to evaluate digital health interventions is still fragmented [3].

The objective of the current study was to explore and provide a systematic overview of the methods used to evaluate the effects of internet-based digital health interventions for citizens. Internet-based digital health interventions covered by this study included patients' access to EHR, patient portals, and other internet-based support programs. Citizens are referred to as the general population, including both healthy individuals and patients having access to health care services.

The following research questions were addressed:

1. Which approaches (study design) have been used to produce knowledge about the effects of internet-based digital health interventions for citizens?
2. Which effects have been measured and reported in studies focusing on internet-based digital health interventions for citizens?
3. Which indicators were used to measure the effects of internet-based digital health interventions for citizens, and

for whom were the effects measured (patients, health care system, society)?

Methods

Study Design

A systematic review of reviews [18] focused on the methods used to measure the effects of internet-based digital health interventions for citizens was conducted in accordance with the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P) [19]. This includes a checklist of recommended items to be addressed in a systematic review protocol [20]. The protocol was developed a priori according to the International Prospective Register of Systematic Reviews (PROSPERO) template and is available upon request. We followed the Cochrane Collaboration methodology for overviews of reviews [21]. The scope of the review and eligibility criteria were formulated using the PICOS approach (participants, interventions, comparisons, outcomes, study designs) [21]. The Assessing the Methodological Quality of Systematic Reviews (AMSTAR) checklist was used to assess the quality of the included reviews [22].

Inclusion and Exclusion Criteria

The systematic review of reviews included qualitative, mixed-method, and quantitative studies published in English or French from January 2010 to October 2016. Review papers were eligible if (1) the primary end-user was the patient or carer or citizen, (2) they were related to internet-based digital health interventions for citizens where there was a direct form of interaction with health care providers, and (3) they evaluated the impact of implementing or using internet-based digital health interventions for citizens. Internet-based digital health interventions for citizens included patients' access to EHR, patient portals, and other internet-based support programs. An EHR is the electronic collection of clinical data relating to one subject of care, including clinical assessments, laboratory results, radiology findings, nursing documentation, allergy information, medication information and discharge letters. Health care organizations can provide online EHR access to patients, relatives or other informal carers. Besides health care organizations, EHR access may also be offered on a national scale [7]. Electronic patient portals are defined as electronic applications (typically web-based) provided and maintained by health care institutions which can offer access to (a subset of) clinical EHR data as well as additional services, including medication refills, appointment scheduling, access to general medical information such as guidelines, or secure messaging [7]. Internet-based support programs refer to interventions, such as social support groups, online therapy for psychosocial or physical symptoms, online systems integrating information, support and coaching services [23], which can promote collaboration and help individuals with chronic conditions. Review papers were excluded if (1) they were related to interventions used solely by health care professionals (eg, clinical decision support systems), (2) they were related to interventions designed for patients without direct interaction with health care providers (eg, mHealth, self-management tools, educational platforms), and (3) they were focused on patient

access to health records which were not digital (eg, paper records).

Data Sources and Search Strategy

We searched for reviews published in the following electronic bibliographic databases: PubMed, EMBASE, The Cochrane Database of Systematic Reviews, CINHAL and Epistemonikos. A structured search strategy was developed using the thesaurus terms of each database and using some keywords included in the titles and abstracts of the reviews. The search strategy included terms relating to or describing internet-based digital health interventions for citizens ([Multimedia Appendix 1](#)). The results of each database search were stored in a single reference database (Endnote). Duplicate references were removed. The electronic search on the mentioned databases was performed by one research team member (PN).

Study Selection

Titles and abstracts of the review papers retrieved using the search strategy were screened by two reviewers (PN, GM). Studies that did not meet the inclusion criteria were excluded. The full texts of the selected studies were then retrieved and independently assessed for eligibility by a review team consisting of six members (PZ, TRS, TB, MPG, PN, GM). Any disagreement over the eligibility of particular studies was resolved through discussion and the involvement of another reviewer if necessary.

Data Extraction

A standardized data extraction form was developed, piloted and used to extract data from the full text of the included reviews for evidence synthesis ([Multimedia Appendix 2](#)). Extracted data used to categorize review papers included first author, year of publication, language, type of review, rationale, objectives, eligibility criteria, and the fields of the AMSTAR checklist. Additional information extracted from each review paper included: study selection, interventions, populations, settings, effects measured (types of outcome, perspectives, indicators), study design, and main findings.

Three papers were chosen to pilot the data extraction process and form. All the review team members reviewed them. A meeting was organized around data extraction to make sure that all review team members had the same understanding of the information to extract.

The full texts of the included reviews were equally and randomly divided among the review members to minimize bias. For each paper, data were extracted systematically and reported on the data extraction form. Papers that did not meet the inclusion criteria were pointed out along with the reasons for exclusion. Another member then cross-checked all the papers reviewed by one member in order to agree on the selection.

Quality Assessment

The AMSTAR checklist was used to assess the quality of the included studies [22]. AMSTAR is an 11-item checklist from which reviewers assign 1 point when 1 criterion is met. This tool characterizes the quality of a systematic review at 3 levels: a score 8-11 is considered high quality (ie, minor or no methodological limitations), a score 4-7 is medium quality (ie,

moderate methodological limitations), while a score 0-3 is low quality (ie, significant methodological limitations). Because this study was aimed at describing and synthesizing a body of both quantitative and qualitative literature, and not determining an effect size, no additional methods for risk of bias were conducted as they would not have affected the interpretative synthesis of the findings.

Data Synthesis

We conducted a thematic analysis of the outcomes of the included papers. Thematic analysis is the most common method adopted within narrative reviews to produce a synthesis of findings arising from a body of literature. It seeks to identify systematically emerging conceptual themes across multiple studies. The themes identified are shaped by the specific review questions [24]. The findings were analysed and structured around the study designs and indicators used to measure the effects of different interventions, the types of interventions, the impact of interventions and the perspectives for whom the outcomes were measured to answer the three research questions. Additional information was also extracted and analyzed to describe the evidence and direction of the effects reported in the included papers. The results were summarized in the form of a textual, narrative understanding of the findings supported by tabular summaries.

Results

Search Results

A total of 2,054 papers were identified from the search strategy. After removing duplicates and initial screening of titles and

abstracts, 42 articles were retrieved for detailed evaluation. Following further inspection of their full-texts, 23 papers met the eligibility criteria (Multimedia Appendix 3), while the remaining 19 articles were excluded (Multimedia Appendix 4). Of the 23 papers included, there were 3 publications [25,26,27] which were part of another main study [7,28,29], leading to a total of 20 unique reviews [4,7,23,28-44]. The overall process of review selection was summarized according to the PRISMA study flow diagram (Figure 1).

Review Characteristics

The general characteristics of the studies, including the type of review [45], population, intervention, setting, and quality, were summarized by the information extracted from the analysis of the full-text articles (Table 1). There were 14/20 (70%) systematic reviews and 1/20 (5%) systematic review which also included a meta-analysis. Among the remaining reviews, 3/20 (15%) were literature, 1/20 (5%) was realist, and 1/20 (5%) was narrative. The most common internet-based digital health interventions for citizens included in the present study were patient portals and patients' access to their EHR, covered by 10/20 (50%) and 6/20 (30%) reviews, respectively. Internet-based support programs were described in 4/20 (20%) studies. A total of 8/20 (40%) did not have any restriction regarding the study settings. Where the setting was specified, 6/20 (30%) reviews were focused on primary care, while the remaining (6/20, 30%) included outpatient and inpatient settings. According to the AMSTAR score 13/20 (65%) reviews were medium quality, while 6/20 (30%) were high quality. Only 1/20 (5%) review scored low quality. Systematic reviews had on average a higher quality than other types.

Figure 1. Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) study flow diagram.

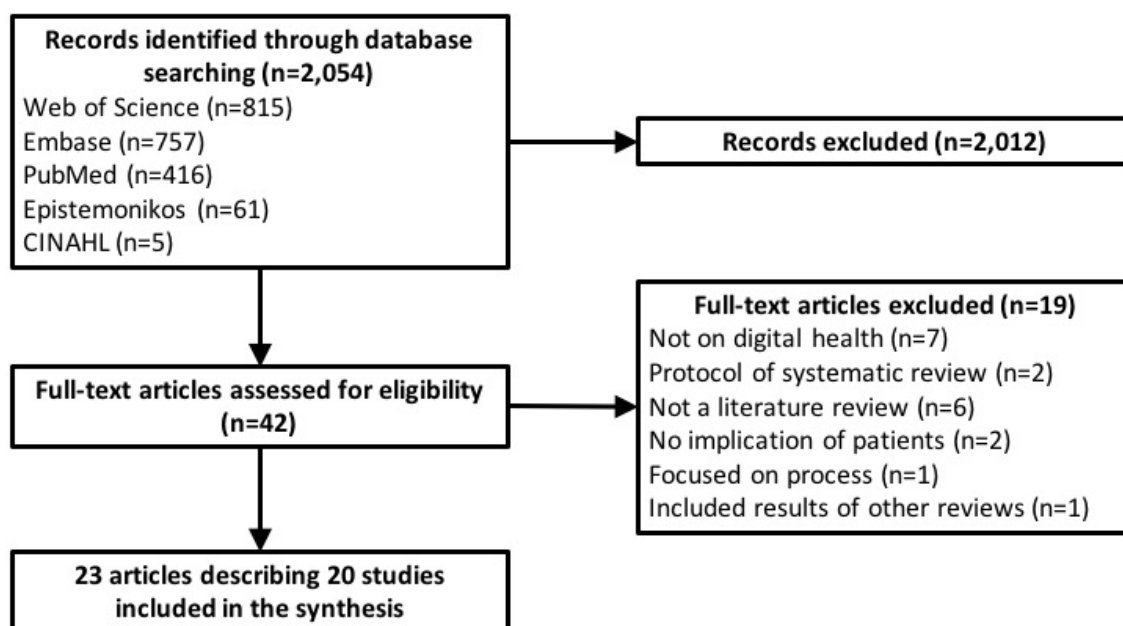


Table 1. Characteristics of included studies.

Reference	Year	Studies, n	Type of review	Population	Intervention	Setting	AMSTAR ^a
Alkureishi et al [30]	2016	53	Systematic review	Adult and paediatric patients	Patients' access to EHR ^b	Outpatient and inpatient	7
Amante et al [31]	2014	16	Systematic review	Patients with diabetes	Patient portals	Primary care	8
Ammenwerth et al [7]	2012	5	Systematic review	Patients	Patient portals	Outpatient and inpatient	8
Bouma et al [23]	2015	16	Literature review	Cancer patients	Internet-based support programs	No restrictions	8
Bush et al [32]	2016	31	Systematic review	Paediatric patients	Patient portals	No restrictions	4
Davis Giardina et al [33]	2014	27	Systematic review	Patients	Patients' access to EHR	No restrictions	5
Davis et al [34]	2014	16	Systematic review	Adult patients	Internet-based support programs	Primary care	6
Goldzweig et al [35]	2013	46	Systematic review	Patients	Patient portals	Primary care	8
Irizarry et al [4]	2014	120	Literature review	Patients	Patient portals	No restrictions	6
Kruse et al [29]	2015	27	Systematic review	Patients with chronic conditions	Patient portals	Outpatient and inpatient	4
Liu et al [36]	2013	8	Systematic review	Patients	Patients' access to EHR	Primary care, emergency, outpatient	6
Mold et al [28]	2015	17	Systematic review	Patients	Patients' access to EHR	Primary care	9
Osborn et al [37]	2010	26	Systematic review	Patients with diabetes	Patient portals	Primary care	6
Otte-Trojel et al [38]	2014	32	Realist review	Patients	Patient portals	Outpatient and inpatient	4
Price et al [39]	2015	23	Systematic review	Patients with chronic conditions	Patients' access to EHR	Outpatient	5
Stellefson et al [40]	2013	15	Systematic review	Patients with chronic conditions	Internet-based support programs	No restrictions	7
Tao and Or [41]	2013	36	Meta-analysis	Patients with diabetes	Internet-based support programs	No restrictions	10
Tulu et al [42]	2016	23	Literature review	Patients with pulmonary conditions	Patient portals	Pulmonary practice	4
Turner et al [43]	2016	12	Narrative review	Patients with HIV	Patients' access to EHR	No restrictions	3
Vimalananda et al [44]	2015	27	Systematic review	Patients, specialty care	Patient portals	No restrictions	6

^aAMSTAR: Assessing the Methodological Quality of Systematic Reviews. Studies are classified as high (scoring 8-11), medium (4-7), or low quality (0-3).

^bEHR: electronic health record.

Overview of Research Methods

Extracted data from the studies included in this review were then analysed to address the first research question by providing an overview of the types of study design used to measure the effects of different internet-based digital health interventions for citizens (Table 2). Each review could be focused on one type of study only (eg, RCT) or include different study designs. Overall, 13/20 (65%) reviews summarised results from both quantitative and qualitative studies (including mixed methods), while the remaining 7/20 (35%) reviews contained only studies using quantitative methods (RCTs, quasi-experimental studies with control, cohort, pre-post studies, retrospective studies, cross-sectional studies, and surveys).

A total of 13/20 (65%) of the included reviews reported the use of RCTs to evaluate patients' access to EHR, patient portals, and internet-based support programs. Quasi-experimental designs, in form of non-randomized controlled trials where subjects are allocated to intervention and control groups without a randomization method, were described in 9/20 (45%) reviews. Another method, which does not imply randomization and does not necessarily require a control group, is a pre-post study (or before-after study). Pre-post studies measure a specific outcome before and after an intervention. However, due to the lack of a control group, this study design is considered weak as it is difficult to conclude whether changes occurred due to the intervention or would have occurred anyway. Pre-post studies were reported only in 6/20 (30%) reviews.

Table 2. Overview of research methods. No data is shown as N/A (not applicable).

Study design	Interventions		
	Patients' access to EHR ^a	Patient portals	Internet-based support programs
RCT ^b	[28,33,36,39]	[4,7,31,35,37,38]	[23,40,41]
Quasi experimental with control	[28,43]	[4,29,35,37,44]	[23,40]
Pre-post	[29,33,36]	[35,37,44]	N/A ^c
Cohort	[28,33,39]	[4,31,38,44]	N/A
Retrospective	[33,39]	[29,33,44]	N/A
Cross-sectional or surveys	[28,30,33,36,39,43]	[4,29,31,32,35,42,44]	[34,40]
Qualitative	[30,43]	[4,29,31,32,35,37,38,42,44]	[34,40]
Mixed methods	[30,43]	[29,31,37,38]	[34]
Other (pilot study, simulation, usability)	N/A	[4,32,37,44]	N/A

^aEHR: electronic health record.

^bRCT: randomized controlled trial.

^cN/A: not applicable.

As an alternative to experimental and quasi-experimental designs, 15/20 (75%) reviews reported using observational studies to evaluate internet-based digital health interventions for citizens. These include prospective cohort studies, retrospective studies, cross-sectional studies, and surveys. Observational studies were reported for all the types of digital health interventions included in this review.

Finally, 13/20 (65%) reviews included evaluation of internet-based digital health interventions through qualitative studies or mixed methods. Use of qualitative methods was described in all the types of digital health interventions covered by this review. Some (4/20, 20%) reviews also referred to the use of other research methods including pilot studies, simulation or usability testing.

Overview of Effects and Indicators

To address the second and third research questions, we extracted and analyzed study data regarding which effects were measured and reported, which methods (regarding indicators) were used to measure the effects, and for whom they were measured (patients, providers, health care system, or society). These data are summarized in [Table 3](#).

A large number of effects were measured when evaluating internet-based digital health interventions for citizens. These were classified into five main categories (1) health and clinical outcomes, (2) psychological and behavioral outcomes, (3) health care utilization, (4) system adoption and use, and (5) system attributes. The first two categories are mainly related to effects perceived by patients as main end users. While there was limited evidence of the clinical benefits for patients resulting from the implementation of internet-based digital health interventions, such as access to their EHR or patient portals, health and behavioral outcomes were often included as an object of evaluation. Health care utilization refers to a range of indicators

which can affect both patients and the health care system in general. Adoption and use is another critical category related to the users of digital health interventions, namely patients and providers. Finally, system attributes refer to other effects focusing on the evaluation of the systems themselves, which can impact on patients, providers, and the health system in general.

Health and clinical outcomes were measured with both general indicators (eg, improvement in health status, quality of life, medication management, mortality, physical activity) and disease-specific indicators (eg, related to diabetes or hypertension). Overall, there were 11/20 (55%) studies which reported health and clinical outcomes.

Psychological and behavioral outcomes include a wide variety of indicators reflecting the impact on patients in changing their behavior towards the way they manage their health or a specific disease. Patient-provider communication was by far the most investigated indicator. Internet-based digital health interventions are claimed to impact on the quality of the communication between patients and health providers, and 13/20 (65%) studies examined this effect. Satisfaction was another widely used indicator. Patient satisfaction was documented in 12/20 (60%) reviews, while only 1/20 (5%) review additionally reported satisfaction for providers. Self-efficacy and self-management represent other important indicators which were included in 5/20 (25%) reviews. However, similarly to clinical outcomes, evidence of the effects on self-efficacy is limited. Other psychological and behavioral outcomes included: adherence to therapy, potential harms, perceived benefits, and perceived social support. Specific indicators measuring a change in the role of patients towards health services were: improved access to information, attitudes, empowerment, acceptance, and endorsement.

Table 3. Overview of effects and indicators. No data is shown as N/A (not applicable).

Outcomes and indicators	Perspective			
	Patients	Providers	Health system	Society
Health and clinical outcomes				
Health status	[29,35,37,38,40,44]	N/A ^a	N/A	N/A
Quality of life	[23,39,40]	N/A	N/A	N/A
Safety or medication management	[28,39]	N/A	N/A	N/A
Disease-specific measures	[32,39,41]	N/A	N/A	N/A
Mortality or risk factors	[7]	N/A	N/A	N/A
Physical activity or nutrition outcomes	[40]	N/A	N/A	N/A
Psychological and behavioral outcomes				
Self-management or self-efficacy	[33,39,40,42,43]	N/A	N/A	N/A
Satisfaction	[7,28-30,33,35-39,44]	[34]	N/A	N/A
Patient activation	[39]	N/A	N/A	N/A
Patient-provider communication	[4,7,28-30,33,37,38,42,43]	[4,30,37]	[31,34]	N/A
Patient access to information	[39]	N/A	N/A	N/A
Acceptance or endorsement	N/A	[4,44]	N/A	N/A
Health literacy	[4,37]	N/A	N/A	N/A
Awareness and knowledge	[39,43]	N/A	N/A	N/A
Perceived benefits	[28,33,43]	N/A	N/A	N/A
Concerns (privacy, security)	[28,43]	N/A	N/A	N/A
Perceived social support	[23,40]	N/A	N/A	N/A
Adherence to treatment	[7,31,33,35,38,43]	N/A	N/A	N/A
Empowerment	[7,29,31,37,38]	N/A	N/A	N/A
Attitudes	[35]	[4]	N/A	N/A
Harms (distress, stress, anxiety)	[23,33,39,40]	N/A	N/A	N/A
Health care utilization				
Outpatient or clinic visits	[28]	N/A	[7,33,38,39,44]	N/A
Access or wait time	[28,42,44]	N/A	N/A	N/A
Hospitalization rate or urgent care utilization	[7]	N/A	[7,29,33,40]	N/A
System adoption and use				
Patient adoption	[4,32]	N/A	N/A	N/A
Professional practice	N/A	[31,34]	N/A	N/A
Patient utilization	[33,40]	N/A	N/A	N/A
System attributes				
Usability	[4,32,37,38,42,43]	[4,38]	N/A	N/A
Utility	[4]	N/A	N/A	N/A
Personalization	[4]	N/A	N/A	N/A
Efficiency	[35]	N/A	[34]	N/A

^aN/A: not applicable.

Health care utilization refers to the impact of digital health interventions on the resources involved, including time used by patients and providers and use of the health care system (eg, hospitalizations and outpatient care). Evaluation of health care

utilization was included in 8/20 (40%) reviews, most of which focused on the economic effects on the health care system.

System adoption and use was measured through three indicators, two of which were defined from the patient's perspective.

Patient adoption and patient utilization refer to how patients decide to use a digital health intervention and to what extent this is used in practice. Professional practice indicates the degree to which a digital health service implies organizational changes for health care professionals. Each of these indicators was covered by 2/20 (10%) reviews.

Other indicators were also used to measure different system attributes. Of these, the most relevant was usability for patients and providers, which was included in 6/20 (30%) reviews. Each of the indicators was measured with specific tools, including data recorded from patients, questionnaires, data collected from databases or registries, and interviews.

Synthesis of the Evidence of Benefits

We finally summarized the main findings regarding the outcomes of different internet-based digital health interventions included in this review to provide a narrative description of the evidence of benefits. A total of 3/6 (50%) reviews focusing on the effects of providing patients access to their EHR found improved levels of patient satisfaction [28,33,36]. However, evidence was less clear for effects on health care quality, including measures of safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity [33]. There have been only a few RCTs examining the effectiveness of online access to EHR on improving health outcomes, but some work has been undertaken focusing on the impact on patient decision making and health outcomes [28]. Patients reported some evidence of improvements in safety through identifying medication errors, self-care, communication and engagement with clinicians [28]. There were no reports of harm to patients, such as increased anxiety, a common fear endorsed by physicians [33], nor breaches in privacy [28]. However, some participants were concerned about privacy and security [43]. Effects on workload and system efficiency were unclear [28,33]. Evidence of benefits, including quality of care, access, productivity, self-management, associated with the use of EHR by citizens were found for patients with specific chronic health conditions [39]. In general, low use was associated with socio-demographic factors [43].

Evidence that patient portals improve health outcomes was mixed [35]. There were 3/10 (30%) reviews reporting positive effects of patient portals on better adherence to treatment and medication regimen [7,29,38]. However, overall evidence of improvement in quality of care was limited [7], and there was a lack of studies investigating long-term health outcomes [32]. As most of the patient portal evaluations targeted online interventions for chronic disease management (eg, diabetes, hypertension, depression, chronic musculoskeletal pain, or mobility difficulty), evidence of benefits was found for disease awareness and self-efficacy [29,37,41], empowerment [29,38], patient-provider communication and social support [37,38,40]. There was some evidence that patient portals impact on costs or utilization [35], including a decrease in office visits [7,29]. However, there were also some studies which found a higher health resource use [38]. In general, there was a high satisfaction and acceptance by users of patient portals [29,32,37,38,42]. Facilitators to adoption and use of patient portals were encouragement and assistance from providers and family

members as well as self-engagement, while barriers included lack of skills, desire, or knowledge, technical difficulties, and lack of potential benefits [31]. Patient engagement and use of patient portals was strongly influenced by personal factors (ethnicity, education level, health literacy, health status) and health care delivery factors (provider endorsement and usability) [4]. Patient portals were mainly used by patients with chronic conditions [32]. No studies found serious adverse consequences [38].

Similarly to patient portals, other internet-based support programs for patients with chronic conditions found evidence of benefits regarding increased self-efficacy and better communication with the health care providers [40]. There was 1/4 (25%) review focused on internet support programs for cancer patients that found some positive results on quality of life and social support, as well as on disease-specific indicators including cancer-related fatigue, insomnia, and stress [23]. A meta-analysis of RCTs showed that the use of health information technology was associated with an improved glycaemic control in patients with diabetes [41].

Discussion

Principal Findings

This systematic review provides an overview of the types of study design and methods used to measure the effects of internet-based digital health interventions for citizens which are reported in other reviews. We found 20 relevant reviews published since 2010, indicating a generally growing interest in the evaluation of the effects of digital health interventions for citizens such as patients' access to their EHR and patient portals. The overall quality of the included reviews was also good, with many high-quality studies and only a few studies with methodological limitations. Non-systematic reviews generally had a lower quality score on the AMSTAR scale because this tool was specifically designed to assess systematic reviews of quantitative studies, and there is no equivalent tool for other types of reviews [46].

Although there is some evidence of benefits from observational studies and surveys, with many studies reporting value in patients having access to more information through internet-based digital health interventions, there is still little reliable evidence from experimental studies of proven effectiveness in improved patient health outcomes [47]. One reason for this may be that not all health conditions are sensitive to patients' access to EHR as an intervention [39]. Conditions with evidence of clinical benefits for patients accessing EHR include chronic diseases (eg, diabetes) with an aspect of monitoring, either by the clinician or the patient (self-monitoring). RCTs are needed to test assumptions about the comparative effectiveness on outcomes for various patient populations [47]. However, their applicability to evaluate the complex multifaceted nature of digital health interventions has been widely debated [13]. RCTs remain an essential method for determining the impact of digital health interventions concerning efficacy and cost-effectiveness but are best undertaken when the services are highly likely to lead to clinically meaningful benefits [1]. As a consequence, patient

portals, online services or internet support programs specifically designed for chronic disease management seem to be more suitable to the use of RCTs as a research method to measure the effects on clinical outcomes. In case of digital health interventions offered to citizens, RCTs might also be useful to measure the impact on those sub-populations of people who are more likely to obtain improved outcomes.

RCTs are also best undertaken once the services are stable and can be implemented with high fidelity [1]. However, digital health interventions are typically complex interventions with multiple components and multiple aims [1]. As a consequence, evaluations present unique methodological challenges. The successful development, integration, and implementation of digital health interventions require a radical shift from traditional, and single-disciplinary academic and clinical approaches [2]. In this respect, the Normalisation Process Theory addresses the factors needed for successful implementation and integration of interventions into routine work, enabling researchers to think through issues of implementation while designing a complex intervention and its evaluation. [48]. Implementation research is another growing field which seeks to understand and work within real-world conditions, rather than trying to control for these conditions or to remove their influence as causal effects [49]. Implementation research provides a framework for using the research question as the basis for selecting among a wide range of qualitative and quantitative methods. Implementation specific research methods include non-traditional studies such as pragmatic trials, effectiveness-implementation hybrid trials, quality improvement studies, participatory action research, and mixed methods [49]. The types of study design used to evaluate internet-based digital health interventions were experimental studies, observational studies, and qualitative studies. We found only a few mixed methods studies in our review, but no other types of implementation research methods were reported. Future evaluations of internet-based digital health interventions should be appropriately designed, and where RCTs are not appropriate, make more use of implementation research methods.

The studies included in this review allowed identifying a wide range of effects and indicators, which can affect patients, providers and the health system in general. These methods can be used in evaluations of new internet-based digital health interventions implemented in the future or to monitor them over time after their introduction. Most indicators were focused on measuring direct impact on the patients as main end users. However, evaluations can also benefit from including indicators used to measure the effects for providers (such as acceptance, communication, and usability) and the health system (such as health care utilization). There is insufficient use of indicators that measure the impact on society, indicating that societal effects have minor importance, or are simply more difficult to measure.

The high number of studies measuring satisfaction and acceptance by users suggests that internet-based digital health interventions might have an important impact on patients in changing their behavior towards the way they manage their health or a specific disease. One reason might be that digital health interventions designed and implemented to improve

health services for citizens/patients are more likely to have an impact on their level of satisfaction. Another reason might be that satisfaction is an indicator relatively easy to measure and monitor over time.

Measuring system adoption and use are vital to understand how patients and providers decide to adopt internet-based digital health interventions and implement them in practice. These indicators are relevant to monitor the overall impact of digital health interventions (ie, scalability) over time. People with serious chronic conditions, individuals with disabilities, parents with small children, people with a keen interest in maintaining healthy lifestyles, and the elderly or their caregivers seem more likely to adopt internet-based digital health interventions [47]. System usability for both patients and providers can in turn impact on satisfaction and use and must, therefore, be included in evaluation and monitoring activities. Patients' interest and ability to use digital health interventions is strongly influenced by personal factors such as age, ethnicity, education level, health literacy, health status, and role as a caregiver [35,37,43]. Future research should focus on identifying specific characteristics associated with a higher degree of patient engagement [4].

Evidence of clinical outcomes was still unclear, for both patients' access to EHR and patient portals. There were, however, some positive results for improved health status and better medication management. Several studies included measurement of psychological and behavioral indicators, with evidence that internet-based digital health interventions can improve self-efficacy. Effects were more significant for patients with specific chronic health conditions. Additional research is needed to identify which features are most influential in changing health behaviors [43].

Effects on workload and system efficiency were also unclear, with some studies reporting savings, while other studies reported increased health service use when using patient portals. This suggests that patient portals might be seen as complements rather than substitutes to existing health services [38]. The inclusion of cost-effectiveness evaluations, or simpler indicators to evaluate the impact of internet-based digital health interventions on health care utilization, is necessary as the economic impact can represent an incentive for both patients and providers to use the systems.

Study Strengths and Limitations

The findings of the current study emerge from the analysis of reviews focused on the evaluation of the effects of different internet-based digital health interventions for citizens, rather than from the analysis of individual studies. Moreover, the results were summarised in the form of a thematic analysis since a meta-analysis was not possible due to the heterogeneity of the included reviews. However, the objective of this study was to explore and provide a systematic overview of the methods used to evaluate the effects of internet-based digital health interventions for citizens. That is, we did not intend to provide a quantitative synthesis of the effects. As such, we chose to conduct a systematic review of reviews which allowed us to answer our research questions. An a priori protocol was developed per international quality standards including the Cochrane Collaboration methodology for overviews of reviews,

PRISMA-P and PROSPERO templates. The scope and eligibility criteria were formulated using the PICOS approach. Moreover, most of the reviews included in this review were high-quality studies according to the AMSTAR checklist. Although our research did not aim at quantifying the outcomes reported in the included papers, we were still able to describe the primary evidence and direction of the effects.

Another limitation of this study was that it was not focused on a specific intervention. On the contrary, it aimed to gather information on which methods were used in the evaluation of digital health interventions. As a consequence, we purposefully chose to include reviews which were focused on a wide range of internet-based digital health interventions. This allowed us to provide a complete picture of which methods were used across different services and identify similarities or specificities. Moreover, mobile health (mHealth) interventions were beyond the scope of this study and thus not explicitly covered in the review. The term mHealth refers to the use of mobile and wireless technologies for health [50] and includes interventions that are delivered through mobile devices or the new generation of tablet computers [51].

The study protocol was designed to include only studies published from 2010. Reviews published earlier were therefore excluded. Digital health is a relatively recent field, and the majority of reviews focusing on digital health interventions has been published over the past few years. As we might have missed a few studies published before 2010, those studies are likely to be included in recent reviews. Moreover, as digital health technologies are always in rapid change, it appeared to

be relevant to focus on studies which were recently conducted. The terms included in the search strategy might have also been restrictive when referring to citizens and patients. Despite these limitations, more than 2,000 studies were screened, and a total of 23 publications involving 20 unique reviews were included in this review.

Conclusions

We found many relevant reviews, indicating a generally growing interest in the evaluation of the effects of internet-based digital health interventions such as patients' access to EHR and patient portals. Although there is some evidence of benefits and satisfaction for patients from observational studies and surveys, there is still little reliable evidence from RCTs or other experimental studies of proven effectiveness in improved patient health outcomes through the use of digital health interventions. The results of this review show that internet-based digital health interventions for citizens have a higher clinical impact on chronic disease management. Future evaluation studies focused on clinical outcomes should possibly focus on specific populations or chronic conditions which are more likely to achieve clinically meaningful benefits and use high-quality methods for study design, such as RCTs. Researchers should think through issues of implementation while designing and evaluating complex digital health interventions. Moreover, non-traditional approaches such as implementation research methods should be considered as valuable alternatives when evaluating internet-based digital health interventions. Additional research is also needed to identify which personal health record features are most influential in changing health behaviors.

Acknowledgments

This study was funded by the Norwegian Centre for E-health Research as part of the project "Effekter av digitale innbyggertjenester". We thank Monika Johansen for the fruitful discussion and suggestions provided during the development of the a-priori protocol for this review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[[PDF File \(Adobe PDF File\), 35KB - jmir_v20i6e10202_app1.pdf](#)]

Multimedia Appendix 2

Data extraction form.

[[PDF File \(Adobe PDF File\), 35KB - jmir_v20i6e10202_app2.pdf](#)]

Multimedia Appendix 3

Studies included.

[[PDF File \(Adobe PDF File\), 80KB - jmir_v20i6e10202_app3.pdf](#)]

Multimedia Appendix 4

Studies excluded.

[PDF File (Adobe PDF File), 84KB - [jmir_v20i6e10202_app4.pdf](#)]

References

1. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating Digital Health Interventions: Key Questions and Approaches. *Am J Prev Med* 2016 Nov;51(5):843-851. [doi: [10.1016/j.amepre.2016.06.008](#)] [Medline: [27745684](#)]
2. Kostkova P. Grand challenges in digital health. *Front Public Health* 2015;3:134 [FREE Full text] [doi: [10.3389/fpubh.2015.00134](#)] [Medline: [26000272](#)]
3. 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](#)]
4. 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](#)]
5. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS Med* 2013;10(1):e1001362 [FREE Full text] [doi: [10.1371/journal.pmed.1001362](#)] [Medline: [23349621](#)]
6. Mold F, Ellis B, de LS, Sheikh A, Wyatt JC, Cavill M, et al. The provision and impact of online patient access to their electronic health records (EHR) and transactional services on the quality and safety of health care: systematic review protocol. *Inform Prim Care* 2012;20(4):271-282 [FREE Full text] [Medline: [23890339](#)]
7. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. *J Med Internet Res* 2012;14(6):e162 [FREE Full text] [doi: [10.2196/jmir.2238](#)] [Medline: [23183044](#)]
8. Frank SR. Digital health care--the convergence of health care and the Internet. *J Ambul Care Manage* 2000 Apr;23(2):8-17. [Medline: [10848396](#)]
9. Agboola SO, Bates DW, Kvedar JC. Digital Health and Patient Safety. *JAMA* 2016 Apr 26;315(16):1697-1698. [doi: [10.1001/jama.2016.2402](#)] [Medline: [27115372](#)]
10. Garell C, Svedberg P, Nygren JM. A Legal Framework to Support Development and Assessment of Digital Health Services. *JMIR Med Inform* 2016 May 25;4(2):e17 [FREE Full text] [doi: [10.2196/medinform.5401](#)] [Medline: [27226391](#)]
11. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. *J Am Med Inform Assoc* 2016 Jan;23(1):5-11 [FREE Full text] [doi: [10.1093/jamia/ocv099](#)] [Medline: [26377990](#)]
12. Lau F, Kuziemsky C. *Handbook of eHealth Evaluation: An Evidence-based Approach*. Victoria, British Columbia, Canada: University of Victoria; 2016.
13. Kaplan B. Evaluating informatics applications--some alternative approaches: theory, social interactionism, and call for methodological pluralism. *Int J Med Inform* 2001 Nov;64(1):39-56. [Medline: [11673101](#)]
14. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. *Bull World Health Organ* 2012 May 1;90(5):357-364 [FREE Full text] [doi: [10.2471/BLT.11.099424](#)] [Medline: [22589569](#)]
15. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Barriers and facilitators to patient and public engagement and recruitment to digital health interventions: protocol of a systematic review of qualitative studies. *BMJ Open* 2016 Dec 02;6(9):e010895 [FREE Full text] [doi: [10.1136/bmjopen-2015-010895](#)] [Medline: [27591017](#)]
16. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. Geneva: World Health Organization; 2016.
17. Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. *Healthc Q* 2007;10(1):112-6, 118. [Medline: [17326376](#)]
18. 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](#)]
19. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015 Jan;4:1 [FREE Full text] [doi: [10.1186/2046-4053-4-1](#)] [Medline: [25554246](#)]
20. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015 Jan 02;349:g7647 [FREE Full text] [Medline: [25555855](#)]
21. Higgins J. *Cochrane Handbook for Systematic Reviews of Interventions Version 5*. Green S (editors) URL: <http://handbook-5-1.cochrane.org/> [accessed 2018-02-15] [WebCite Cache ID 6xFPALeTJ]
22. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10 [FREE Full text] [doi: [10.1186/1471-2288-7-10](#)] [Medline: [17302989](#)]

23. Bouma G, Admiraal JM, de Vries EGE, Schröder CP, Walenkamp AME, Reyners AKL. Internet-based support programs to alleviate psychosocial and physical symptoms in cancer patients: a literature analysis. *Crit Rev Oncol Hematol* 2015 Jul;95(1):26-37. [doi: [10.1016/j.critrevonc.2015.01.011](https://doi.org/10.1016/j.critrevonc.2015.01.011)] [Medline: [25701515](#)]
24. Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *J Health Serv Res Policy* 2005 Jul;10 Suppl 1:6-20. [doi: [10.1258/1355819054308576](https://doi.org/10.1258/1355819054308576)] [Medline: [16053580](#)]
25. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. *BMJ Open* 2014;4(9):e006021 [FREE Full text] [doi: [10.1136/bmjopen-2014-006021](https://doi.org/10.1136/bmjopen-2014-006021)] [Medline: [25200561](#)]
26. Ammenwerth E, Schnell-Inderst P, Hoerbst A. Patient empowerment by electronic health records: first results of a systematic review on the benefit of patient portals. *Stud Health Technol Inform* 2011;165:63-67. [Medline: [21685587](#)]
27. 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](https://doi.org/10.2196/jmir.3703)] [Medline: [25707035](#)]
28. Mold F, de LS, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. *Br J Gen Pract* 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: [10.3399/bjgp15X683941](https://doi.org/10.3399/bjgp15X683941)] [Medline: [25733435](#)]
29. 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](https://doi.org/10.2196/jmir.3171)] [Medline: [25669240](#)]
30. Alkureishi MA, Lee WW, Lyons M, Press VG, Imam S, Nkansah-Amankra A, et al. Impact of Electronic Medical Record Use on the Patient-Doctor Relationship and Communication: A Systematic Review. *J Gen Intern Med* 2016 May;31(5):548-560 [FREE Full text] [doi: [10.1007/s11606-015-3582-1](https://doi.org/10.1007/s11606-015-3582-1)] [Medline: [26786877](#)]
31. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. *Diabetes Technol Ther* 2014 Nov;16(11):784-793. [doi: [10.1089/dia.2014.0078](https://doi.org/10.1089/dia.2014.0078)] [Medline: [24999599](#)]
32. Bush RA, Connelly CD, Fuller M, Pérez A. Implementation of the Integrated Electronic Patient Portal in the Pediatric Population: A Systematic Review. *Telemed J E Health* 2015 Aug 10:1-16 [FREE Full text] [doi: [10.1089/tmj.2015.0033](https://doi.org/10.1089/tmj.2015.0033)] [Medline: [26258289](#)]
33. Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. *J Am Med Inform Assoc* 2014;21(4):737-741. [doi: [10.1136/amiainjnl-2013-002239](https://doi.org/10.1136/amiainjnl-2013-002239)] [Medline: [24154835](#)]
34. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. *Telemed J E Health* 2014 May;20(5):428-438 [FREE Full text] [doi: [10.1089/tmj.2013.0166](https://doi.org/10.1089/tmj.2013.0166)] [Medline: [24731239](#)]
35. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. *Ann Intern Med* 2013 Nov 19;159(10):677-687. [doi: [10.7326/0003-4819-159-10-201311190-00006](https://doi.org/10.7326/0003-4819-159-10-201311190-00006)] [Medline: [24247673](#)]
36. Liu J, Luo L, Zhang R, Huang T. Patient satisfaction with electronic medical/health record: a systematic review. *Scand J Caring Sci* 2013 Dec;27(4):785-791. [doi: [10.1111/scs.12015](https://doi.org/10.1111/scs.12015)] [Medline: [23181421](#)]
37. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. *Curr Diab Rep* 2010 Dec;10(6):422-435 [FREE Full text] [doi: [10.1007/s11892-010-0151-1](https://doi.org/10.1007/s11892-010-0151-1)] [Medline: [20890688](#)]
38. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. *J Am Med Inform Assoc* 2014 Jul;21(4):751-757. [doi: [10.1136/amiainjnl-2013-002501](https://doi.org/10.1136/amiainjnl-2013-002501)] [Medline: [24503882](#)]
39. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention, a systematic review. *BMC Med Inform Decis Mak* 2015;15:32 [FREE Full text] [doi: [10.1186/s12911-015-0159-1](https://doi.org/10.1186/s12911-015-0159-1)] [Medline: [25927384](#)]
40. 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 [FREE Full text] [doi: [10.2196/jmir.2439](https://doi.org/10.2196/jmir.2439)] [Medline: [23410671](#)]
41. Tao D, Or CK. Effects of self-management health information technology on glycaemic control for patients with diabetes: a meta-analysis of randomized controlled trials. *J Telemed Telecare* 2013 Apr 5:133-143. [doi: [10.1177/1357633X13479701](https://doi.org/10.1177/1357633X13479701)] [Medline: [23563018](#)]
42. Tulu B, Trudel J, Strong DM, Johnson SA, Sundaresan D, Garber L. Patient Portals: An Underutilized Resource for Improving Patient Engagement. *Chest* 2015 Jun 11:272-277. [doi: [10.1378/chest.14-2559](https://doi.org/10.1378/chest.14-2559)] [Medline: [26066707](#)]
43. Turner K, Klamon SL, Shea CM. Personal health records for people living with HIV: a review. *AIDS Care* 2016 Sep;28(9):1181-1187. [doi: [10.1080/09540121.2016.1153594](https://doi.org/10.1080/09540121.2016.1153594)] [Medline: [26917113](#)]
44. Vimalananda VG, Gupte G, Seraj SM, Orlander J, Berlowitz D, Fincke BG, et al. Electronic consultations (e-consults) to improve access to specialty care: A systematic review and narrative synthesis. *J Telemed Telecare* 2015 May 20:323-330. [doi: [10.1177/1357633X15582108](https://doi.org/10.1177/1357633X15582108)] [Medline: [25995331](#)]
45. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009 Jun;26(2):91-108 [FREE Full text] [doi: [10.1111/j.1471-1842.2009.00848.x](https://doi.org/10.1111/j.1471-1842.2009.00848.x)] [Medline: [19490148](#)]

46. Rouleau G, Gagnon M, Côté J, Payne-Gagnon J, Hudson E, Dubois C. Impact of Information and Communication Technologies on Nursing Care: Results of an Overview of Systematic Reviews. *J Med Internet Res* 2017 Apr 25;19(4):e122 [FREE Full text] [doi: [10.2196/jmir.6686](https://doi.org/10.2196/jmir.6686)] [Medline: [28442454](https://pubmed.ncbi.nlm.nih.gov/28442454/)]
47. Archer N, Fevrier-Thomas U, Lokker C, McKibbin KA, Straus SE. Personal health records: a scoping review. *J Am Med Inform Assoc* 2011;18(4):515-522 [FREE Full text] [doi: [10.1136/amiajnl-2011-000105](https://doi.org/10.1136/amiajnl-2011-000105)] [Medline: [21672914](https://pubmed.ncbi.nlm.nih.gov/21672914/)]
48. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC Med* 2010;8:63 [FREE Full text] [doi: [10.1186/1741-7015-8-63](https://doi.org/10.1186/1741-7015-8-63)] [Medline: [20961442](https://pubmed.ncbi.nlm.nih.gov/20961442/)]
49. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. *BMJ* 2013 Nov 20;347:f6753. [Medline: [24259324](https://pubmed.ncbi.nlm.nih.gov/24259324/)]
50. Agarwal S, LeFevre AE, 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;i1174. [doi: [10.1136/bmj.i1174](https://doi.org/10.1136/bmj.i1174)]
51. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *J Med Internet Res* 2011 Dec 31;13(4):e126 [FREE Full text] [doi: [10.2196/jmir.1923](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]

Abbreviations

AMSTAR: Assessing the Methodological Quality of Systematic Reviews

EHR: Electronic Health Records

PICOS: Participants, Interventions, Comparisons, Outcomes, Study Designs

PRISMA-P: Preferred Reporting Items for Systematic review and Meta-Analysis Protocols

PROSPERO: International Prospective Register of Systematic Reviews

RCT: Randomized Controlled Trial

Edited by G Eysenbach; submitted 23.02.18; peer-reviewed by T Irizarry, B Tulu, S O'Connor; comments to author 22.03.18; revised version received 27.04.18; accepted 15.05.18; published 07.06.18.

Please cite as:

Zanaboni P, Ngangue P, Mbemba GIC, Schopf TR, Bergmo TS, Gagnon MP

Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic Review of Reviews

J Med Internet Res 2018;20(6):e10202

URL: <http://www.jmir.org/2018/6/e10202/>

doi: [10.2196/10202](https://doi.org/10.2196/10202)

PMID:

©Paolo Zanaboni, Patrice Ngangue, Gisele Irène Claudine Mbemba, Thomas Roger Schopf, Trine Strand Bergmo, Marie-Pierre Gagnon. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 07.06.2018. 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.

Review

The Social Effects of Exergames on Older Adults: Systematic Review and Metric Analysis

Jinhui Li¹, PhD; Mojisola Erdt¹, PhD; Luxi Chen², PhD; Yuanyuan Cao¹, PhD; Shan-Qi Lee¹, BSocSci; Yin-Leng Theng¹, PhD

¹Centre for Healthy and Sustainable Cities, Nanyang Technological University, Singapore, Singapore

²Centre for Family and Population Research, National University of Singapore, Singapore, Singapore

Corresponding Author:

Jinhui Li, PhD

Centre for Healthy and Sustainable Cities

Nanyang Technological University

31 Nanyang Link

Singapore,

Singapore

Phone: 65 67915214

Email: lijinhuihust@gmail.com

Abstract

Background: Recently, many studies have been conducted to investigate the effects of exergames on the social well-being of older adults.

Objective: The aim of this paper is to synthesize existing studies and provide an overall picture on the social effects of exergames on older adults.

Methods: A comprehensive literature search with inclusive criteria was conducted in major social science bibliographic databases. The characteristics of exergames, participants, methodology, as well as outcome measurements were extracted from the relevant studies included in the review. The bibliometric and altmetric outreach of the included studies were also investigated.

Results: A total of 10 studies were included in the review, with 8 studies having used the Nintendo Wii platform. Most of the studies recruited healthy older adults from local communities or senior activity centers. Three groups of social-related outcomes have been identified, including emotion-related, behavior-related, and attitude-related outcomes. A metric analysis has shown that the emotion-related and behavior-related outcomes received high attention from both the academic community and social media platforms.

Conclusions: Overall, the majority of exergame studies demonstrated promising results for enhanced social well-being, such as reduction of loneliness, increased social connection, and positive attitudes towards others. The paper also provided implications for health care researchers and exergame designers.

(*J Med Internet Res* 2018;20(6):e10486) doi:[10.2196/10486](https://doi.org/10.2196/10486)

KEYWORDS

active video games; psychosocial well-being; ageing; literature review; citation analysis

Introduction

Significant population aging has been experienced by countries worldwide. In 2015, a report from the United Nations [1] indicated the number of older adults aged 60 and above was 901 million, which equated to 13% of the entire population. The report predicted that this number is expected to double by 2050, reaching nearly 2.1 billion people. Population ageing has become a major global demographic trend and has subsequently raised many public concerns on the well-being of older adults

[2,3]. Older adults often suffer from several common negative events, such as a lack of close family ties (eg, living alone), loss of a loved one, a decline in mobility, or a reduction in active participation in social activities. The accumulation of these negative events could result in inadequate social support or impaired social interaction [4,5]. Some studies have indicated that a lack of social interaction led to frequently experienced social problems and disorders, such as social isolation and loneliness [6-8]. Given the potential harmful effects of social isolation and loneliness, it is important to develop social

interventions to reduce emotional damage to older adults and inappropriate health and social service usage.

With the advent of digital technology, exergames, which combine digital gaming and physical exercise, are commonly used as daily exercise programs [9]. Despite being originally designed for entertainment, exergames are increasingly used for health promotion. There is a rapid growth in the popularity and use of exergames as health programs in public settings, such as in communities [10], school, and work environments [11]. Many previous studies [12,13] have assessed the potential benefits of exergames on participants' physical, cognitive, and psychological well-being. For example, evidence from a 2-week pilot study demonstrated that exergames were able to significantly improve upper extremity function for poststroke patients [14]. A pilot study from Chan et al [15] showed that older adults in virtual reality cognitive training programs had better improvements in repetition and memory retention than those in usual programs. Albores et al [16] reported that older patients with chronic obstructive pulmonary disease showed significant improvements in their emotional well-being after a 12-week Wii Fit training program.

The social effects of exergames have drawn considerable attention from researchers [17-19]. Exergaming is a social experience which gives the players opportunities to interact with each other. This can in turn foster social networking and friendships among the players. Results from a study published by Kooiman and Sheehan [17] showed that exergaming over the internet increased students' social relatedness in physical education. Social interaction was reported as the most important motivation for adolescents in a 20-week exergaming intervention [20]. In addition to the positive effects experienced by the younger generations, recent research on exergames has also extended to investigate of the social effects of exergaming to the older population [21,22].

Considering the major concerns regarding older adults with social disorders, it is important to have an overview on whether exergames may serve as an effective intervention for the social well-being of this group of people. In the literature on exergaming, many systematic reviews examined only on the physical and cognitive benefits of exergaming on older adults [23-25]. Some studies have reviewed the psychosocial effects of exergames, however, these studies are focused on psychological changes, such as the effect of exergaming on depression, mood, and enjoyment of exercise [26-29]. Therefore, an overall picture on the social effects of exergames on older adults is required. With increasing research efforts in the field of exergaming, the current systematic review was conducted with the aim of synthesizing the existing literature and to provide implications for improving social well-being in older adults using exergaming. Additionally, the review also investigated the bibliometric and altmetric outreach of the included studies in this systematic review, to understand their impacts in both academic and nonacademic (social media) platforms.

Methods

The current review adopted the definition of an exergame from Oh and Yang [9] who defined it as "an experiential activity in

which playing exergames or any videogames requires physical exertion or movements that are more than sedentary activities and also include strength, balance, and flexibility activities." The studies included in the systematic review should thus involve exergames, according to the above definition, as the primary intervention of the study. Other inclusion criteria were: (1) the study should incorporate measures of social outcomes such as social connection, social bonding, or loneliness; (2) the study should target participants aged 55 or above; (3) the study should report original research in English. The term "older adult" commonly refers to a person having a chronological age of 65 years and older [30] but extensive studies indicate that insights into the needs of future older adults could be provided if pre-elderly adults aged 55 to 64 were included in the investigations [31,32]. Therefore, the current review included studies with participants aged 55 and above. The current review has a restriction to English-only articles because non-English publications do not appear in major bibliographic databases. In order to achieve a complete picture of exergaming effects on social outcomes, there were no constraining criteria applied with regard to the study design. Thus, the review included studies which used both qualitative and quantitative methods.

A comprehensive literature search was conducted in major social science bibliographic databases, including PsycINFO, PubMed, CINAHL, and ScienceDirect. Potential studies were identified by the combination of exergame terms (exergame OR Wii OR Kinect OR active video game), social terms (social* OR social support OR social interaction OR social bonding OR communication*), and ageing terms (aging OR aged OR elderly OR older OR senior). Reference lists of the included studies and relevant reviews were also inspected for additional studies to be included in the current systematic review. A total of 319 articles published before 22 January 2017 were retrieved for review and analysis. All articles were assessed using either the title, abstract, or full text to determine their eligibility in the systematic review conducted in this paper. The articles without full texts in any of the databases listed above excluded if the full texts could not be retrieved using online search engines or by contacting the authors directly. After identifying the final list of included studies, the characteristics of exergames, participants (country, sample size, age, and profile), methodology (study design and duration), as well as outcome measurements were extracted from the studies. Screening and data extraction was performed mainly by one reviewer, while a second reviewer was assisted by checking and editing the extracted data.

To investigate the bibliometric score (in terms of citation count), the altmetric score, and the social media presence of the articles included in the systematic review, we collected citation counts from Scopus [33], as well as usage and capture data from PlumX [34] before 31 April 2017. We also collected Tweet counts, number of Mendeley readers, and the Altmetric Attention Score from Altmetric [35] for each included article. Altmetrics can be described as new or alternative measures of the impact of research objects, based mainly on social media data sources [36]. The Altmetric Attention Score is a weighted aggregate metric comprising diverse online sources from news outlets, policy documents, blogs, Wikipedia, Twitter, Facebook,

YouTube, and other social media sources. Usage data from PlumX is a combined metric incorporating counts from downloads, views, library holdings, video plays, clicks, collaborators, and other usage metrics. Capturing data from PlumX comprises counts from bookmarks, favorites, followers, readers, subscribers, watchers, exports or saves, and code forks. The 2015 QS world university rankings [37] were used to determine the prestigious universities. Prestigious universities were defined as those listed in the 2015 QS world university rankings. We used a logarithmic scale for a better visualization of the data. These metrics gave us an insight into the outreach and impact of the included studies in this systematic review.

Results

Study Selection

According to the inclusive criteria, a total of 10 studies were eligible to be included in the final review process. Figure 1 illustrates the flowchart of the systematic review process for the selection of the included studies. Tables 1 and 2 outline the key characteristics of these 10 studies.

Characteristics of the Studies Included

Exergame Types

Of the 10 studies investigated, 8 investigated social effects of exergames using Nintendo Wii, while 2 studies used Microsoft Xbox Kinect. Both platforms are the most popular exergaming platforms in the current market, and both offer console-based devices and games which make exergaming possible in the home setting. In terms of game topics, it is interesting to note that half of the studies ($n=5$) applied games from Nintendo Wii Sports package [39-43]. The Wii Sports game package allows participants to play virtual sport games (such as tennis, bowling, baseball, golf, or boxing) by performing body motions that they would do in actual sports. Another study from Wu, Li, and Theng [45] also used a virtual bowling game, but from the Microsoft Kinect Sports game package. Two studies used exercise games from Nintendo Wii Fit, or its successor Nintendo Wii Fit U [21,38]. Wii Fit exergames are different to Wii Sports games as they aim to improve players' physical fitness through exercise activities such as strength training, aerobics, yoga, and balance games. Besides those simulating actual exercise, exergames with topics from daily life activities were also found in two studies [39,43], such as cookery simulation-styled *Cooking Mama* and party simulation-styled *Wii Party*.

Participants

Most of the studies recruited healthy older adults from local communities or senior activity centers. There were 2 studies [38,44], however, that investigated the social effects of exergames on older adults with physical or social problems, such as those with impaired balance, with a disability, or those who were socially isolated. Six studies included participants with Western cultural backgrounds, including the US, Australia, and Canada. Among the studies focused on Western cultural

backgrounds, one study focused on African American participants [21]. Four studies were conducted in the context of Asia, all of which were conducted in Singapore. The majority of the studies had a small sample size with less than 50 participants.

Methodology

Five of the 10 studies applied poststudy qualitative methods to assess the social effects of exergames, such as semistructured interview, semistructured group interview, or focus group discussion. Among those applying quantitative methods, 4 studies tested the effects between exergames and other control conditions. For example, Wu et al [45], Jung et al [39], and Kahlbaugh et al [40] compared exergames with traditional activities such as playing board games, watching television programs, or performing normal exercise. Xu et al's [22] study is an exception which compared the effects among 3 exergame conditions (playing alone vs playing with elderly vs playing with youths). One study [43] applied a within-group experiment method to compare the effects before and after the exergame intervention. The duration of the intervention period ranged from 1 to 12 weeks, while most of the studies involved 8 sessions or more. Two reviewers independently applied the risk of bias tool from Cochrane Collaboration [46] to assess the methodological quality of 4 studies with control conditions. Table 3 shows the results of the quality assessment of the 4 controlled studies. According to the Cochrane recommendations [46], 2 studies [22,45] were identified to have "High Risk of Bias," while the other 2 studies [39,40] were identified to have "Moderate or Unclear Risk of Bias."

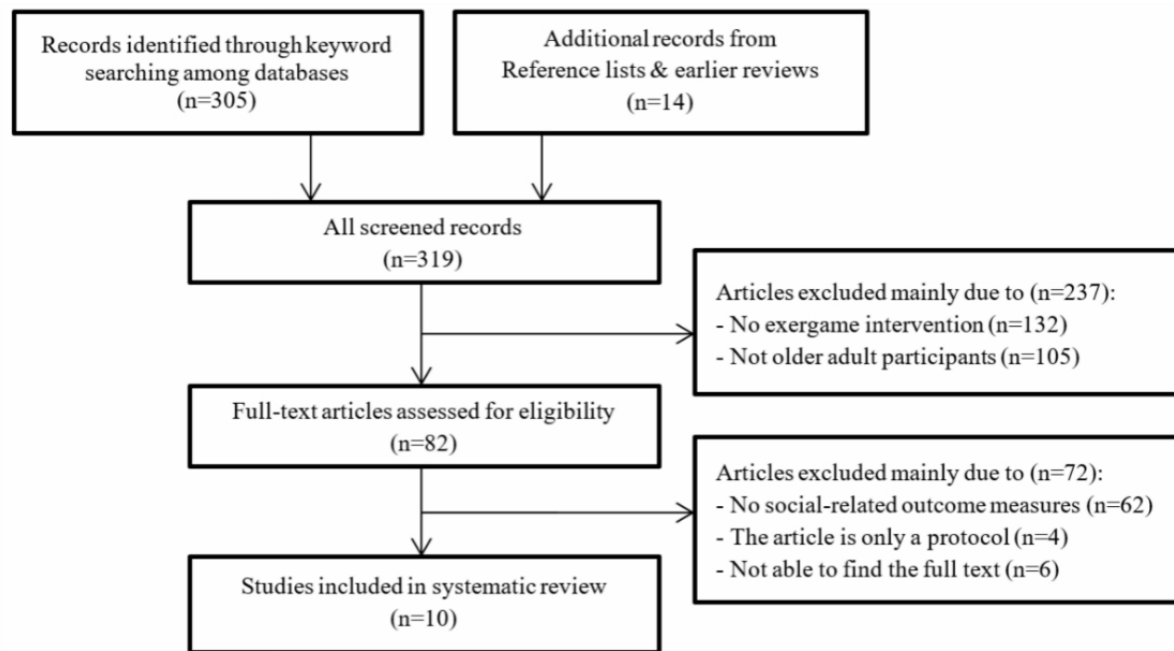
Social Outcomes

The findings of the included studies have identified several social-related outcomes. Based on the different natures of the outcomes, they were categorized into three groups: emotion-related, behavior-related, and attitude-related.

Emotion-Related

Loneliness was identified to be the main emotion-related social outcome affected by exergames. Jung et al conducted a study to assess the potential of Nintendo Wii in improving the quality of life among older adults in a long-term care facility [39]. Their results indicated that elderly participating in the Wii condition group had a significantly lower level of loneliness than those participating in the other condition group, who played traditional board games. Similarly, another between-group study in the US also reported that playing Wii rather than watching television programs led to a lower level of loneliness [40].

Additionally, Xu et al found a significant decrease in loneliness among older adults after exergaming, although little differences were found across different play types or age groups (young-old vs old-old) [22]. In the same study, social anxiousness was also found to have significantly declined, however, this was noticed only in the young-old participants who played exergames with youths [22].

Figure 1. Flowchart of the systematic review process for the included studies.**Table 1.** Characteristics of the included studies (exergames and participants).

Study	Exergames	Participants			
		Country	Sample size	Age in years	Profile
Agmon et al [38]	Wii Fit Exergames (basic step, soccer heading, ski slalom, and table tilt)	US	7	84 (5) ^a	<ul style="list-style-type: none"> Older adults with impaired balance From care retirement communities
Chao et al [21]	Wii Fit U (balance games, yoga poses, strength training, aerobics, and dance games)	US	12	64.17 (6.74) ^a	<ul style="list-style-type: none"> Healthy older female adults aged 55 years and above From African American community
Jung et al [39]	Wii Sports (tennis, bowling, baseball and boxing) and Cooking Mama	Singapore	45	56-92	<ul style="list-style-type: none"> Local healthy older adults From senior activities centers
Kahlbaugh et al [40]	Wii game (Wii bowling)	US	35	82 (9.8) ^a	<ul style="list-style-type: none"> Healthy older adults From independent living residential apartments
Keogh et al [41]	Nintendo Wii Sports	Australia	34	83 (8) ^a	<ul style="list-style-type: none"> Healthy older adults From residential aged care centers
Millington [42]	Exergame such as Wii Bowling	Canada	8	N/A ^b	<ul style="list-style-type: none"> Healthy older persons From retirement centers
Theng et al [43]	Wii games such as “Wii Sports,” “Cooking Mama,” and “Wii Party”	Singapore	28	>60	<ul style="list-style-type: none"> Healthy older adults From a seniors’ activity center
Wollersheim et al [44]	Wii games	Australia	11	73.5 (9) ^a	<ul style="list-style-type: none"> Older women with a disability or who are socially isolated
Wu et al [45]	Kinect Sport Bowling with a partner	Singapore	113	>55	<ul style="list-style-type: none"> Healthy old adults From senior activity centers and community clubs
Xu et al [22]	Three Kinect exergames	Singapore	89	75	<ul style="list-style-type: none"> Local healthy older adults From senior activities centers

^aAge presented as mean (SD).^bN/A: not available.

Table 2. Characteristics of the included studies (methodology and outcome).

Study	Methodology		Outcome	
	Study design ^a	Duration	Measurement	Effect
Agmon et al [38]	Within-group, poststudy interview	3 sessions per week; 12 weeks	Socialization: semistructured interview	Six out of 7 participants described that they enjoyed playing Wii Fit with their grandchildren
Chao et al [21]	Within-group, poststudy interview	2 sessions per week; 12 weeks	Social connection: semistructured interviews	The program encouraged participants to get connected with others
Jung et al [39]	Between-group, 2 conditions: <ul style="list-style-type: none"> • Playing exergames (N=30) • Playing traditional board games (N=15) 	3 sessions per week; 6 weeks	Loneliness: UCLA ^b Loneliness Scale	Exergame group versus control group: $t_{43}=5.34$, $P<.01$
Kahlbaugh et al [40]	Between-group, 3 conditions: <ul style="list-style-type: none"> • Playing exergames with a partner (N=16) • Watching television programs with a partner (N=12) • No visits (N=7) 	1 session per week; 10 weeks	Loneliness: UCLA Loneliness Scale	Exergame group versus television group: $F_{2,30}=6.24$, $P<.005$
Keogh et al [41]	Within-group, poststudy interview	8 weeks	Socialization: semistructured group interview	“Several (P5 and P4) found that having a ‘new face’ to interact with and someone who would sit and listen was something to look forward to.”
Millington [42]	Within-group, poststudy interview	Wii constant use at one center; 1 to 2 times per month for another two centers	Social engagement: Interview	“Virtual bowling can bring people together in communal spaces while also ‘getting them up’ and active”
Theng et al [43]	Within-group, pre- and poststudy measurement	6 sessions	Positive attitude: semantic differential scale	Mean positive attitude towards youth: increased from 4.06 (SD 0.78) to 4.27 (SD 0.43)
Wollersheim et al [44]	Poststudy focus group discussion	2 sessions per week; 6 weeks	Social bonding: focus group discussion	“Many of the women noted that being more technologically adept allowed them to be more connected to their grandchildren.”
Wu et al [45]	Between-group, 4 conditions: <ul style="list-style-type: none"> • Playing collaborative exergame (N=26) • Playing competitive exergame (N=24) • Playing collaborative traditional exercise (N=25) • Playing competitive traditional exercise (N=20) 	2 sessions per week; 4 weeks	Social presence: The Social Presence in Gaming questionnaire	Exergame group versus traditional exercise group: $\beta=-.20$, $P<.10$ (in general intention model)
Xu et al [22]	Between-group, 3 conditions: <ul style="list-style-type: none"> • Playing exergames with their peers (N=31) • Playing with an adolescent (N=26) • Playing alone (N=31) 	3 sessions per week; 1 week	3 measurements: <ul style="list-style-type: none"> • Loneliness: UCLA Loneliness Scale • Social anxiousness: the interaction anxiousness scale • Sociability: sociability scale 	3 effects: <ul style="list-style-type: none"> • Loneliness: significantly decreased after playing exergames, $F_{1,83}=.57$, $P<.05$ • Social anxiousness: did not change significantly, $F_{1,83}=1.58$, $P=.212$ • Sociability: significantly increased after playing exergames, $F_{1,83}=3.95$, $P=.050$

^aFor social outcomes.^bUCLA: University of California, Los Angeles.

Table 3. Results of quality assessment of four controlled studies included in the review.

Citation	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
Jung et al [39]	Low bias	Unclear	Unclear	Unclear	Low bias	Low bias	Low bias
Kahlbaugh et al [40]	Low bias	Unclear	Unclear	Unclear	Low bias	Low bias	Low bias
Wu et al [45]	High bias	High bias	Unclear	Unclear	Low bias	Low bias	Low bias
Xu et al [22]	High bias	Unclear	Unclear	Unclear	Low bias	Low bias	Low bias

Attitude-Related

Wu et al [45] presented a study which examined the exergame effects on social presence, which was defined as the sense of connecting or being with others in a media-mediated environment. Their results found that older adults in the exergame setting had a significant lower social presence than those in traditional exercise. Another Singapore study from Theng et al [43] showed that playing exergames with youths led to improvement in older adults' positive attitude toward the younger age group.

Metric Analysis

Currently, the study from Agmon et al [38] has received the most attention from the scholarly community to date, with a total of 105 citations, 27 of which came from prestigious universities. The study published by Wollersheim et al [44] has also received a good amount of attention with a total of 63 citations, of which 8 were from prestigious universities. Most

of the citations of these two studies came from articles published between 2014 and 2016, and mainly from papers published in the fields of Medicine and Computer Science. Since the studies from Xu et al [22] and Chao et al [21] were only recently published (they were published in December 2016 and January 2017 respectively), no citations could yet be found for these articles. Figure 2 shows the bibliometric outreach of the exergame studies.

Figure 3 gives an overview of the altmetric outreach of the exergame studies. The study with the highest Altmetric Attention Score was Theng et al with a score of 29 [43]. This high score was attributed to 3 mentions on news outlets in March 2017, naming this study as an example of how Nintendo's motion control system has helped to make gaming accessible to new groups of users. Kahlbaugh et al had a very high PlumX usage count of 5300, and a high PlumX capture count of 498 [40]. These were mainly due to abstract views, clicks on outbound links, and exports or saves on EBSCO [47].

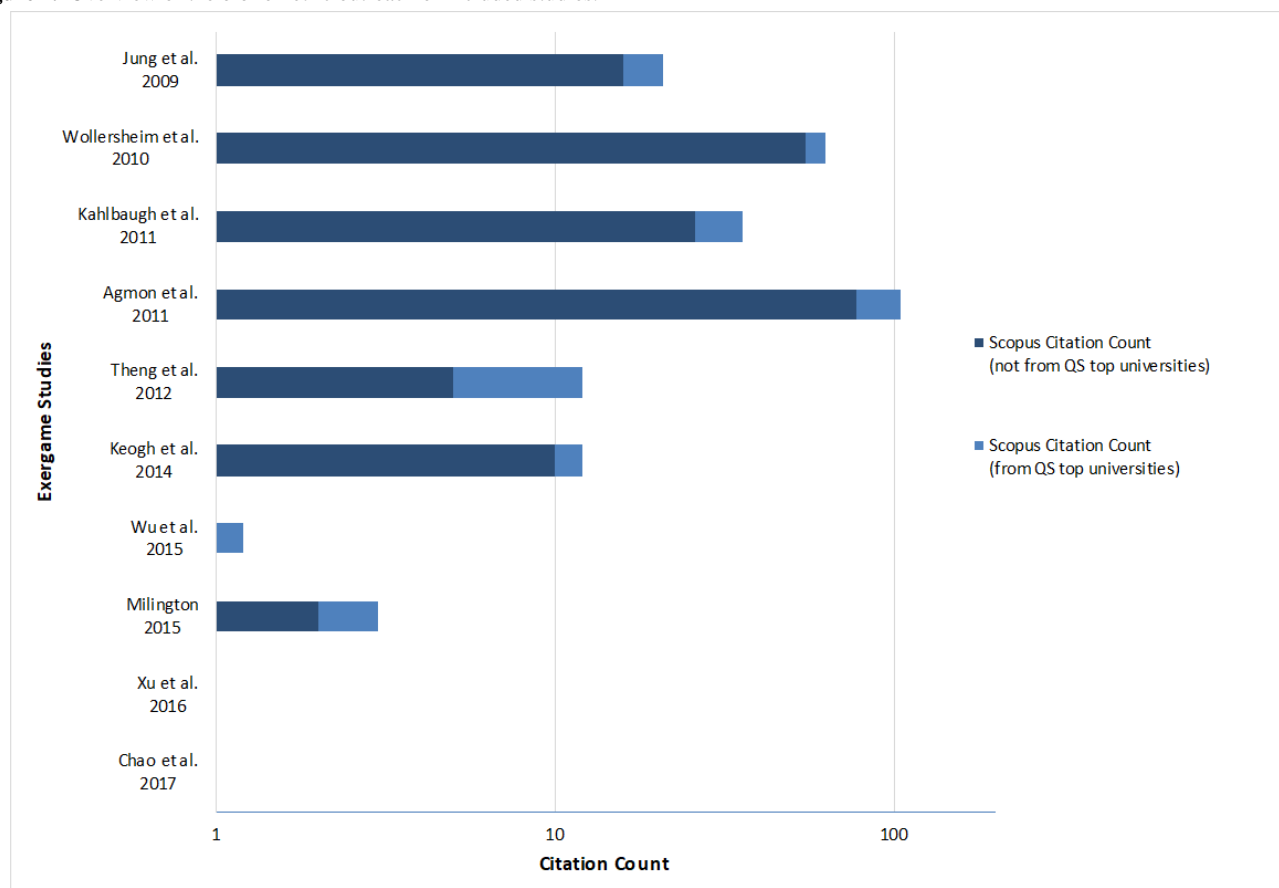
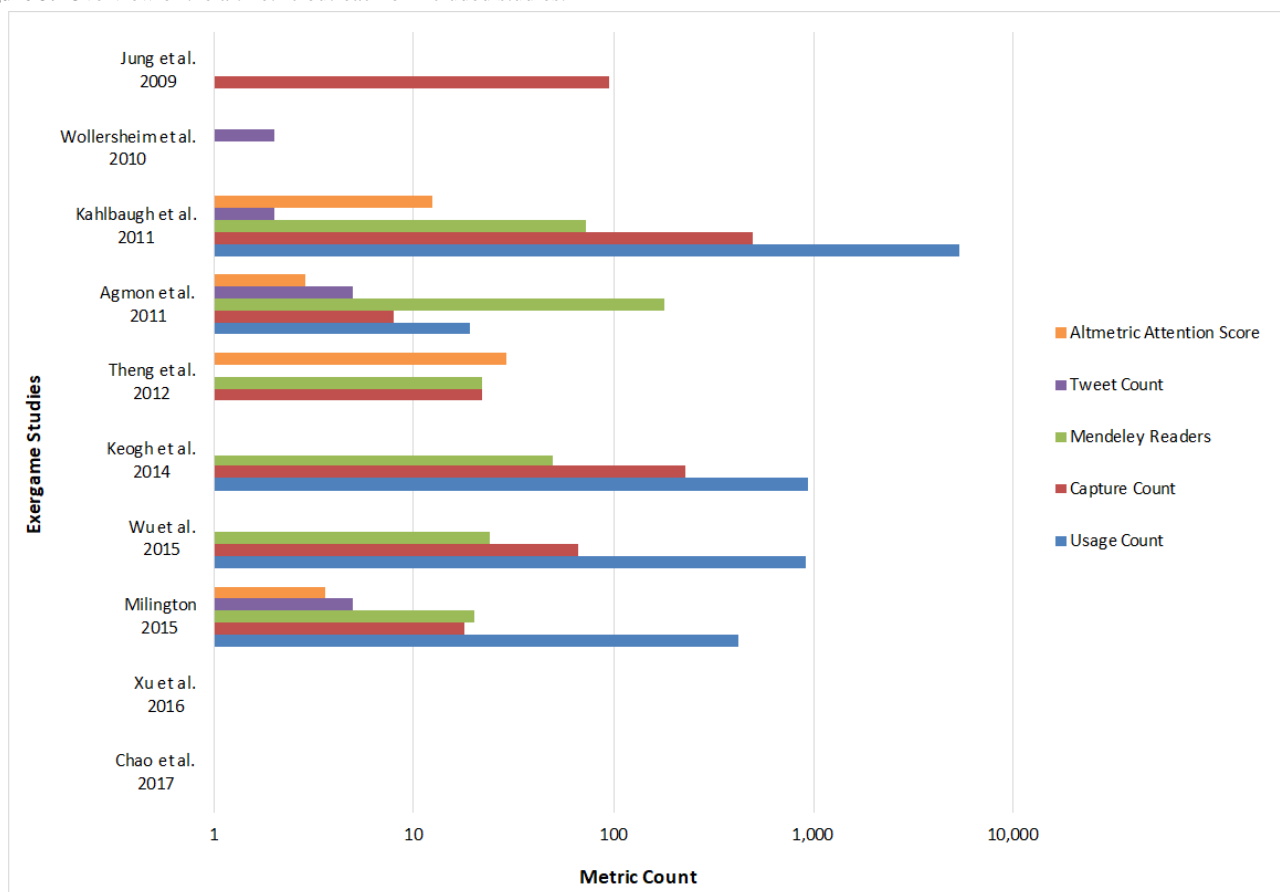
Figure 2. Overview of the bibliometric outreach of included studies.

Figure 3. Overview of the altmetric outreach of included studies.

Agmon et al had a high count of 180 Mendeley readers, but since Altmetric.com does not include Mendeley readers in its score, this is not reflected as this study had a score of 3 [38]. Again, two recent studies [21,22] did not yet have any altmetrics, however, one tweet could already be found for the article from Chao et al [21], giving this article an Altmetric Attention Score of 0.5. Tweet counts were however low across all studies, with Milington [42] and Agmon et al [38] having the highest count of 5 tweets each.

Discussion

Principal Findings

While previous reviews have synthesized the psychosocial effects of exergames [26,28,29], the current review focused specifically on social benefits and extended to the ageing population. The systematic review shows an increasing interest in using exergames to improve the social well-being among older adults, with 9 out of 10 included studies published after the year 2010. Nevertheless, the small number of identified studies also calls for more investigation in this new research area. Several systematic reviews had similar findings in the area of exergames for mental health, with 12 and 9 studies found in the reviews published by Spek et al [48] and Li et al [28] respectively. Overall, the majority of exergame studies demonstrated promising results for enhanced social well-being in older adults, such as a reduction of loneliness, an increase in social connection, and positive attitudes towards others.

Social Benefits of Exergames

Although the social benefits of exergames are often discussed in the literature on exergames, there has been no particular review found on this topic. By summarizing the existing original studies, the findings from the current review showed that exergames could be an effective intervention for social improvements among older adults. The review supported the finding that exergames were able to reduce the loneliness level among older adults. The decrease in loneliness was perhaps not due to playing the exergame itself, but rather due to the increased interactions between the participants and other players. In a large sample survey study, Lee and Ishii-Kuntz [49] indicated that doing an activity with other people reduced loneliness among older adults. Furthermore, many of the included studies suggested that exergames provide opportunities for social interaction and connectedness with peers and family members [38,41,44]. In addition to loneliness, older adults often lack the motivation to engage in exercise. Chao et al indicated that these behavior-related social outcomes of exergames may increase exercise motivation and adherence among older adults [27]. A metric analysis has shown that the emotion-related and behavior-related outcomes of exergames received a lot of attention in the academic community as well as on social media.

Attitude-related social outcomes are a new finding which have not been reported in previously published exergame reviews (eg, those published by Chao et al [27] or Matallaoui et al [29]). Exergames were found to affect a sense of being with others [45] and positive attitudes towards others [43], similar to findings in recent research on persuasive video gaming [50,51].

An experiment from a study published by Alhabash and Wise [50] found that video game role-play led to a change in students' explicit and implicit attitudes toward Palestinians and Israelis. In another study, students who played the persuasive social impact game had an increased positive attitude towards the homeless [51]. Results from this review further supports that active video games, such as exergames, have the potential to affect older adults' attitudes towards other groups of people. Although attitude-related social outcomes currently have a low academic impact in the exergaming research area, they have begun to receive a certain amount of discussion on social media.

Implications for Future Study

This review showed that that Nintendo Wii was the most frequently used exergaming platform in the included studies. This finding is supported by another review [27], which reported that the Nintendo Wii is one of the most accessible and popular exergames for seniors. Chao et al [27] further indicated the high attendance rates among older players in Wii exergames programs. Although the evidence may suggest Wii to be a suitable platform for older adults to perform exergames, there have been no studies conducted which investigate the difference in effect between exergaming using a Wii and other platforms, such as the Microsoft Kinect consoles. More studies are needed to compare the effects of different exergaming platforms. Sport games were identified to be the favored type of games used in the included studies, and bowling was tested in 3 studies [40,42,45]. According to the American College of Sports Medicine [52], older adults are encouraged to perform physical activities that maintain or increase their balance and flexibility through slow movements. The bowling games exergames allow for slow movements which match the typical physical activities recommended for older adults. Additionally, bowling is a self-paced exercise in which older adults could take the time they needed to perform the moves [27]. Crucially, all the studies applied commercial exergames available on the market and none of the interventions were integrated with social theories. This highlights the need to combine social-related theories with exergame programs in order to optimize the effectiveness of social improvements.

Although most of studies targeted healthy older adults, two studies examined the social effects on older adults with physical disabilities [38,44]. Physical disability, particularly low mobility, has often been identified as a risk factor for social isolation among older adults [53]. Low mobility prevents seniors from participating in active social engagement and connection, leading to common social disorders such as loneliness. The two

included studies with older adult participants with physical disabilities showed that playing exergames improved their social well-being by increasing social bonding with their peers and grandchildren. However, the physical limitations of this group of older adults may have led to some difficulties in interacting with the exergames. They may have been exposed to frustrating experiences or even accidents if the exergames were performed without proper human or technical assistance. As a result, health care providers and exergame designers should take this into consideration when implementing future social exergame programs for older adults with disabilities. In terms of cultural background, the studies were conducted in both Western and Asian contexts. It appears that exergames might have social effects on older adults with various cultural backgrounds, but knowledge is lacking on whether the social outcomes would be affected by cultural factors. Future studies are recommended to compare the social effects between different cultural contexts.

The study designs included in this study varied in rigor, with 6 studies applying a within-group design, and 4 studies applying between-group design by comparing exergames with a control condition. Although the majority of studies showed promising results for the use of exergames for social enhancement, the conclusions need to be interpreted with caution due to the limited number of randomized controlled trials. The included studies were either predominantly small pilot trials or feasibility studies; they lacked the adequate sample sizes needed for a powered efficacy trial. Meanwhile, half of the studies used qualitative methods for data collection. Without validated quantitative instruments, their findings do not have the capacity to detect significant changes in social outcomes.

Limitations

There are some limitations in the review. Due to the limited number of identified studies, the systematic review included articles with both qualitative and quantitative analyses. The quality assessment of the included studies was difficult to conduct, and it was not possible to produce mean effect sizes via a meta-analysis. However, the review was broad in scope and included a diversity of study conditions and social outcome measures. Another limitation is that the key conclusions should be interpreted with caution due to the small number of included studies. Furthermore, relevant studies may have been unintentionally excluded because of the specific keywords used and the databases selected. Lastly, a publication bias, particularly language bias, might have occurred because we restricted the search to English language publications.

Acknowledgments

This research was supported by (1) the National Research Foundation Singapore under its National Innovation Challenge on Active and Confident Ageing (Award #MOH/NIC/CAHIG03/2016) and administered by the Singapore Ministry of Health's National Medical Research Council; and (2) the National Research Foundation, Prime Minister's Office, Singapore under its Science of Research, Innovation and Enterprise programme (SRIE Award #NRF2014-NRF-SRIE001-019).

Conflicts of Interest

None declared.

References

1. United Nations. New York: United Nations; 2015. World Population Ageing URL: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf [accessed 2018-06-14] [WebCite Cache ID 70A5W23VI]
2. Cheng ST, Fung HH, Li LW, Li T, Chi I. Successful Aging: Concepts, Reflections and Its Relevance to Asia. In: Cheng ST, Chi I, Fung HH, Li LW, Woo J, editors. Successful Aging: Asian Perspectives. Netherlands: Springer; 2015.
3. Read S, Grundy E, Foverskov E. Socio-economic position and subjective health and well-being among older people in Europe: a systematic narrative review. *Aging Ment Health* 2016 Mar;20(5):529-542 [FREE Full text] [doi: [10.1080/13607863.2015.1023766](https://doi.org/10.1080/13607863.2015.1023766)] [Medline: [25806655](https://pubmed.ncbi.nlm.nih.gov/25806655/)]
4. Bøen H, Dalgard OS, Bjertness E. The importance of social support in the associations between psychological distress and somatic health problems and socio-economic factors among older adults living at home: a cross sectional study. *BMC Geriatr* 2012 Jun 08;12:27 [FREE Full text] [doi: [10.1186/1471-2318-12-27](https://doi.org/10.1186/1471-2318-12-27)] [Medline: [22682023](https://pubmed.ncbi.nlm.nih.gov/22682023/)]
5. Alexandrino-Silva C, Alves TF, Tófoli LF, Wang Y, Andrade LH. Psychiatry: life events and social support in late life depression. *Clinics* 2011;66(2):233-238. [doi: [10.1590/S1807-59322011000200009](https://doi.org/10.1590/S1807-59322011000200009)]
6. Singh A, Misra N. Loneliness, depression and sociability in old age. *Ind Psychiatry J* 2009 Jan;18(1):51-55 [FREE Full text] [doi: [10.4103/0972-6748.57861](https://doi.org/10.4103/0972-6748.57861)] [Medline: [21234164](https://pubmed.ncbi.nlm.nih.gov/21234164/)]
7. Holmén K, Furukawa H. Loneliness, health and social network among elderly people—a follow-up study. *Archives of Gerontology and Geriatrics* 2002 Nov;35(3):261-274. [doi: [10.1016/S0167-4943\(02\)00049-3](https://doi.org/10.1016/S0167-4943(02)00049-3)]
8. Grenade L, Boldy D. Social isolation and loneliness among older people: Issues and future challenges in community and residential settings. *Australian Health Review* 2008;32(3):468-478 [FREE Full text]
9. Oh Y, Yang S. Defining exergames & exergaming. 2010 Presented at: Meaningful Play 2010; October 21 - 23; East Lansing, MI, USA.
10. Baranowski T, Maddison R, Maloney A, Medina E, Simons M. Building a Better Mousetrap (Exergame) to Increase Youth Physical Activity. *Games Health J* 2014 Apr;3(2):72-78. [doi: [10.1089/g4h.2014.0018](https://doi.org/10.1089/g4h.2014.0018)] [Medline: [26196047](https://pubmed.ncbi.nlm.nih.gov/26196047/)]
11. Maddison R, Simons M, Straker L, Witherspoon L, Palmeira A, Thin AG. Active video games: An opportunity for enhanced learning and positive health effects? *Cognitive Technology Journal* 2013;18(1):6-13 [FREE Full text]
12. Daley AJ. Can exergaming contribute to improving physical activity levels and health outcomes in children? *Pediatrics* 2009 Aug;124(2):763-771. [doi: [10.1542/peds.2008-2357](https://doi.org/10.1542/peds.2008-2357)] [Medline: [19596728](https://pubmed.ncbi.nlm.nih.gov/19596728/)]
13. Kamel Boulos MN. Xbox 360 Kinect Exergames for Health. *Games for Health Journal* 2012 Oct;1(5):326-330. [doi: [10.1089/g4h.2012.0041](https://doi.org/10.1089/g4h.2012.0041)]
14. Mouawad MR, Doust CG, Max MD, McNulty PA. Wii-based movement therapy to promote improved upper extremity function post-stroke: a pilot study. *J Rehabil Med* 2011 May;43(6):527-533 [FREE Full text] [doi: [10.2340/16501977-0816](https://doi.org/10.2340/16501977-0816)] [Medline: [21533334](https://pubmed.ncbi.nlm.nih.gov/21533334/)]
15. Chan CLF, Ngai EKY, Leung PKH, Wong S. Effect of the adapted Virtual Reality cognitive training program among Chinese older adults with chronic schizophrenia: a pilot study. *Int J Geriatr Psychiatry* 2010 Jun;25(6):643-649. [doi: [10.1002/gps.2403](https://doi.org/10.1002/gps.2403)] [Medline: [19806599](https://pubmed.ncbi.nlm.nih.gov/19806599/)]
16. Albores J, Marolda C, Haggerty M, Gerstenhaber B, Zuwallack R. The use of a home exercise program based on a computer system in patients with chronic obstructive pulmonary disease. *J Cardiopulm Rehabil Prev* 2013;33(1):47-52. [doi: [10.1097/HCR.0b013e3182724091](https://doi.org/10.1097/HCR.0b013e3182724091)] [Medline: [23235321](https://pubmed.ncbi.nlm.nih.gov/23235321/)]
17. Kooiman BJ, Sheehan DP. The efficacy of exergames for social relatedness in online physical education. *Cogent Education* 2015 Jun 3;2(1):1045808. [doi: [10.1080/2331186X.2015.1045808](https://doi.org/10.1080/2331186X.2015.1045808)]
18. Marker AM, Staiano AE. Better Together: Outcomes of Cooperation Versus Competition in Social Exergaming. *Games Health J* 2015 Feb;4(1):25-30 [FREE Full text] [doi: [10.1089/g4h.2014.0066](https://doi.org/10.1089/g4h.2014.0066)] [Medline: [26181677](https://pubmed.ncbi.nlm.nih.gov/26181677/)]
19. Kim J, Timmerman CE. Effects of Supportive Feedback Messages on Exergame Experiences. *Journal of Media Psychology* 2018 Jan;30(1):29-40. [doi: [10.1027/1864-1105/a000175](https://doi.org/10.1027/1864-1105/a000175)]
20. Staiano AE, Abraham AA, Calvert SL. Motivating effects of cooperative exergame play for overweight and obese adolescents. *J Diabetes Sci Technol* 2012 Jul 01;6(4):812-819 [FREE Full text] [doi: [10.1177/193229681200600412](https://doi.org/10.1177/193229681200600412)] [Medline: [22920807](https://pubmed.ncbi.nlm.nih.gov/22920807/)]
21. Chao Y, Musanti R, Zha P, Katigbak C. The Feasibility of an Exergaming Program in Underserved Older African Americans. *West J Nurs Res* 2017 Jan 01:193945916687529. [doi: [10.1177/0193945916687529](https://doi.org/10.1177/0193945916687529)] [Medline: [28322649](https://pubmed.ncbi.nlm.nih.gov/28322649/)]
22. Xu X, Li J, Pham TP, Salmon CT, Theng Y. Improving Psychosocial Well-Being of Older Adults Through Exergaming: The Moderation Effects of Intergenerational Communication and Age Cohorts. *Games Health J* 2016 Dec;5(6):389-397. [doi: [10.1089/g4h.2016.0060](https://doi.org/10.1089/g4h.2016.0060)] [Medline: [27976953](https://pubmed.ncbi.nlm.nih.gov/27976953/)]
23. Choi S, Guo L, Kang D, Xiong S. Exergame technology and interactive interventions for elderly fall prevention: A systematic literature review. *Appl Ergon* 2017 Nov;65:570-581. [doi: [10.1016/j.apergo.2016.10.013](https://doi.org/10.1016/j.apergo.2016.10.013)] [Medline: [27825723](https://pubmed.ncbi.nlm.nih.gov/27825723/)]
24. Larsen LH, Schou L, Lund HH, Langberg H. The Physical Effect of Exergames in Healthy Elderly-A Systematic Review. *Games Health J* 2013 Aug;2(4):205-212. [doi: [10.1089/g4h.2013.0036](https://doi.org/10.1089/g4h.2013.0036)] [Medline: [26192224](https://pubmed.ncbi.nlm.nih.gov/26192224/)]
25. Tahmosybayat R, Baker K, Godfrey A, Caplan N, Barry G. A systematic review and meta-analysis of outcome measures to assess postural control in older adults who undertake exergaming. *Maturitas* 2017 Apr;98:35-45. [doi: [10.1016/j.maturitas.2017.02.003](https://doi.org/10.1016/j.maturitas.2017.02.003)] [Medline: [28274326](https://pubmed.ncbi.nlm.nih.gov/28274326/)]

26. Zeng N, Pope Z, Lee J, Gao Z. A systematic review of active video games on rehabilitative outcomes among older patients. *Journal of Sport and Health Science* 2017 Mar;6(1):33-43. [doi: [10.1016/j.jshs.2016.12.002](https://doi.org/10.1016/j.jshs.2016.12.002)]
27. Chao Y, Scherer YK, Montgomery CA. Effects of using Nintendo Wii™ exergames in older adults: a review of the literature. *J Aging Health* 2015 Apr;27(3):379-402. [doi: [10.1177/0898264314551171](https://doi.org/10.1177/0898264314551171)] [Medline: [25245519](https://pubmed.ncbi.nlm.nih.gov/25245519/)]
28. Li J, Theng Y, Foo S. Effect of Exergames on Depression: A Systematic Review and Meta-Analysis. *Cyberpsychol Behav Soc Netw* 2016 Jan;19(1):34-42. [doi: [10.1089/cyber.2015.0366](https://doi.org/10.1089/cyber.2015.0366)] [Medline: [26716638](https://pubmed.ncbi.nlm.nih.gov/26716638/)]
29. Matallaoui A, Koivisto J, Hamari J, Zarnekow R. How Effective Is “Exergamification”? A Systematic Review on the Effectiveness of Gamification Features in Exergames. 2017 Presented at: Proceedings of the 50th Annual Hawaii International Conference on System Sciences; 2017 January; Hawai'i, USA p. 3316-3325. [doi: [10.24251/HICSS.2017.402](https://doi.org/10.24251/HICSS.2017.402)]
30. World Health Organization. Definition of an older or elderly person. 2013. URL: <http://www.who.int/healthinfo/survey/ageingdefnolder/en/> [accessed 2018-06-14] [WebCite Cache ID 70A7aZ4vy]
31. Bures R. Migration and the life course: is there a retirement transition? *Int. J. Popul. Geogr* 1997 Jun;3(2):109-119. [doi: [10.1002/\(SICI\)1099-1220\(199706\)3:2<109::AID-IJPG65>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1099-1220(199706)3:2<109::AID-IJPG65>3.0.CO;2-I)]
32. Ministry of Community Development, Youth and Sports. Committee on Ageing Issues: Report on the Ageing Population. 2006. URL: https://www.msf.gov.sg/publications/Documents/CAI_report.pdf [accessed 2018-06-14] [WebCite Cache ID 70A7tFs50]
33. Scopus. URL: <https://www.scopus.com/home.uri> [accessed 2018-03-23] [WebCite Cache ID 6y7h9dUXs]
34. PlumX. URL: <https://plumanalytics.com/> [accessed 2018-03-23] [WebCite Cache ID 6y7imgjOa]
35. Altmetric. URL: <https://www.altmetric.com/> [accessed 2018-03-23] [WebCite Cache ID 6y7jIIM4A]
36. Priem J, Taraborelli D, Groth P, Neylon C. Almetrics. 2010. Almetrics: A manifesto URL: <http://altmetrics.org/manifesto/> [accessed 2018-03-23] [WebCite Cache ID 6y7jbThMt]
37. QS world university rankings. 2015. URL: <https://www.topuniversities.com/university-rankings/world-university-rankings/2015> [accessed 2018-03-23] [WebCite Cache ID 6y7jAlUeu]
38. Schmidt BJ, Carvalho N, Krynski S, Ortega C, Liberman J, Kamei ME. Studies on thyroid and hypophyseal thyrotrophic hormone (TSH) in Down syndrome. *Arq Neuropsiquiatr* 1977 Mar;35(1):1-5. [Medline: [139143](https://pubmed.ncbi.nlm.nih.gov/139143/)]
39. Jung Y, Li K, Janissa N, Gladys W, Lee K. Games for a better life: effects of playing Wii games on the well-being of seniors in a long-term care facility. Proceedings of the Sixth Australasian Conference on Interactive Entertainment; Sydney, Australia. : ACM; 2009 Presented at: Proceedings of the Sixth Australasian Conference on Interactive Entertainment; December 17 - 19, 2009; Sydney, Australia. [doi: [10.1145/1746050.1746055](https://doi.org/10.1145/1746050.1746055)]
40. Kahlbaugh PE, Sperandio AJ, Carlson AL, Hauselt J. Effects of Playing Wii on Well-Being in the Elderly: Physical Activity, Loneliness, and Mood. *Activities, Adaptation & Aging* 2011 Oct;35(4):331-344. [doi: [10.1080/01924788.2011.625218](https://doi.org/10.1080/01924788.2011.625218)]
41. Keogh JWL, Power N, Wooller L, Lucas P, Whatman C. Physical and psychosocial function in residential aged-care elders: effect of Nintendo Wii Sports games. *J Aging Phys Act* 2014 Apr;22(2):235-244. [doi: [10.1123/japa.2012-0272](https://doi.org/10.1123/japa.2012-0272)] [Medline: [23752164](https://pubmed.ncbi.nlm.nih.gov/23752164/)]
42. Millington B. Exergaming in retirement centres and the integration of media and physical literacies. *J Aging Stud* 2015 Dec;35:160-168. [doi: [10.1016/j.jaging.2015.08.005](https://doi.org/10.1016/j.jaging.2015.08.005)] [Medline: [26568225](https://pubmed.ncbi.nlm.nih.gov/26568225/)]
43. Theng Y, Chua P, Pham T. Wii as entertainment and socialisation aids for mental and social health of the elderly. : ACM; 2012 Presented at: CHI '12 Extended Abstracts on Human Factors in Computing Systems; May 05 - 10, 2012; Austin, Texas, USA p. 691-702. [doi: [10.1145/2212776.2212840](https://doi.org/10.1145/2212776.2212840)]
44. Chao Y, Scherer YK, Montgomery CA, Wu Y, Lucke KT. Physical and psychosocial effects of Wii Fit exergames use in assisted living residents: a pilot study. *Clin Nurs Res* 2015 Dec;24(6):589-603. [doi: [10.1177/1054773814562880](https://doi.org/10.1177/1054773814562880)] [Medline: [25488422](https://pubmed.ncbi.nlm.nih.gov/25488422/)]
45. Wu Z, Li J, Theng Y. Examining the Influencing Factors of Exercise Intention Among Older Adults: A Controlled Study Between Exergame and Traditional Exercise. *Cyberpsychol Behav Soc Netw* 2015 Sep;18(9):521-527. [doi: [10.1089/cyber.2015.0065](https://doi.org/10.1089/cyber.2015.0065)] [Medline: [26348812](https://pubmed.ncbi.nlm.nih.gov/26348812/)]
46. Higgins J, Green S. The Cochrane Collaboration. 2011. Cochrane handbook for systematic reviews of interventions URL: <http://training.cochrane.org/handbook> [accessed 2018-06-14] [WebCite Cache ID 70A8eAb0M]
47. EBSCO. URL: <https://www.ebsco.com/> [accessed 2018-03-23] [WebCite Cache ID 6y7oX5Zrb]
48. Spek V, Cuijpers P, Nyklíček 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](https://doi.org/10.1017/S0033291706008944)] [Medline: [17112400](https://pubmed.ncbi.nlm.nih.gov/17112400/)]
49. Lee GR, Ishii-Kuntz M. Social interaction, loneliness, and emotional well-being among the elderly. *Res Aging* 1987 Dec;9(4):459-482. [doi: [10.1177/0164027587094001](https://doi.org/10.1177/0164027587094001)] [Medline: [3438563](https://pubmed.ncbi.nlm.nih.gov/3438563/)]
50. Alhabash S, Wise K. Playing their game: Changing stereotypes of Palestinians and Israelis through videogame play. *New Media & Society* 2014 Mar 02;17(8):1358-1376. [doi: [10.1177/1461444814525010](https://doi.org/10.1177/1461444814525010)]
51. Ruggiero D. The effect of a persuasive social impact game on affective learning and attitude. *Computers in Human Behavior* 2015 Apr;45:213-221 [FREE Full text] [doi: [10.1016/j.chb.2014.11.062](https://doi.org/10.1016/j.chb.2014.11.062)]

52. American COSM, Chodzko-Zajko WJ, Proctor DN, Fiatarone SMA, Minson CT, Nigg CR, et al. American College of Sports Medicine position stand. Exercise and physical activity for older adults. *Med Sci Sports Exerc* 2009 Jul;41(7):1510-1530. [doi: [10.1249/MSS.0b013e3181a0c95c](https://doi.org/10.1249/MSS.0b013e3181a0c95c)] [Medline: [19516148](https://pubmed.ncbi.nlm.nih.gov/19516148/)]
53. Rosso AL, Taylor JA, Tabb LP, Michael YL. Mobility, disability, and social engagement in older adults. *J Aging Health* 2013 Jun;25(4):617-637 [FREE Full text] [doi: [10.1177/0898264313482489](https://doi.org/10.1177/0898264313482489)] [Medline: [23548944](https://pubmed.ncbi.nlm.nih.gov/23548944/)]

Abbreviations

UCLA: University of California, Los Angeles

Edited by G Eysenbach; submitted 23.03.18; peer-reviewed by L Li, S Parsekar; comments to author 11.04.18; accepted 01.05.18; published 28.06.18.

Please cite as:

Li J, Erdt M, Chen L, Cao Y, Lee SQ, Theng YL

The Social Effects of Exergames on Older Adults: Systematic Review and Metric Analysis

J Med Internet Res 2018;20(6):e10486

URL: <http://www.jmir.org/2018/6/e10486/>

doi: [10.2196/10486](https://doi.org/10.2196/10486)

PMID: [29954727](https://pubmed.ncbi.nlm.nih.gov/29954727/)

©Jinhui Li, Mojisola Erdt, Luxi Chen, Yuanyuan Cao, Shan-Qi Lee, Yin-Leng Theng. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.06.2018. 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.

Viewpoint

Using Social Media to Target Cancer Prevention in Young Adults: Viewpoint

Urmimala Sarkar¹, MD, MPH; Gem M Le¹, MHS, PhD; Courtney R Lyles¹, PhD; Danielle Ramo², PhD; Eleni Linos³, MD; Kirsten Bibbins-Domingo^{1,4}, MD, PhD

¹Center for Vulnerable Populations, Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA, United States

²Weill Institute of Neurosciences, Department of Psychiatry, University of California, San Francisco, San Francisco, CA, United States

³Department of Dermatology, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

⁴Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:

Urmimala Sarkar, MD, MPH

Center for Vulnerable Populations

Division of General Internal Medicine

University of California, San Francisco

Building 10, 3rd Floor

1001 Potrero Ave

San Francisco, CA, 94110

United States

Phone: 1 415 206 4273

Fax: 1 415 206 5586

Email: urmimala.sarkar@ucsf.edu

Abstract

Focusing on primary cancer prevention can reduce its incidence. Changing health behaviors is critical to cancer prevention. Modifiable cancer risk factors include lifestyle behaviors related to vaccination, physical activity, weight control and maintenance, alcohol consumption, and tobacco use. These health habits are often formed in young adulthood, a life stage which currently intersects with the growing population of *digital natives* whose childhood occurred in the internet era. Social media is a critical communication medium to reach this population of digital natives. Using a life course perspective, the purpose of this viewpoint paper is to describe the current landscape of nascent research using social media to target cancer prevention efforts in young adults and propose future directions to strengthen the scientific knowledge supporting social media strategies to promote cancer prevention behaviors. Leveraging social media as a health promotion tool is a promising strategy to impact modifiable behavioral risk factors for cancer and warrants further research on developing effective communication strategies in young adults to prevent cancer in the future generations.

(*J Med Internet Res* 2018;20(6):e203) doi:[10.2196/jmir.8882](https://doi.org/10.2196/jmir.8882)

KEYWORDS

cancer; prevention & control; young adult; behavior; social media

Introduction

Cancer is a leading cause of death in the United States and a major growing public health burden. Primary prevention is an important strategy of focus as the burgeoning scientific research supports the notion that a large portion of cancer is preventable [1,2]. Although the etiology of cancer is multifactorial and complex and differs across specific types of cancer, it has been well established that approximately 50% to 60% of all cancers can be reduced with behavior change such as vaccination, physical activity, weight control and maintenance, reducing

alcohol consumption, and smoking cessation [3,4]. Given this context, it is critical for public health efforts to prioritize the fostering of positive health behaviors to reduce the future burden of cancer.

Many of these health behaviors are considered modifiable risk factors, and to an extent, may be more susceptible to change and influence during critical age periods over one's life course. Cancer prevention efforts have traditionally focused on older adults aged 40 years and over, who tend to be eligible for most cancer screenings and have more health awareness as they naturally experience more health issues with aging. However,

much less attention has been paid to cancer prevention strategies targeted to younger age demographics, such as those aged 18-29 years, and, in particular, to strategies tailored through the use of new media. It is imperative to target young adults to promote cancer prevention behaviors before cancer develops. This younger age group is a critical developmental period that can set the stage for forming mindsets and worldviews that will ultimately shape future health habits and lifestyles [5,6]. Although cancer does not commonly occur in this age group, it is important to focus on prevention earlier in life, as cancer exposures are generally thought to occur earlier in life and contribute to cancers that are more commonly diagnosed among those 40 years and older (eg, lung, breast, colorectal, and prostate). Cancer prevention behaviors include these upstream behaviors, which can be modified earlier in life and directly relevant to young adults, as well as the more proximal action of completing recommended cancer screening, which is generally not relevant to young adults for the most common cancers (breast, prostate, and colorectal cancers).

The generation of young adults born from 1995 onwards are considered *digital natives* and defined as people “born or brought up during the age of digital technology and therefore familiar with computers and the internet from an early age” [7]. Young adults aged 18-29 years are the most frequent users of social media; in 2016, 86% of them used at least one social media site [8] and 92% engaged with 2 or more devices simultaneously including mobile phones, tablets, PC, and TV [9]. Social media must be considered as a public health strategy in young adults, simply because it is embedded in their everyday lives. To effectively reach them, health communication must occur where they are, engaging in online platforms, and must also be tailored using effective cancer prevention messaging uniquely suited for particular online platforms. For example, Twitter messages are limited to 280 characters and cancer prevention messaging to younger populations must take into design the linguistic and cultural factors in how to effectively communicate and engage young adults through Twitter.

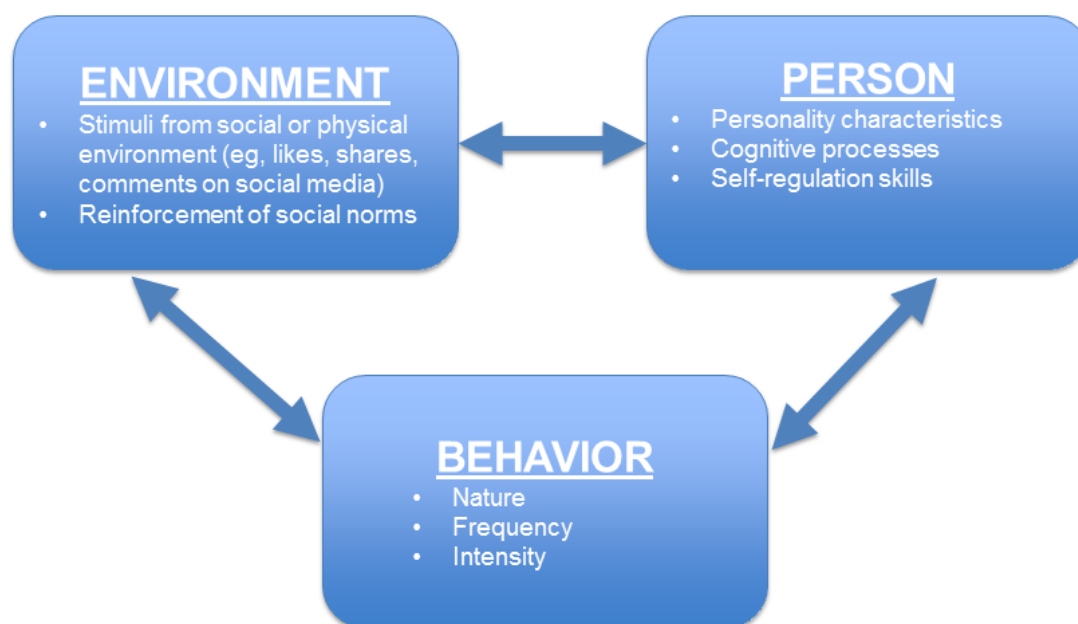
In this viewpoint paper, we focus on social media and past use in primary cancer prevention in the general population and discuss how these studies can be applied to young adults to reduce the burden of cancer in the next generation of older adults. We reflect on the current state of the field and offer discussion on how previous research has implications for considering measurement and theoretical issues in future directions of research. Specifically, we provide an example of theoretical considerations from our current work (Lyson et al. Social media as a tool to promote health awareness: results from an online cervical cancer prevention study. Under review, submitted April 2018), describe various types of studies using social media for health communication with young adult digital natives with supporting examples, highlight methodological considerations in conducting studies in this field, and propose to integrate the life course perspective of cancer prevention with

new forms of media, both of which overlap in the focus on young adults and lifestyle behavior change to present a unique opportunity for researchers to test effective cancer prevention strategies using social media.

Theoretical Considerations

Theoretical considerations are an important component in conducting rigorous research in social media and health. Specifically, behavior change interventions are most effectively guided and tested by conceptual frameworks appropriate for the target audience. As an example, in our past and current work, Bandura’s social cognitive theory, an interpersonal-level health behavior theory [10], has been the most relevant theory to apply to research questions focused on social media influences on health behaviors. This theory encompasses social influences on health in a wide variety of settings and can naturally be extended to the social media environment. Social cognitive theory is used to explain how people learn behaviors by observing others and through vicarious reinforcement. It emphasizes reciprocal causation of behaviors between the self and society, in which personal factors in the form of cognitive, affective, and biological events, behavioral patterns, and environmental events all operate as interacting determinants that influence each other bidirectionally, that is, “reciprocal determinism” (Figure 1). As part of the environment, Web-based social media frames and reinforces social norms; social media sites have their own “rules” for reinforcement of messages and content in terms of likes, shares, and comments that are much more explicit than in everyday life.

When applied to social media communication, social cognitive theory suggests that new ideas, values, behavior patterns, and social practices are rapidly diffused worldwide through observational learning, in part through social networks. The concept of reciprocal determinism is critical to behavior change via Web-based social networks. Not only do individuals learn facts and information from social media but they are also actively shaping the social media sites to be broader networks for social change or political movements through their participation. This reciprocity sets the stage for peer-to-peer influence, as in studies in which groups interact via Web-based social media to address health issues. Furthermore, social media enriches the availability of public health data in the environment; in Bandura’s model, social media provides a “socially mediated pathway” to disseminate communication by linking people to social networks and community settings that provide natural incentives and continued personalized guidance for desired change. The social media activities of public health organizations, such as vaccination campaigns from the Centers for Disease Control (CDC) delivered via Twitter, allow for dissemination and reinforcement of health behaviors. The concept of “observational learning,” that individuals learn from watching others perform a given behavior, informs how behavior can spread via Web-based social media.

Figure 1. Reciprocal determinism in Bandura's social cognitive theory for behavior change.**Figure 1. Social Cognitive Theory
Applied to Research in Social Media and Health:
Reciprocal Determinism**

Current Research in Social Media and Health

Public health research using social media takes place on a spectrum ranging from using social media as a real-time data source to engaging target populations online to influence health behaviors. Automated analysis of passively collected social media data can be used for disease or behavioral surveillance, including for early identification of disease outbreaks [11]. Public health organizations also deliver health information and health promotion messages using social media. In fact, the CDC has a social media toolkit intended to facilitate public health communication efforts via social media by partners and stakeholders [12]. This approach is unidirectional; experts deliver content to lay participants. The assumption in this approach is that populations at risk are willing and able to engage with health-related content and subsequently modify behavior. More recently, public health researchers have used social media to deliver health interventions that harness the immediacy of Web-based communication as well as the influence of Web-based social networks [13]. In social media intervention research, researchers interact with participants online, and participants may interact with each other. To augment our viewpoint discussion, we highlight various study designs that have been employed using social media data sources, provide supporting examples from the literature, and discuss implications for future research in social media and health.

Observational Studies Using Social Media Data Sources

Because individuals, especially young adults, publicly share health information online, social media data can provide a robust

data source for behaviors that are difficult to characterize and health data that are unavailable through traditional surveillance methods. This method of “mining” social media data for public health purposes is perhaps the most widely developed type of social media health research [14]. This type of observational research may be less prone to bias as people on social media typically do not act as though they are being observed for the purpose of research, in contrast to traditional research methods that explicitly recruit people to participate in research in academic settings and ask people to report on health-related behaviors. Myriad examples of this type of work exist across disparate public health domains including substance use [15], body weight-associated stigma [16], and infectious disease surveillance [11,17]. For example, Lyles et al performed this observational type of analysis for cervical cancer prevention discussions among young women on Twitter [18]. The analysis demonstrated that women do share publicly their experiences with cervical cancer screening, often with language encouraging peers to undergo screening as well. These user-generated health promotion messages are useful for characterizing public sentiment and informing public health messaging content. More recently, we analyzed Instagram data to characterize misuse of codeine on social media and found that codeine misuse was commonly represented with the ingestion of alcohol, cannabis, and/or benzodiazepines [19]. Our findings suggested that codeine misuse was represented as normalized behavior and found in mainstream commercialization of music and cartoons on social media. Because health behaviors are often difficult to capture in traditional observational research studies that rely on self-reported survey data, social media provides a unique lens through which stigmatized behaviors can be observed through a “fly on the wall” perspective.

This literature demonstrates that public health professionals can learn about community perceptions of cancer prevention-relevant behaviors by examining social media content. There is still much unrealized potential for connecting social media content and sentiment to real-world health behaviors. Thus far, one effective use of social media data has been in the area of “infoveillance” such as in influenza forecasting [20] and real-time outbreak identification [21]. Using geocode tags from social media data content could likewise be used to geographically pinpoint challenges and opportunities in cancer prevention behaviors; this methodology has been previously applied to infectious disease outbreak research. However, as a typical methodological concern for all self-reported data, the information on the user’s location is largely based on what is provided in their user profile, which may not be complete or accurate information. For Twitter data, it is estimated that about 1% to 2% of tweets are shown to be geotagged [22].

Observational studies using social media data have the advantage of accessing vast amounts of public data readily available in real time. This immediacy is a major advantage of using social media data to inform public health surveillance. However, methodological challenges remain in conducting rigorous and unbiased studies using social media data. Social media data are user-driven data and depend on the population who chooses to publicly share information. This is a self-selected group and may not represent the general population. Access to the internet and privacy concerns influence the likelihood of posting information online [23]. Internet access, particularly on mobile devices, is growing rapidly among young adults, and mobile internet is well suited to social media use. Privacy concerns are common, but younger adults compared with older adults are more likely to have shared personal information online [23], potentially enhancing generalizability in this age group. A second limitation of social media content as a public health data source is its unstructured nature, making comparisons across platforms or even individual messages challenging. Moreover, it is often impossible to verify the identity or other relevant details about individuals who post online. In general, social media posts often lack identification, demographic information, and other details. Social media data analysis must be interpreted in light of these inherent limitations.

Unidirectional Mass Communication Health Promotion via Web-Based Social Media

Governmental organizations such as the CDC and the National Cancer Institute have used social media marketing strategies to deliver a wide array of health promotion content through multiple dissemination channels and platforms, such as blogs, Twitter, and Facebook [24]. Researchers have also used an online marketing approach for cancer prevention. As an example, Cidre-Serrano et al used Google AdWords to display skin cancer prevention messages on individuals’ search results page when users searched for tanning beds [25]. These prevention advertisements were displayed over 200,000 times over 2 months with a click-through ratio of 1%, which is generally considered sufficient for commercial purposes. Google for NonProfits and other Web-based platforms provide a limited amount of free advertising for nonprofit organizations, making

this a low-cost approach for qualifying organizations. In general, the unidirectional strategy of “pushing” content at individuals has the advantages of being low-cost with a significant reach, as well as the ability to target content to specific high-risk groups (eg, young women who use tanning beds). However, data are lacking about the effect of health promotion messages delivered online. An example of planned work to address this gap would be to learn whether a Facebook advertising-based intervention aimed at reducing indoor tanning would shift knowledge and attitudes about indoor tanning and reduce individual intent to use tanning beds to ultimately prevent melanoma in high-risk groups.

Web-based social media is a powerful advertising and marketing tool as 88% of businesses use social media [26]; however, commercial entities have been shown to use social media to promote unhealthy behaviors. For example, Ricklefs et al documented the indoor tanning industry’s use of social media as a strategy for maintaining relationships with customers and to offer pricing deals that promote high-frequency tanning [27]. Similarly, e-cigarette advertising is prevalent on Twitter, particularly in states that limit other forms of tobacco advertising [28].

Provision of public health information or promotion via social media is subject to many of the same limitations as mass-media public health campaigns [29]. Social media messages are well integrated into the lives of users and can be easily accessible when they need it the most. The potential for health campaigns to go “viral,” increasing the audience size and impact, is a theoretical advantage of social media campaigns compared with traditional approaches, but it cannot be predicted or planned. Inasmuch as content is easily accessible, it is, however, also easy to turn off. As with billboards, it remains unclear whether health content is reaching its intended audience. On social media, for example, many public health and medical professionals follow CDC on Twitter, but the extent of dissemination to the lay public is unknown. As with all unidirectional public health messaging, it is challenging to accurately assess the effect of such campaigns on health outcomes amidst the many other health influences in the individual’s environment. To measure the effectiveness of public health messages, innovative sampling and methods and proxy outcomes may be needed.

Web-Based Social Media Interventions for Cancer Prevention Behaviors

Social media can also be used as a delivery platform for conducting intervention studies aimed at promoting health and wellness [13,30]. Web-based interventions have significant advantages: cost, ease of participation, and ability to scale up. These interventions can also harness the interaction dynamics of Web-based social networks and create positive peer-to-peer momentum for behavior change. For instance, in the area of smoking cessation, the Tobacco Status Project (TSP) is a Facebook intervention for young adult smokers combining messaging, peer-to-peer interaction, online counseling sessions, and group cognitive-behavioral sessions. A feasibility trial achieved 72% follow-up rates and an 18% rate of reported 7-day abstinence at 12 months (9% verified) [31]. Importantly,

engagement in the intervention was high, with 92% participation in the full 3-month intervention [32]. These results demonstrate that Web-based social media platforms can be used to deliver behavioral interventions; however, the content, mode of delivery, and network structure all require careful planning and evaluation [33,34]. A clinical trial testing the efficacy of TSP on biochemically verified smoking abstinence is underway [35]. We believe that conducting interventions via social media platforms requires further study to understand the specific components that contribute to intervention effectiveness, such as the ideal intensity/timing/duration of the intervention, how/which Web-based social networks to access (general social networking vs disease-specific sites), the mix of peer-to-peer versus expert support for behavior change, how to escalate to “real-life” interventions such as pharmacologic treatment for tobacco (eg, nicotine replacement), and how to address Web-based misinformation and foster trust of information.

One of the major challenges in social media research is the rapid pace at which social media platforms evolve online and gain and lose popularity for certain segments of society. For example, Facebook has gained more users in the older age groups and has lost favor with younger age groups who have migrated to other platforms such as Snapchat. Research involving specific social media platforms can quickly become outdated as it can take several years for research studies to be funded, implemented, and ultimately published. This can be a frustrating challenge for researchers engaged in social media and health studies; although there are no easy solutions to this, there are possibilities to reframe the research questions to be more platform-agnostic and thus more widely applicable to the understanding how social media affects health behaviors. A more conceptual approach, driven by conceptual frameworks, to the research question can shed insights on constructs underlying social interactions that influence health behavior, as opposed to relying on specific platform. In considering the choice of platforms, researchers should prepare to be nimble and course-correct if they realize that the target audience or research question does not match the intended platform. Funded research should consider alternate platforms as part their research strategy and anticipate potential problems and alternative solutions to meet the needs of the research question.

Although social media interventions have the significant advantage of reaching people where they are, more complex health behaviors such as quitting smoking may require more intensive interventions beyond online social interactions. For instance, replacing in-person tobacco cessation counseling with online counseling allows participants to receive content without consuming transportation time, and at their convenience; however, there is a concern that delivering interventions online may dilute their effectiveness, especially because of the lack of

personal connection. Moreover, many evidence-based interventions developed to be delivered in-person or via telephone require significant adaptation for Web-based social media [33], and reach, efficacy, and implementation may differ significantly. Future studies should incorporate rigorous methodological approaches in the design and evaluation of social media interventions by drawing on appropriate conceptual frameworks and evaluation methods from implementation sciences [36] regardless of whether they are newly developed or are adopted from existing interventions, because the “rules of engagement” online are so different from traditional health intervention environments.

Measuring outcomes is a methodological challenge in all types of studies. For social media and health research, there are various ways in which outcome measures can be obtained: (1) enrolling participants online and obtaining informed consent to follow participants for behavior change, (2) partnering with platforms to examine online actions (social media analytics such as click-throughs, page-viewing behaviors, purchases, etc), and (3) partnering with health systems for data linkage and online/clinic-hybrid interventions (linkage with electronic health records). These approaches combine traditional research methods of data collection (ie, direct data collection from participants through surveys) with innovative partnerships with social media platforms and health systems to provide a more comprehensive collection of outcome data to ascertain intervention effectiveness.

Conclusion

In this new era of communication, social media has tremendous potential to improve public health as it has permeated society across all socioeconomic strata and races/ethnicities [37]. Young adults comprise a diverse population on social media, which has implications for addressing future disparities in cancer. The range of research described in this viewpoint paper harnesses a variety of disciplines, ranging from data science to social science. There is a need to ensure that multidisciplinary research teams have the appropriate expertise to conduct the research; the team’s composition should be driven by the expertise needed for the proposed research questions (data science, disease-specific/clinical expertise, behavioral science, communication sciences, public health professionals, social marketing experts, and qualitative and quantitative methods). Furthermore, research is needed to understand the effects as well as risks of using social media for cancer prevention in young adults to determine the impact on reducing the future burden of cancer. Use of social media as a health promotion tool seems most relevant to modifiable behavioral risk factors in young adults and warrants further research to prevent cancer in the next generation.

Acknowledgments

The authors would like to acknowledge funding support from NCI (K24CA212294) for the preparation of this manuscript.

Conflicts of Interest

DR has consulted to Carrot, Inc, which makes a tobacco cessation device.

References

1. Song M, Giovannucci E. Preventable incidence and mortality of carcinoma associated with lifestyle factors among white adults in the United States. *J Am Med Assoc Oncol* 2016 Sep 01;2(9):1154-1161 [FREE Full text] [doi: [10.1001/jamaoncol.2016.0843](https://doi.org/10.1001/jamaoncol.2016.0843)] [Medline: [27196525](https://pubmed.ncbi.nlm.nih.gov/27196525/)]
2. Colditz GA, Sutcliffe S. The preventability of cancer: stacking the deck. *J Am Med Assoc Oncol* 2016 Sep 01;2(9):1131-1133. [doi: [10.1001/jamaoncol.2016.0889](https://doi.org/10.1001/jamaoncol.2016.0889)] [Medline: [27195793](https://pubmed.ncbi.nlm.nih.gov/27195793/)]
3. Colditz GA, Sellers TA, Trapido E. Epidemiology - identifying the causes and preventability of cancer? *Nat Rev Cancer* 2006 Dec;6(1):75-83. [doi: [10.1038/nrc1784](https://doi.org/10.1038/nrc1784)] [Medline: [16372016](https://pubmed.ncbi.nlm.nih.gov/16372016/)]
4. Colditz GA, Wei EK. Preventability of cancer: the relative contributions of biologic and social and physical environmental determinants of cancer mortality. *Annu Rev Public Health* 2012 Apr;33:137-156 [FREE Full text] [doi: [10.1146/annurev-publhealth-031811-124627](https://doi.org/10.1146/annurev-publhealth-031811-124627)] [Medline: [22224878](https://pubmed.ncbi.nlm.nih.gov/22224878/)]
5. Lassi ZS, Salam RA, Das JK, Wazny K, Bhutta ZA. An unfinished agenda on adolescent health: opportunities for interventions. *Semin Perinatol* 2015 Aug;39(5):353-360. [doi: [10.1053/j.semperi.2015.06.005](https://doi.org/10.1053/j.semperi.2015.06.005)] [Medline: [26162972](https://pubmed.ncbi.nlm.nih.gov/26162972/)]
6. White MC, Peipins LA, Watson M, Trivers KF, Holman DM, Rodriguez JL. Cancer prevention for the next generation. *J Adolesc Health* 2013 May;52(5 Suppl):S1-S7 [FREE Full text] [doi: [10.1016/j.jadohealth.2013.02.016](https://doi.org/10.1016/j.jadohealth.2013.02.016)] [Medline: [23601606](https://pubmed.ncbi.nlm.nih.gov/23601606/)]
7. Oxford Dictionaries. Digital native URL: https://en.oxforddictionaries.com/definition/digital_native [accessed 2018-04-30] [WebCite Cache ID 6z4STUd7S]
8. Pew Research Center. 2017. Social Media Fact Sheet URL: <http://www.pewinternet.org/fact-sheet/social-media/> [accessed 2018-04-30] [WebCite Cache ID 6z4SmiP6d]
9. Think with Google.: Google; 2012 Nov. The Media Habits of Teens and Young Adults Infographic URL: <https://www.thinkwithgoogle.com/consumer-insights/the-media-habits-of-teens-and-young-adults/> [accessed 2018-04-30] [WebCite Cache ID 6z4SxQCCm]
10. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. New Jersey: Prentice-Hall; 1986.
11. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. *PLoS Curr* 2014;6:1-2 [FREE Full text] [doi: [10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117](https://doi.org/10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117)] [Medline: [25642377](https://pubmed.ncbi.nlm.nih.gov/25642377/)]
12. Centers for Disease Control and Prevention. 2011. The Health Communicator's Social Media Toolkit URL: https://www.cdc.gov/healthcommunication/ToolsTemplates/SocialMediaToolkit_BM.pdf [accessed 2018-04-30] [WebCite Cache ID 6z4TDNKZ1]
13. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. *J Med Internet Res* 2014;16(2):e40 [FREE Full text] [doi: [10.2196/jmir.2952](https://doi.org/10.2196/jmir.2952)] [Medline: [24550083](https://pubmed.ncbi.nlm.nih.gov/24550083/)]
14. Kim AE, Hansen HM, Murphy J, Richards AK, Duke J, Allen JA. Methodological considerations in analyzing Twitter data. *J Natl Cancer Inst Monogr* 2013 Dec;2013(47):140-146. [doi: [10.1093/jncimonographs/igt026](https://doi.org/10.1093/jncimonographs/igt026)] [Medline: [24395983](https://pubmed.ncbi.nlm.nih.gov/24395983/)]
15. Chan B, Lopez A, Sarkar U. The canary in the coal mine tweets: social media reveals public perceptions of non-medical use of opioids. *PLoS One* 2015;10(8):e0135072 [FREE Full text] [doi: [10.1371/journal.pone.0135072](https://doi.org/10.1371/journal.pone.0135072)] [Medline: [26252774](https://pubmed.ncbi.nlm.nih.gov/26252774/)]
16. Chou WY, Prestin A, Kunath S. Obesity in social media: a mixed methods analysis. *Transl Behav Med* 2014 Sep;4(3):314-323 [FREE Full text] [doi: [10.1007/s13142-014-0256-1](https://doi.org/10.1007/s13142-014-0256-1)] [Medline: [25264470](https://pubmed.ncbi.nlm.nih.gov/25264470/)]
17. Araz OM, Bentley D, Muelleman RL. Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska. *Am J Emerg Med* 2014 Sep;32(9):1016-1023. [doi: [10.1016/j.ajem.2014.05.052](https://doi.org/10.1016/j.ajem.2014.05.052)] [Medline: [25037278](https://pubmed.ncbi.nlm.nih.gov/25037278/)]
18. Lyles CR, Godbehre A, Le G, El Ghaoui L, Sarkar U. Applying sparse machine learning methods to twitter: analysis of the 2012 change in pap smear guidelines. A sequential mixed-methods study. *JMIR Public Health Surveill* 2016 Jun 10;2(1):e21 [FREE Full text] [doi: [10.2196/publichealth.5308](https://doi.org/10.2196/publichealth.5308)] [Medline: [27288093](https://pubmed.ncbi.nlm.nih.gov/27288093/)]
19. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of codeine misuse on Instagram: content analysis. *JMIR Public Health Surveill* 2018 Mar 20;4(1):e22 [FREE Full text] [doi: [10.2196/publichealth.8144](https://doi.org/10.2196/publichealth.8144)] [Medline: [29559422](https://pubmed.ncbi.nlm.nih.gov/29559422/)]
20. Malik MT, Gumel A, Thompson LH, Strome T, Mahmud SM. "Google flu trends" and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. *Can J Public Health* 2011 Aug;102(4):294-297. [Medline: [21913587](https://pubmed.ncbi.nlm.nih.gov/21913587/)]
21. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One* 2013;8(12):e83672 [FREE Full text] [doi: [10.1371/journal.pone.0083672](https://doi.org/10.1371/journal.pone.0083672)] [Medline: [24349542](https://pubmed.ncbi.nlm.nih.gov/24349542/)]
22. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. Menlo Park, CA: Association for the Advancement of Artificial Intelligence; 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media; July 8, 2013; Cambridge, Massachusetts p. 400-408.
23. Madden M. Pew Research Center. 2012. Privacy management on social media sites URL: <http://www.pewinternet.org/2012/02/24/privacy-management-on-social-media-sites/> [accessed 2018-04-30] [WebCite Cache ID 6z4Vc93nY]

24. Miller AS. The zombie apocalypse: the viral impact of social media marketing on health. *J Consum Health Internet* 2013 Oct;17(4):362-368. [doi: [10.1080/15398285.2013.833447](https://doi.org/10.1080/15398285.2013.833447)]
25. Cidre Serrano W, Chren MM, Resneck Jr JS, Aji NN, Pagoto S, Linos E. Online advertising for cancer prevention: Google ads and tanning beds. *J Am Med Assoc Dermatol* 2016 Jan;152(1):101-102. [doi: [10.1001/jamadermatol.2015.3998](https://doi.org/10.1001/jamadermatol.2015.3998)] [Medline: [26444220](https://pubmed.ncbi.nlm.nih.gov/26444220/)]
26. Bennett S. Adweek. 2014 Oct 20. 88% of Companies are Using Social Media for Marketing [STUDY] URL: <http://www.adweek.com/digital/social-media-companies/> [accessed 2018-04-30] [WebCite Cache ID 6z4VmkXEy]
27. Ricklefs CA, Asdigian NL, Kalra HL, Mayer JA, Dellavalle RP, Holman DM, et al. Indoor tanning promotions on social media in six US cities #UVTanning #tanning. *Transl Behav Med* 2016 Jun;6(2):260-270 [FREE Full text] [doi: [10.1007/s13142-015-0378-0](https://doi.org/10.1007/s13142-015-0378-0)] [Medline: [27356996](https://pubmed.ncbi.nlm.nih.gov/27356996/)]
28. Dai H, Deem MJ, Hao J. Geographic variations in electronic cigarette advertisements on Twitter in the United States. *Int J Public Health* 2017 May;62(4):479-487. [doi: [10.1007/s00038-016-0906-9](https://doi.org/10.1007/s00038-016-0906-9)] [Medline: [27742923](https://pubmed.ncbi.nlm.nih.gov/27742923/)]
29. Evans WD, Uhrig J, Davis K, McCormack L. Efficacy methods to evaluate health communication and marketing campaigns. *J Health Commun* 2009 Jun;14(4):315-330. [doi: [10.1080/10810730902872234](https://doi.org/10.1080/10810730902872234)] [Medline: [19466645](https://pubmed.ncbi.nlm.nih.gov/19466645/)]
30. Valente TW. *Social Networks and Health: Models, Methods, and Applications*. New York, NY: Oxford University Press; 2010.
31. 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](https://doi.org/10.2196/jmir.5209)] [Medline: [26721211](https://pubmed.ncbi.nlm.nih.gov/26721211/)]
32. Thrul J, Klein AB, Ramo DE. Smoking cessation intervention on Facebook: which content generates the best engagement? *J Med Internet Res* 2015 Nov 11;17(11):e244. [Medline: [26561529](https://pubmed.ncbi.nlm.nih.gov/26561529/)]
33. 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](https://doi.org/10.2196/jmir.5086)] [Medline: [26825969](https://pubmed.ncbi.nlm.nih.gov/26825969/)]
34. Centola D. The spread of behavior in an online social network experiment. *Science* 2010 Sep 3;329(5996):1194-1197 [FREE Full text] [doi: [10.1126/science.1185231](https://doi.org/10.1126/science.1185231)] [Medline: [20813952](https://pubmed.ncbi.nlm.nih.gov/20813952/)]
35. Ramo DE, Thrul J, Delucchi KL, Ling PM, Hall SM, Prochaska JJ. The Tobacco Status Project (TSP): Study protocol for a randomized controlled trial of a Facebook smoking cessation intervention for young adults. *BMC Public Health* 2015 Sep 15;15:897 [FREE Full text] [doi: [10.1186/s12889-015-2217-0](https://doi.org/10.1186/s12889-015-2217-0)] [Medline: [26374203](https://pubmed.ncbi.nlm.nih.gov/26374203/)]
36. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999 Sep;89(9):1322-1327. [Medline: [10474547](https://pubmed.ncbi.nlm.nih.gov/10474547/)]
37. Pew Research Center. 2014. Social Networking Fact Sheet URL: <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/> [accessed 2018-04-30] [WebCite Cache ID 6z4WFELA0]

Abbreviations

CDC: Centers for Disease Control

TSP: Tobacco Status Project

Edited by G Eysenbach; submitted 05.09.17; peer-reviewed by G Daal, A Koskan, A Leader, T Fazzino; comments to author 29.11.17; revised version received 01.03.18; accepted 10.03.18; published 05.06.18.

Please cite as:

Sarkar U, Le GM, Lyles CR, Ramo D, Linos E, Bibbins-Domingo K
Using Social Media to Target Cancer Prevention in Young Adults: Viewpoint
J Med Internet Res 2018;20(6):e203
URL: <http://www.jmir.org/2018/6/e203/>
doi: [10.2196/jmir.8882](https://doi.org/10.2196/jmir.8882)
PMID:

©Urmimala Sarkar, Gem M Le, Courtney R Lyles, Danielle Ramo, Eleni Linos, Kirsten Bibbins-Domingo. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.06.2018. 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.

Tutorial

Tutorial for Using Control Systems Engineering to Optimize Adaptive Mobile Health Interventions

Eric B Hekler^{1,2}, PhD; Daniel E Rivera³, PhD; Cesar A Martin^{3,4}, PhD; Sayali S Phatak², MS; Mohammad T Freigoun³, MS; Elizabeth Korinek², PhD; Predrag Klasnja^{5,6}, PhD; Marc A Adams², PhD; Matthew P Buman², PhD

¹Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA, United States

²School of Nutrition & Health Promotion, Arizona State University, Phoenix, AZ, United States

³School for Engineering of Matter, Transport, and Energy, Ira A Fulton Schools of Engineering, Arizona State University, Tempe, AZ, United States

⁴Facultad de Ingenieria en Electricidad y Computacion, Escuela Superior Politecnica del Litoral (ESPOL Polytechnic University), Guayaquil, Ecuador

⁵Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

⁶School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:

Eric B Hekler, PhD

Department of Family Medicine & Public Health

University of California, San Diego

9500 Gilman Drive

Atkinson Hall (Mail Code 0811, Office 6113)

La Jolla, CA, 92093-0811

United States

Phone: 1 858 822 7482

Fax: 1 858 534 9338

Email: ehekler@ucsd.edu

Abstract

Background: Adaptive behavioral interventions are individualized interventions that vary support based on a person's evolving needs. Digital technologies enable these adaptive interventions to function at scale. Adaptive interventions show great promise for producing better results compared with static interventions related to health outcomes. Our central thesis is that adaptive interventions are more likely to succeed at helping individuals meet and maintain behavioral targets if its elements can be iteratively improved via data-driven testing (ie, optimization). Control systems engineering is a discipline focused on decision making in systems that change over time and has a wealth of methods that could be useful for optimizing adaptive interventions.

Objective: The purpose of this paper was to provide an introductory tutorial on when and what to do when using control systems engineering for designing and optimizing adaptive mobile health (mHealth) behavioral interventions.

Overview: We start with a review of the need for optimization, building on the multiphase optimization strategy (MOST). We then provide an overview of control systems engineering, followed by attributes of problems that are well matched to control engineering. Key steps in the development and optimization of an adaptive intervention from a control engineering perspective are then summarized, with a focus on why, what, and when to do subtasks in each step.

Implications: Control engineering offers exciting opportunities for optimizing individualization and adaptation elements of adaptive interventions. Arguably, the time is now for control systems engineers and behavioral and health scientists to partner to advance interventions that can be individualized, adaptive, and scalable. This tutorial should aid in creating the bridge between these communities.

(*J Med Internet Res* 2018;20(6):e214) doi:[10.2196/jmir.8622](https://doi.org/10.2196/jmir.8622)

KEYWORDS

adaptive interventions; mHealth; eHealth; digital health; control systems engineering; behavior change; optimization; multiphase optimization strategy; physical activity; behavioral maintenance

Introduction

Background

Overwhelming evidence suggests health behaviors such as smoking, physical activity (PA), and diet are key to preventing noncommunicable diseases such as many forms of cancer, heart disease, and diabetes [1-4]. Across interventions (eg, human-delivered and community-based), statistically significant changes in health behaviors relative to control can be found, but these differences rarely meet clinical targets such as 10,000 steps/day for PA, particularly when focused on behavioral maintenance [5-7]. Mobile health (mHealth) interventions show promise for promoting behavior change [8], but further work is needed to realize their potential for meeting and maintaining behavioral and clinical targets. To accomplish the goal of meeting and maintaining clinically meaningful targets, many have argued for adaptive mHealth interventions that are individualized and vary the intervention based on an individual's evolving needs [9-18].

Adaptive interventions are complex interventions [19], which, like static interventions (meaning interventions that are delivered the same way to everyone and do not adjust provision of support over time), often include multiple active ingredient components meant to facilitate behavior change, such as goal setting or problem solving. Adaptive interventions include additional elements [9,11,12]. Since adaptive interventions adjust provision of support over time, an additional element is *decision points*, which are the meaningful windows of time when the selection of intervention type or dose (henceforth labeled intervention option) occurs (eg, daily or monthly). Adaptive interventions also include *tailoring variables*, which are the baseline (eg, demographics) and time-varying information (eg, stress, affect, and weather) that informs intervention option selection at each decision point. Finally, *decision rules* operationalize the adaptation by specifying which intervention option to select at a given decision point based on known information such as tailoring variables.

For example, we have been developing an adaptive PA intervention, *Just Walk*, which includes goal setting, positive reinforcement, and self-monitoring components [20-22]. The end goal for this intervention is to help individuals meet and maintain PA guidelines of 10,000 steps per day by developing an intervention that is responsive to the idiosyncratic and dynamic nature of steps (see Case Study Overview section for more details). *Just Walk* includes a target of daily decision making and, thus, the decision point is each morning. *Just Walk* includes multiple tailoring variables (eg, stress, mood, weather, and self-efficacy) that can be used to inform the decision made at each daily decision point. One intervention component within *Just Walk* is a suggested daily step goal that can be adjusted each day depending on a person's changing needs. A second intervention component is positive reinforcement for achieving goals, which, in this case, involves provision of points that translate into gift cards. For this component, available points can vary each day, thus enabling a mechanism for increasing motivation to meet a given goal on any given day.

As this example illustrates, there are many elements within this seemingly simple adaptive intervention. The central thesis of this work is that adaptive interventions are more likely to succeed at helping individuals meet and maintain behavioral targets if its elements can be iteratively improved via data-driven testing of the elements. The classic evaluation strategy for behavioral interventions is the randomized controlled trial (RCT). An RCT provides information about whether an intervention package can produce an effect relative to a meaningful comparator (eg, current clinical best practice) but limited information about how, when, where, and for whom each element functions to produce the desired effect. As such, an RCT does not provide sufficient insights for supporting data-driven improvement (also called optimization) of the elements of an adaptive intervention such as *Just Walk*.

Control systems engineering is a field that focuses on decision making in systems that change over time. Control engineering is pervasive (eg, pacemakers, climate control, and robotics) but often goes unnoticed as a hidden technology [23] and to date, has only been minimally applied for use in testing and improving behavioral interventions [24-27]. The methods of control systems engineering are well suited to iteratively improving elements of adaptive interventions for real-world health behavior change. For example, control engineering methods can be used to account for and test the value of multiple tailoring variables simultaneously when selecting interventions and can adapt frequently (eg, every second, minute, hour, and day).

The purpose of this paper was to provide an introductory tutorial on when and what to do when using control systems engineering for designing and optimizing adaptive mHealth behavioral interventions. We start with a review of the need for optimization, building on the multiphase optimization strategy (MOST) [28]. Next, we provide a brief overview of control systems engineering with a particular focus on defining key terms and highlighting the general logic that guides control systems engineering. Following this, we describe attributes of problems that are well matched to control engineering, and then we summarize steps to take to design and optimize an adaptive intervention via control systems engineering. We ground this tutorial in our on-going case study, *Just Walk*.

Optimization: Unpacking Complex Interventions

In classic RCT's, all elements are combined into a unified package relative to another package. On the basis of this, limited information about each element, such as the tailoring variables to use for individualization or the decision rules to use for adaptation, is available. If an intervention package produces suboptimal results, it will be difficult, empirically, to localize what elements or interaction between elements could be further improved upon to produce a more potent intervention within RCTs.

Collins et al have been pioneering MOST, which provides structure for thinking about optimization of complex interventions [28,29]. MOST is a comprehensive, principled-engineering-inspired framework for optimizing and evaluating behavioral interventions. The framework includes an RCT to conduct summative evaluations of an optimized complex intervention relative to a meaningful comparator, such as current

best practice interventions [30], but also includes other experimental designs for iterative improvement and thus, data-driven optimization of a behavioral intervention. Optimization is accomplished on the basis of optimization criteria. Optimization criteria include measures and clinically meaningful trade-offs such as cost, time, or minimal effectiveness targets, with success (or failure) on each metric determined before running an optimization trial. For example, one optimization criteria could be that each intervention component must be significantly better, statistically, than a comparator, and the entire intervention package must be deliverable for less than US \$500. Optimization criteria can also include constraints that limit the actions or feasible ranges of the optimization procedure. For example, an adaptive intervention component cannot make drastic changes such as a large jump in suggested step goals from one day to the next.

The most common optimization trial used in MOST (indeed, sometimes inappropriately labeled a MOST trial) is the use of a factorial or fractional factorial design [31]. This optimization trial can be used to optimize static complex interventions to, for example, eliminate ineffective components [32] or test for interaction effects between components [33] in relation to optimization criteria such as cost-effectiveness [31]. As adaptive interventions include additional elements beyond static interventions, methods are required that can support data-driven optimization of these elements.

One approach for optimizing adaptive interventions is the sequential multiple assignment randomized trial (SMART) [34]. SMART is a method that mimics clinical practice and supports the study of decision rules of adaptation, such as what to do with nonresponders. As clinical visits are often separated by weeks or even months, SMART was designed with relatively infrequent decision points (eg, once every 3 months) as plausible moments of adaptation. Furthermore, SMART can only account for relatively few tailoring variables within a given decision rule. As such, SMART is not well matched to adaptive interventions that monitor multiple tailoring variables simultaneously and with frequent decision points, such as daily, as is the case with *Just Walk*.

There is another emerging method for optimizing adaptive interventions called the microrandomization trial (MRT) [35]. MRT involves randomizing provision of support, not between individuals but, instead, at each decision point. For example, if we used MRT for our *Just Walk* intervention, we could randomize whether a suggested step goal was provided each morning or not to test for each day if a step goal increases steps for that day compared with days without a step goal. There are great opportunities for optimizing adaptive interventions via MRT, particularly when coupled with adaptation strategies that are broadly derived from the computer science method called reinforcement learning (RL) [36]. In particular, MRT and RL are well matched to the emerging intervention class called “just-in-time adaptive interventions,” which provide support during “just-in-time” states, meaning, when a person has the opportunity to engage in a positive behavior (or vulnerability to engage in a negative behavior) and the receptivity to want to be provided support [11].

As demonstrated by our publication record [35], we are supportive of the MRT approach. With that said, we contend that there is great opportunity for taking advantage of the rich history and methods from control engineering when optimizing adaptive mHealth interventions. These methods are complementary for optimizing adaptive interventions that, we argue, should both be part of the repertoire of optimization trial methods that health and behavioral scientists could use for optimizing adaptive interventions, what Almiral et al have called the “optimization toolkit” [37,38]. In the remainder of the paper, we highlight the unique value of control systems engineering for optimizing adaptive interventions.

Control Systems Engineering Overview

Control engineering has a long history dating back nearly a century and is pervasive (eg, pacemakers, artificial pancreas systems, and supply management) [23]. Control systems engineering focuses on decision making in systems that change over time. An algorithm called a controller defines decision rules (often called policies in control parlance) that attempt to balance mathematical equations related to predicted error, which, in this context is deviation from a desired state. For example, a desired state may be 10,000 steps/day, but if a person currently walks 6000 steps/day, then the error is 4000 steps/day. *Controllers* perform the same task as classic tailoring decision rules [11,12] but with important differences. Classic tailoring uses if-then structures, such as if stage of change=X then Intervention=A; if stage of change=Y then intervention=B [9,10]. Although controllers can use if-then structures, they can use other structures, particularly mathematical equations and optimization algorithms that can account for multiple tailoring variables, intervention options, and responses of the person simultaneously.

One can think of this like accounting. The controller keeps a ledger of measurements. In all controllers, including nonmodel-based controllers such as Proportional-Integral-Derivative (PID) controllers [39,40], this ledger includes measurements of provision of intervention options, called *controlled input variables* in control parlance, and outcome variable(s) (or outputs) that can define how close a system (eg, a person) is in relation to the desired state (error), particularly in response to intervention options. Decisions are made based on the dynamic interrelationship between the intervention options and outcome measure in the past (P portion of a PID controller), present (I portion of a PID controller), or the anticipated rate of change in the future (D; note some controllers only include parts of this such as I or PI controllers). For sake of clarity, we label this class of controller as nonmodel-based controllers.

In more advanced controllers that include a dynamical model, such as model-predictive controllers [41], other variables are also measured, including (1) *inputs*, which include endogenous variables that influence the outcome variables (eg, stress and self-efficacy) and (2) *disturbance variables*, which are exogenous variables (eg, weather) the system cannot control and are not attributes of the person but impact the state of the person and, thus, could influence intervention option selection. In the ledger for model-based controllers, not only are

intervention options and outcomes tracked, but their interrelationships are defined via a *dynamical model*, which is like a structural equation model but can model dynamics and incorporate a wider range of response options via difference and differential equations [15]. With this information, model-based controllers can simultaneously monitor a wide array of important issues related to individualization and adaptation, such as variables that are particularly influential (eg, one person's steps are influenced by stress and another by day of the week [22]) or a person's changing responsiveness to interventions, such as an intervention option only being useful for a limited time via the novelty effect [42].

Both model and nonmodel-based controllers conduct a series of simulations to predict responses to intervention options in the near and distant future, either based on deviations between the desired state and the state of the system alone (nonmodel-based controllers) or via dynamical models (model-based controllers). These forecasts are used to make decisions. The intervention option predicted to most likely foster movement toward the desired state within prespecified constraints (eg, only small daily changes to step goals allowed) is selected. In contrast, if-then rules require knowledge of the match between tailoring variables and intervention options before specification [43]. This difference means that, relative to if-then structures, mathematical equations can manage more complex decision environments (eg, more tailoring variables and interventions options), can function with limited a priori knowledge about an individual, and can perform when a person's responses fall outside of expectations and thus, are feasibly more responsive to each individual's changing needs.

Control engineering includes a wealth of methods for optimizing adaptive interventions by managing and mitigating lack of knowledge related to intervention elements. Lack of knowledge can take various forms from sensor noise (eg, measurement noise when inferring steps [44,45]) to incorrect models (eg, inaccurate predictions). *System identification* is an experimental and analytic suite of methods to generate or validate dynamical models for future predictions, [46-49] or, to put it in more behavioral terms, it can be used for rigorous theory testing. System identification "excites" variance with a person via plausible intervention options to test what happens in different states and contexts of the person over time. For example, if a control engineer wished to generate a dynamical model to understand factors that impact a person's steps, she may vary a person's daily suggested step goals in different states, such as different days of the week or when stressed vs not [50]. System identification can occur using both open loop and closed loop experimentation. An open loop experiment is "open loop" because the intervention options that are provided to a person are specified a priori and, thus, a person's responsiveness to the intervention options are *not* taken into account when selecting future intervention options. When a person's data are taken into account to adjust support, this is called a closed loop experiment. *Dynamical systems modeling* analyzes what occurred following the intervention options over time during different conditions to generate a dynamical model for each person and, ideally, a generic dynamical model structure such as a *semiphysical model*, which is useful across individuals.

Key concepts related to testing controllers are performance and robustness. Performance involves how well the controller can produce the desired effect as efficiently as possible. Robustness involves how well the controller can produce desired performance when issues such as poor measurement, models, or interventions or changing responsiveness to interventions arise [43]. It is quite common for controllers with high performance to be less robust vs robust controllers to have poorer performance (eg, take longer to achieve the desired state). As such, a central focus of controller design and testing is to define the right balance between performance and robustness, which can occur via *closed loop experimentation* and *robustness testing* [43,50,51].

A closed loop experiment can be used to test the controller in relation to optimization targets, such as meeting and maintaining PA guidelines. It is closed loop because, like in closed loop system identification, a person's response to each selected intervention option provided is documented and then taken into account when selecting the next intervention option, thus closing the loop. This type of experiment can include a variety of strategies to test the controller. For example, one could systematically vary providing the predicted optimal vs nonoptimal intervention option to test the controller, if appropriate for the research and intervention. This sort of strategy maps on to the computer science concept of exploring vs exploiting [36]. Exploring involves including some randomness to see what will happen when a predicted nonoptimal option was provided. Exploitation, in contrast, involves using all that is known about a person to select the predicted optimal intervention option. Thus, comparing explore vs exploit options is one way to test controllers, particularly related to performance. In contrast, robustness testing [43] involves examining how well the controller can function when issues such as poor measurement or models arise and, thus, is complementary and often done in tandem with closed loop experimentation.

Within control systems engineering, it is common to use all of these methods (ie, system identification, closed loop experimentation, robustness testing) within a single system or individual. In particular, system identification experiments (ie, theory testing) and closed loop experiments (ie, testing with controller actions present to support, among other things, testing of the controller) can be offered sequentially to a single person and, indeed, decision rules can be defined on when to switch from one method to the next. For example, a closed loop experiment might be used to test a controller striving toward helping a person to meet PA guidelines. If the person meets the behavioral target for a prespecified time (eg, 2 weeks), this could trigger the switch to different optimization criteria, such as targeting maintenance of steps and minimization of interactions between the intervention and person (ie, a second controller optimization algorithm for maintenance). The combined study that includes system identification, closed loop experimentation, and robustness testing is what we call a *control optimization trial* that can balance the competing demands of performance (eg, quickly helping a person meet goals) vs robustness (eg, being responsive to individual differences and changing needs). These methods enable a rigorous and efficient

approach to optimize elements of an adaptive intervention for each individual.

Attributes of Problems That Are Well-Suited to Control Systems Engineering

In this section, we describe attributes of problems that are well-matched to control engineering. We ground our discussion within the concrete case study of *Just Walk*.

Case Study Overview: Just Walk

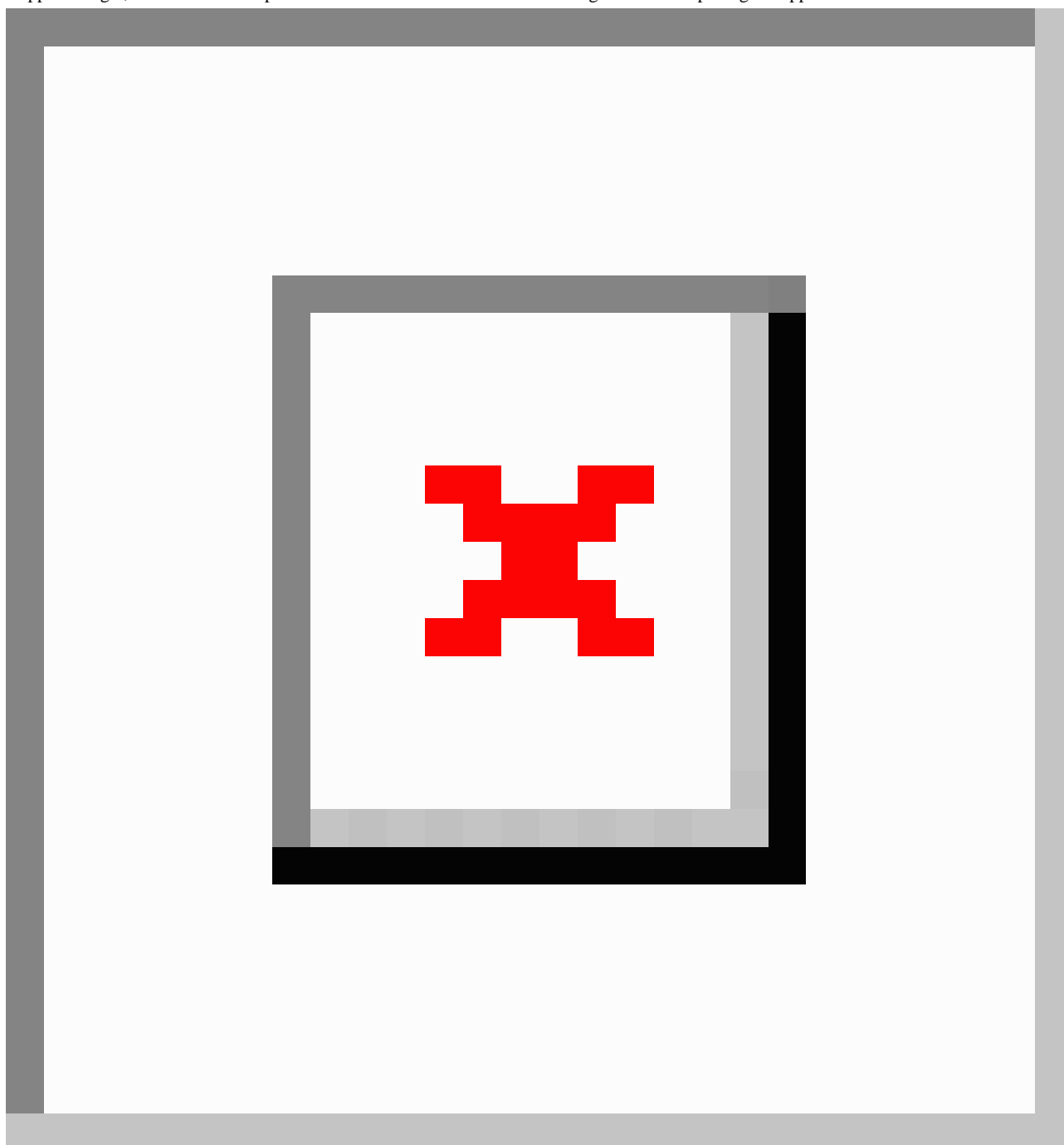
Convincing evidence indicates PA is valuable for reducing risk of certain types of cancer [52,53], cardiovascular disease [54], and for improving glycemic control [55]. Walking or taking steps is important for all adults but in particular, those who are sedentary, overweight or obese, and in the age range of 40 to 65 years because they are at increased risk of chronic diseases and because this group can safely walk and fit it into their lives [56,57]. With an aging population, step interventions could help prevent chronic diseases, reduce health care costs, and improve functional life years and quality of life [52-55,58-70]. Guidelines for steps suggest 7100 to 10,000 steps/day [56,57], but only one-third of this group meet the guidelines [71-81]. Across PA interventions for older adults (eg, human-delivered and digital), results show 620 steps/day increases, which translate to individuals walking, on average, 5388 steps/day, which is below guidelines [5]. Findings are similar among healthy adults with 496 steps/day achieved, and even high-impact interventions peak at 1363 steps/day increases; both below guidelines [6]. Even among interventions that produce an effect, maintenance is rarely measured and, when it is, it is not achieved for a large number of participants [82-85]. There is a strong need for

interventions that can help individuals meet and maintain PA guidelines.

One reason meeting and maintaining PA is hard may be because of the idiosyncratic and dynamic nature of steps. Specifically, taking steps occurs in both incidental and purposeful ways [86-88] such as commuting, leisure walking, or sports and is engaged in differently by different people. Furthermore, when and where individuals fit steps in also changes (eg, weekend warriors vs evening gym rats) over time and also can vary between individuals. Our prior work [22,89] shows that individual variables (eg, stress and busyness) and contextual factors (eg, weekend or weekday) have different relationships to steps for different people. These idiosyncratic determinants change over time. Walking routines change based on a variety of factors such as small stressors (eg, pressing deadlines) to large ones (eg, changing careers and retirement) and context changes (eg, changes in season) [71-73,75,76,80]. It is also common that interventions lose their potency, thus suggesting reduced responsiveness [8,84,90,91].

We have been developing *Just Walk* as an mHealth adaptive walking intervention, specifically to account for the inherently idiosyncratic and dynamic nature of walking behavior. Our intervention includes individualized step goal suggestions, self-monitoring (measured via a wearable device), and contingent reinforcement (ie, points and gift cards) that are provided when daily goals are met. In addition, we will supplement our behavior change active ingredients via a range of engagement-supporting tools such as suggestions for weather-appropriate ways to be active. The mHealth system includes, at present, a front-end mobile phone app (Figure 1), a back-end server, and integration with wearable devices (eg, Fitbit) to objectively measure PA.

Figure 1. Screenshots of the Just Walk App. The image on the left is the view inside the app, which includes the suggested step goal for the day (in the red box), available points (in gold medal in the middle) and current steps (in green box). Below is the person's step history. The image on the right is the app's "widget," which enables a person to receive feedback relative to their goal without opening the app.



Problems That Are Well-Suited to Control Systems Engineering

We turn to a discussion on the types of problems that are well suited for control engineering (see [Textboxes 1](#) and [2](#)). We discuss each below and will use *Just Walk* to illustrate.

First, the problem is dynamic, meaning the input and output variables interact over time. Within *Just Walk*, steps/day is dynamic as it often fluctuates day to day for each person. The factors that impact how many steps a person takes, such as internal states such as stress, busyness, perceived self-efficacy and external states such as weather, also change over time. Any

self-regulatory process that can be measured frequently, such as blood pressure, weight, emotion regulation, or glucose regulation within the body, are dynamic and thus, feasibly appropriate for control engineering. Conversely, if the behavioral or clinical target used to define the problem changes slowly (eg, atherosclerotic plaque formation, mortality), then control engineering is not appropriate.

Second, interventions are available to foster movement from a less desirable to a more desirable state. As part of this, there are concrete decisions that can be made for each decision point. Note that these decisions can include providing or not providing

an intervention or more continuous intervention options (suggested daily step goal).

Within *Just Walk*, the two dynamic interventions are based on Operant Theory [92,93] and the Social Cognitive Theory (SCT) [94]; specifically, the logic of the feedback loop between antecedents, behaviors, and consequences. Within *Just Walk*, the antecedent is a suggested daily step goal, the behavior is steps/day, and the consequence is daily points, which translate into Amazon gift cards. We chose these two dynamic interventions based on past research suggesting that they can influence steps [95,96]. Conversely, if the behavioral or clinical interventions are not particularly dynamic (eg, taking a vaccine that only occurs once) or do not repeat frequently (eg, attempts to facilitate taking a flu vaccine 1x per year), then control engineering is not appropriate.

Third, the target outcome can be measured with sufficient temporal density over an extended period. In the *Just Walk* example, this requirement is met via the use of wearable sensors to track steps. This requirement is available for many of the processes listed above, such as blood pressure, weight, or glucose regulation, along with behavioral targets such as sleep and some forms of diet (eg, chewing as inferred from accelerometry). When there is a lack of a variable that can be measured repeatedly over time, then control engineering methods become less relevant. For example, lack of quality cancer risk metrics, at present, reduces the utility of control engineering for cancer prevention, except for meaningful proximal predictors such as weight or PA for some forms of cancer (at least until more proximal markers of cancer risk can be developed).

Finally, there is a need for definable desirable states for the target outcome(s), which are called *set-points* in control parlance. This is particularly important as it establishes a within-person benchmark of success for the controller and, thus, the optimization criteria for individualization (via tailoring variable selection) and adaptation (via the decision rule). It is important to note that multiple phases, which are labeled states [97] in control engineering to acknowledge the movement between states rather than to imply progression, can be defined, and each state can have its own optimization criteria. Furthermore, multiple levels of success can be defined.

Within *Just Walk*, there are two states: behavioral initiation and maintenance. Our set-point for behavioral initiation is 7100 to 10,000 steps/day based on past work [56,57]. *Just Walk* strives for either 10,000 steps/day per week or, if a person does not seem capable of meeting that goal (ie, a person starts at low steps/day and does not achieve 10,000 steps per day within 6 months), then 3000 steps/day above the person's baseline median steps is used as the set-point (which usually equates to at least 7000 steps). Within the maintenance state, these set-points are used but with added constraints. During initiation, there is a bias toward providing support, unless a person appears to be responding negatively to the intervention (eg, reduced adherence). In maintenance, *Just Walk* switches toward reducing the total number of interactions, with the ideal of no support provided when not needed. *Just Walk*, thus, does not end but, instead, adapts based on perpetual need, which, conceptually could be a highly cost-efficient approach. With the optimization criteria defined, it enables data-driven optimization for individualization and adaptation.

Textbox 1. Required attributes of a problem that are well matched to control engineering.

- Dynamic, input-outputs
- Intervention options are available
- Outcome variables are measurable (or inferable) intensively
- A meaningful target or "state" exist

Textbox 2. Desirable attributes of a problem that are well matched to control engineering.

- Frequent decision points
- Previous theory available to guide model development
- Other feasibly important variables can be intensively measured
- Theorized dynamic interrelationships between inputs or outputs (eg, feedback)

Beyond these requirements, there are several desirable attributes. First, it is advantageous to have frequent decision points, such as every hour, day, or week. Within *Just Walk*, we used a daily timescale. These daily decision points enabled the design of an efficient 12-week study (described below and [98]). Technically, it is possible to develop dynamical models with less frequent decision points; thus, control engineering can be used for stepped care decision making [99]. This longer timescale, however, establishes the need for longer system identification studies. If, for example, we had used a weekly timescale in *Just Walk*, the study would have needed to be 7-times longer.

Determination of the appropriate timescale and minimal number of decision points or observations needed can be achieved using simulation studies, such as the ones we conducted for our *Just Walk* study [15,100].

Second, it is desirable if previous knowledge about the phenomenon is available. Within our example, we used the SCT to inform measurement selection, a model structure for defining our dynamical model and the interrelationships between variables, intervention selection, and the design of our study, discussed below [100,101].

Finally, it is desirable that other variables that could impact the outcome can be measured. Within *Just Walk*, we could infer variables passively, such as weather, and ask participants to complete surveys daily with minimal burden [21].

Finally, if there are strong theoretical reasons to hypothesize feedback loops and lagged effects [102], then the suite of methods used by control engineers might be beneficial. This is because dynamical modeling can mathematically specify and thus model and account for issues such as carryover effects, lagged effects, delayed effects, or feedback loops via the use of difference and differential equations [83]. As delineated by SCT, there are multiple theorized feedback loops that can be modeled via dynamical modeling.

Steps to Take When Using Control Engineering for Adaptive Interventions

Overview

In this section, we highlight suggested steps that could be used when using control engineering methods to optimize adaptive interventions. A full review on exactly how to do each step is beyond the scope of this introductory tutorial. Instead, for each step, we define *why* the step is important, *what* specific tasks are involved in the steps, and *when* to do the step vs possibly skip the whole step or at least some tasks of the step. To provide insights on *how* to do these steps, relevant references are provided. Each step is grounded with the concrete example of our *Just Walk* intervention.

Although the use of the word “step” may imply a linear process, it often is not. For example, it can be highly advantageous to select a general theoretical model (a task within step 1) and to then define optimization criteria (a task in step 4) before moving on to creating or selecting intervention options (step 2) or to even start with optimization criteria as a definition of success, which is advocated for in agile science [103]. In line with our focus on optimizing elements of an adaptive intervention, essential to this overall process is the use of the iterative process and triangulation of methods to clarify one aspect of the adaptive intervention and then examine its impact on other aspects (see Discussion).

An important prestep is to make an initial decision on the type of controller one is targeting. Although there are many considerations involved in the selection of the appropriate controller, at a high level, selection of one controller over another largely hinges on the anticipated complexity of the dynamical system, the degree to which a model can be generated that is actually predictive or useful for making decisions based on forecasted responses, and the degree to which the dynamics can be inferred from the dynamics of one (or a relatively few number) of variables (nonmodel-based), as opposed to the response of multiple interrelated variables (model-based). If the guiding theoretical model implies a complex dynamic system that would not be well represented by monitoring only intervention options and outcomes, then a model-driven controller would likely be most appropriate. If, however, the dynamics can be picked up adequately with intervention options and outcomes, such as the direction a boat is pointed as

measured via a compass as used within a boat autopilot (a classic PID controller), then a nonmodel-based controller is appropriate. There is a lot more subtlety involved in selecting the right controller (eg, the possibility of model-based PID controllers), and interested readers can gain more insights on control options here [39,40,104]. On the basis of the complexity of behavior, we anticipate that it will almost always be best to use model-based controllers. As such, the steps below are suited for model-based controllers.

Suggested steps include the following: (1) derive a preliminary dynamical model; (2) select intervention options (ie, type, frequency, and dosages) and outcomes; (3) conduct system identification (ie, theory testing); (4) design the controller; and (5) conduct a control optimization trial (ie, intervention element testing).

Step 1: Derive a Dynamical Model

This step is important for establishing a well-specified framework for understanding the eventual adaptive intervention and guiding all subsequent work. The tasks involve first specifying a general theoretical model for guiding the work, then translating that into a dynamical model, and finally, the option of vetting this dynamical model either via simulation studies, secondary data analyses, or both.

Like in MOST, a theoretical model is used to provide structure and specification about key intervention options, outcome measures, and other variables that impact the outcome measures. It is strongly advised to almost always engage in this step as it provides the foundation for understanding predictions and decisions made within the eventual adaptive intervention. The one caveat is when an adaptive intervention is being generated when very little is known about the phenomenon, except that it is highly dynamic. When this is the case, it is often more appropriate to do noninterventional work such as conducting more naturalistic studies such as ecological momentary assessment or human-centered design work [105]. Nonetheless, we could imagine the small possibility of the need to use experimentation to gain insights about a system that is not known. This should only be done when the intervention can be done safely.

Although there is not a single way to develop a theoretical model, we suggest thinking clearly through three reference points and using each to triangulate toward, first, a theoretical model and then a dynamical model. These three references are (1) prior theories, particularly those that have been well-validated in the literature among the target group; (2) prior empirical work about what works in general and other key variables to be aware of for the target group; and (3) hands-on experience and interactions with the target group in the form of human-centered design methods such as interviews, observation, codesign, or prototyping, to gain insights about your target group that may not be well understood or encapsulated in prior theories or evidence. For details on exactly how to create or select an appropriate theoretical model, we suggested the following references (see [22,37,106]).

Translation of a theoretical model into a dynamical model requires far clearer specification of the prediction. For

model-based controllers, this task is required. Creating a dynamical model involves clear specification of a variety of issues such as model structure, anticipated directionality and strength of relations between variables, and anticipated dynamics of the interrelationships [97]. For more details on how to do this step, see prior work [97,107,108].

The final optional task within this step is simulation studies or analyses with secondary data to vet a dynamical model. As the previous step highlights, dynamical models often require a high degree of mathematical specification on predictions. The use of simulations, such as changing one variable to see how the other variables might respond within the system, is valuable to gauge if the dynamical model is producing the sorts of effects that would be anticipated. If the simulated changes in one variable produce effects that are not anticipated, this can be used to either check the math or check the assumptions about the problem. Either way, it improves precision and understanding on what is being hypothesized dynamically. Secondary analyses can also be valuable as data can be used to ground the predictions of the dynamical model, again, to see if the dynamical model is working according to both expectations and available evidence. For more details on how to create a dynamical model and do preliminary vetting via simulation studies and secondary analyses, see [15,97,100].

Within our *Kust Walk* example, we chose to use prior theories, Operant Theory and SCT, as one foundation for our adaptive intervention. This is based on extensive prior work illustrating the value of these related theories for supporting robust interventions among our target group. We also have ample experience working with our target group for supporting PA via interventions based on them [109] and thus have prior evidence and user interactions. As such, we grounded our model selection based on the three references of prior theory, evidence, and insights from our target group. As this example also suggests, we hypothesized that the dynamics for understanding steps would be best understood using a model-based controller as opposed to nonmodel-based controllers. On the basis of this, we translated SCT into a dynamical model, with full steps and details about this process described in Riley et al [15] and Martin et al [100]. We identified key variables (eg, self-efficacy and outcome expectancies), defined a model structure (Figure 2), and then specified, mathematically, the anticipated interrelationships between variables. After specifying these attributes of the dynamical model, we then ran a series of

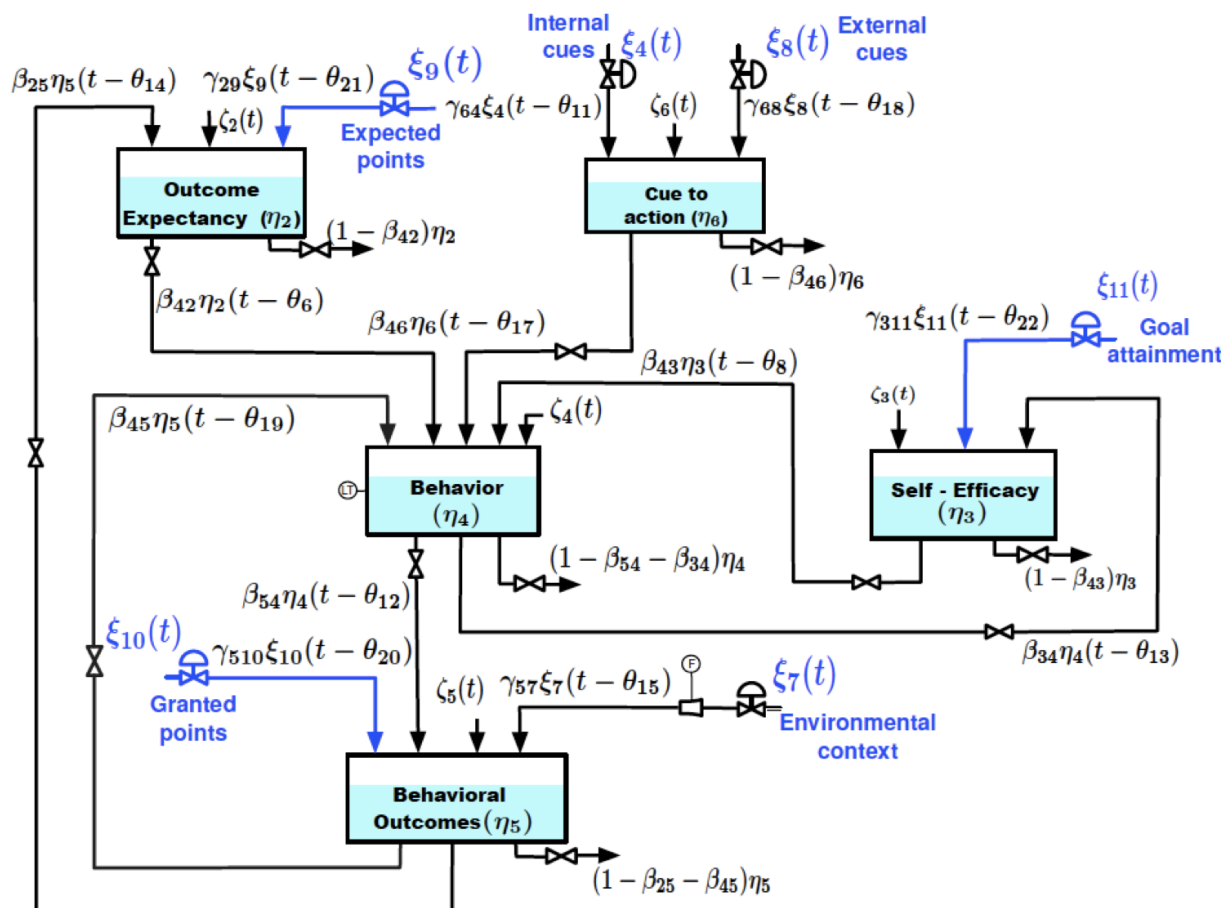
simulations of theoretical predictions to stress test the model with known psychological concepts such as habituation [15]. We also conducted secondary analyses from prior available evidence [110,111]. We decided on a simplified version of the SCT as a dynamical model to guide the rest of the process, based partially on the results of the simulation studies and secondary analyses.

Step 2: Defining Interventions Options and Outcomes

Defining target intervention options and outcome metrics are the defining features of an adaptive intervention and, thus, this step is essential. The key tasks of this step include defining the outcome metric(s) being targeted (which will be translated into optimization criteria in step 4), defining the intervention options and then, optionally, also specifying clear dynamic hypotheses on how these intervention options will dynamically interact with the person to produce desired changes to the outcome(s).

Clearly defining the outcomes is a logical follow-up step from the theory and dynamical modeling work. This is because, within the prior step, it is technically possible to do most of step 1, save the secondary analyses, without any concrete outcome measure defined (eg, steps/day or hours of sleep per night). Defining outcome metrics to target is important as it establishes a grounding on the purpose of the specific adaptive intervention. As discussed in the previous section, outcome metrics are best when they can be measured repeatedly over time to establish the current state of the target person relative to the desired final state. The intervention options can then be defined to impact the outcome metrics dynamically. These intervention options could be thought of as the essential levers the adaptive intervention can use to make adjustments and thus, facilitate movement from a less desirable to a more desirable state.

Although not required, it can be valuable to generate a dynamic hypothesis about the interrelationship between an intervention and an outcome to further ground thinking about the intervention. Although there are many ways to think about dynamic hypotheses, one way is to think in terms of outcome responsivity to the intervention options when a person is in a different state or context, including changing disease state or changes in their readiness for change (eg, stages of change). The Transtheoretical Model (TTM) establishes a basic (albeit slow) dynamic hypothesis in that different processes of change are hypothesized to be needed for different stages of change [112,113].

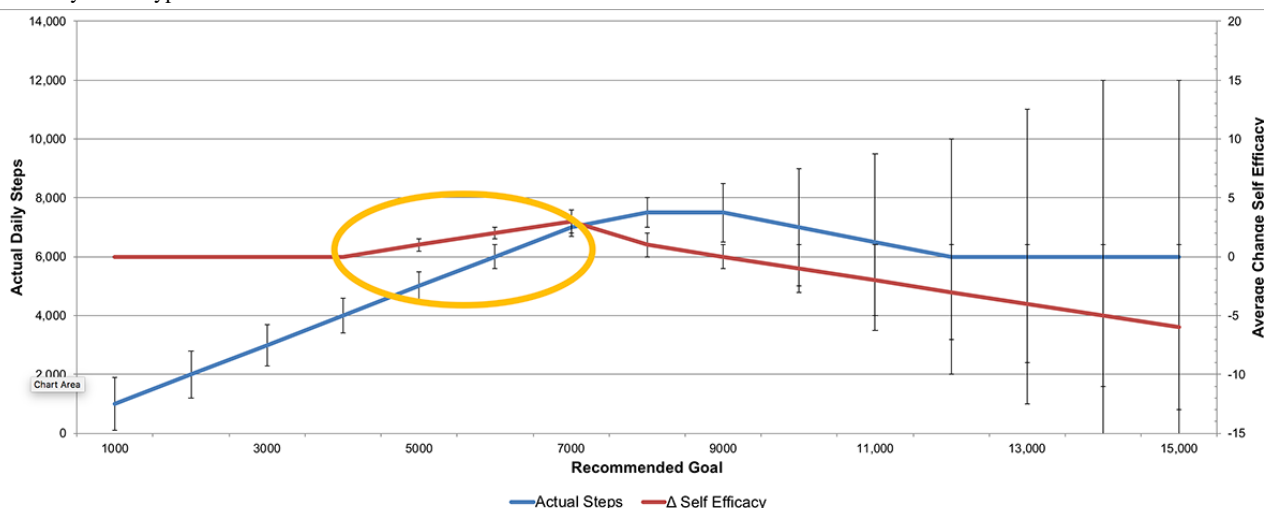
Figure 2. Simplified dynamical model version of Social Cognitive Theory.

Similarly, one might theorize that a person will respond differently to an intervention if it is provided to them when they are stressed vs not, or at home vs at work. These variations, which are further described in detail elsewhere using the modeling logic of state-spaces [97], provide a structure for thinking through the dynamic interrelationships between interventions, outcomes, and changes in the person and context over time.

These dynamic hypotheses can take various forms. A relatively simple dynamic hypothesis could be to specify if-then statements for different states or contexts of the individual, which is the implicit structure used in the TTM (eg, if stage of change=X then process of change=Y). This could be useful for stepped care interventions (eg, see [99,114-116]). As discussed earlier, control engineering uses mathematical equations for prediction to support dynamic decision making. On the basis of this, dynamic hypotheses are not required to conform to if-then statements, but instead can be defined more mathematically related to predicted changes in key variables. This latter more complex structure is what we use within *Just Walk*.

We chose to focus on individualized step goal suggestions and provision of points as our two dynamic intervention components (grounded in self-monitoring as the third, but we assumed that to be a constant intervention component). Our key outcome measure in *Just Walk* is steps per day, as measured via a

wearable device. To help define our eventual controller design (step 4), we postulated a dynamic hypothesis that can be specified mathematically but not as an if-then structure. A common hallmark of goal setting includes strategies that help a person define what might be called an *ambitious but doable* target [117]. Within *Just Walk*, we have encapsulated this mathematically as a dynamic hypothesis that is influenced by suggested goals and points (see Figure 3). The figure is meant to visualize the dynamic interrelationship between recommended step goals (x-axis), actual steps taken (y-axis, left side), and the impact on self-efficacy, on average (y-axis, right side). The yellow circle is the hypothesized *ambitious but doable* recommended step goal range that is hypothesized to be optimal for fostering increases over time in self-efficacy. Below this range, and any time a person meets their step goals, we hypothesize will not impact their self-efficacy. Above this dynamic range and we hypothesize that, on average, the person will not attain goals as regularly and, thus, with particularly high goals, will result in an, on average, reduction in self-efficacy. For the person/day represented in Figure 3, if a person's goal is below 4500 steps, we hypothesize no change in self-efficacy when the goal is attained. If the goal is too high, we expect goal attainment to happen less frequently, on average, which would result in an overall reduction in self-efficacy over time.

Figure 3. Dynamic hypothesis.

For example, if a goal of 9000 steps or greater is suggested for our Figure 3 example, goal attainment will occur less often, on average, resulting in a progressive reduction in self-efficacy as a person reduces their confidence that they can meet challenging step goals. Finally, as self-efficacy and individual and contextual factors change (eg, stress and weather), along with available points, so will the target range (eg, blue line moving up or down). This dynamic hypothesis conforms with the SCT, the rationale of graded goals, and with previous evidence suggesting high goals (ie, 10,000 steps, arguably ambitious for sedentary individuals) resulted in more variable steps [96].

Step 3: Conduct a System Identification Experiment (That is, Theory Testing)

This step is most distinct from other forms of testing and optimization within behavioral interventions. As such, we include a great deal more information here to highlight the logic and overall approach.

From a control systems perspective, the primary goal of this step is to estimate and validate dynamical models. This is valuable in and of itself, regardless of any subsequent controller, because it is explicitly focused on understanding a “system” or phenomena, such as, in this case, an individual human. To put it in more behavioral terms, system identification is a form of dynamic theory testing. System identification is also important for later steps, if there is a desire to use model-based controllers. Although not commonly the focus in control engineering, this type of study can also be used to select tailoring variables for individuals (ie, data-driven individualization), test dynamic hypotheses, or develop a benchmark comparator for optimization criteria when prior work provides limited insights on a meaningful benchmark. It is also technically possible to conduct a rigorous system identification experiment while also pilot testing aspects of the intervention and other protocols to test feasibility issues, if needed, because system identification is an inherently n-of-1 or idiographic approach, though this last strategy is not necessarily recommended.

As a reminder, within a *system identification* experiment, excitation of variance (ie, providing different intervention options) is provided to the system (in this case, a person) to test

what happens in different states and contexts of the system over time. *Dynamical systems modeling* analyzes what occurred following the intervention options over time during different conditions to generate a dynamical model for each person in a scalable fashion. System identification can occur using both open loop and closed loop experimentation but, from the perspective of system identification, these variations are used to validate the dynamical models (ie, theory testing) as opposed to testing the controller (ie, decision rule testing), which is the emphasis in step 5.

Suggested tasks that could be included in this step include the following: (1) design of the system identification experiment and analytic plan and (2) data analyses. If pilot testing of the technology is also needed, other optional tasks could include the following: (1) human-centered design work [103,106,118] to facilitate creation of a useful and usable intervention and (2) creation or selection of the technology tools needed to conduct the intervention (eg, digital health intervention platform). These optional tasks should be conducted when no prior adaptive intervention is available, but otherwise should be skipped as this step is primarily focused on theory testing, not pilot testing. If these optional tasks are conducted, current best practices for human-centered design and feasibility testing should occur [37,103,106,118,119]. The system identification experiment should be conducted when there is inadequate secondary data or theory (often thought of as first principles in control parlance) available about the topic to generate robust dynamical models, when the research question is clearly about dynamic interrelationships within a person (ie, theory testing), or when there is a clear dynamic hypothesis to test about the system or person. The system identification experiment does not necessarily need to be done if the targeted controller is not model-based, such as some forms of PID (and their derivations) controllers [39,40].

In terms of system identification, there is a rich literature, including toolkits within MATLAB (MathWorks), on procedures and best practices for the design of a system identification experiment and analytic plans [47,48]. As a system identification experiment excites variance within a system, the study design involves carefully defining intervention options with a particular

eye toward having variability of excitation. Excitation can occur both by varying the amplitude of the differences between intervention options and also the repeatability of intervention options. For example, in our *Just Walk* study, we chose meaningful ranges in terms of the “dosages,” from baseline median steps to double a person’s baseline median steps for step goals and providing between 100 to 500 points for the reinforcement component, with 500=US \$ 1. Note also that system identification experiments can include binary intervention options as well (eg, provision of support or not) [120].

With a sense of the amplitude defined, the next task involves designing for adequate excitation over time, which minimizes error for model estimation. This involves the length of a “cycle” and the number of cycles needed in a study to achieve sufficient minimization of error in estimation and validation. A cycle is a deterministic, repeatable pattern that defines provision of intervention options to an individual. Intervention options can be provisioned to mimic randomness via pseudorandom signals that can achieve the valuable properties of randomness for causal inference, while still being deterministic and, thus, repeatable (for more details see [121]). The primary purpose of a cycle is to enable both estimation and validation of dynamical models in terms of their predictive capacity across cycles. Within our *Just Walk*, it was determined that five 16-day cycles would produce sufficient excitation over time to minimize error in estimation and validation with our two interventions delivered orthogonally (discussed in greater detail below on how this was determined [121]). As a side note, it is possible to do estimation and validation with purely random signals, but pseudorandom cycles facilitate aspects of model validation [120,122,123]. The design of a system identification experiment can be done with a number of different toolkits that support simulation of estimation and validation based on different sources of noise or variance in the model [47].

Once data are collected, the process of data analyses takes place. A central logic of dynamical systems modeling, as with other aspects of control systems engineering, is triangulation. In particular, system identification toolkits (eg, those available via MATLAB) include a wide range of strategies to examine time series data produced from system identification experiments, such as different visualizations, step-response curves (ie, the unique influence of each variable on the outcome, much like a partial r^2), or model fits for both estimation and validation. Each one of these provides a different understanding on the overall reliability and validity of the dynamical models produced. As such, they are all used with the goal of defining dynamical models that work according to expectations across these tests.

Beyond the criteria used to evaluate the models, there are also different analytic techniques that can be used as part of dynamical systems modeling. For the sake of simplicity, we describe black-box dynamical modeling vs semiphenomenological or grey-box modeling. Central to these different modeling efforts is the degree to which prior theory and evidence is taken into account when defining a dynamical model structure. On one end are black box models from methods such as Auto-regressive model with exogenous input (ARX) modeling, which are much like generalized linear models. These models include no model

structure to define the interrelationships between structures beyond ordinary linear regression accounting for repeated measures. Semiphenomenological modeling, on the other hand, includes theorized model structures, predicted dynamics, and other factors that are either known or theorized to be true in terms of the interrelationship between variables. One could think of semiphenomenological modeling as a dynamical version of structural equation modeling [22,89]. In brief, theorized model structures, such as Figure 2, along with predicted dynamics (eg, feedback loops are the ways in which the relationships occur dynamically) are articulated within a mathematical model [97,124]. These models can then be compared with the initial black box models on a variety of criteria related to reliability and validity of the models for each person, such as overall model fit, which provides insights on the percentage of variance explained by each dynamical model (for more details see Study Design). This process, thus, enables a rigorous strategy for iteratively developing models of progressively improved predictive capacity for each person, while simultaneously enabling incremental theory testing. Furthermore, particularly related to theory testing, generic model structures can be defined if they prove reliable across individuals, thus providing a structure for translating insights drawn about an individual to be generalized to other individuals and also more generic theory testing and development that is grounded first in individuals rather than starting first in the aggregate.

The final task is to define *good enough* predictive capacity to establish an optimization criterion. If little to no information about what is good enough is available, the above strategy of comparing data-driven vs theory-driven models is a good start. If, however, other parameters or benchmarks are available and meaningful from the literature relevant for the problem domain, then those can be used as starting benchmarks on factors such as model fit. When good enough predictive capacity is reliably being shown across individuals (or at least a large enough portion of individuals, which also can and should be defined), this establishes justification for the development (step 4) and testing (step 5) of a model-driven controller. If not, a nonmodel-based controller could be explored, or the team should examine earlier steps in the process or other optimization trials (eg, between-person factorial trials, SMART, or MRT).

With these tasks defined in abstract, we turn to the *Just Walk* example. In our previous work [21,22,89], we conducted human-centered design work to develop an app for adults who are midlife and older, overweight, and sedentary. We then conducted a 12-week system identification open loop experiment, which is described below. In this context, because we did not have a previous platform, we decided to do the optional feasibility work. For the feasibility aim, this study design could be thought of as a modified variation of a single case experimental design, particularly an ABBBBB trial design with the “A” representing the baseline phase and each “B” representing an intervention cycle that was repeated five times [21]. This design supports testing feasibility issues including limited efficacy, which is defined as within-person changes in steps. Our results suggest (1) our intervention increased steps; (2) good demand, acceptability, implementation, and practicality; and (3) our system identification experiment

produces valuable data for dynamical models (for more details see [21]).

To support the eventual controller, we chose to run an open loop system identification experiment. This was because, although we did conduct secondary data analyses to vet our dynamical model [100], the secondary data analyses did not enable us to do rigorous estimation and validation of our dynamical model. On the basis of this, and our desire to develop an eventual model-based controller, we conducted an open loop system identification experiment [89,121,125].

In *Just Walk*, we devised our open loop system identification experiment to estimate and validate our simplified SCT dynamical model (Figure 2) and to test an approach for individualized tailoring variable selection [15,22,100]. As implied by our dynamic hypothesis (see step 2), from an excitation standpoint, this hypothesis requires that we include goals that are *doable*, *ambitious but doable*, and *too ambitious* for individuals. Furthermore, as the hypothesis includes specification that individual differences (eg, stress and busyness) and context (eg, weekday or weekend and weather) that could feasibly influence what is ambitious but doable on any given day for a person, it also established the requirement of repeated observations that are in the three broad category ranges within different states of the individual (eg, high vs low stress) and contexts (eg, weekdays vs weekends) for tailoring variable selection purposes. As states and context cannot be randomized, we instead chose to run the experiment over a 12-week period to increase the likelihood of observing variations in these individual and contextual factors in relation to different suggested step goals and for excitation purposes [121]. Beyond this, expected points was also hypothesized to interact with these other factors and thus varied over time.

A full description on the design of the study is beyond the scope of this paper but has been described elsewhere, which includes concrete strategies for achieving the equivalent of “power” calculations for an open loop system identification experiment [89,121,125]. In brief, our study design involved the pseudorandom suggestion of daily step goals and expected points one could receive if they met their goals as defined in repeated 16-day cycles (Figure 4). On the basis of analyses that are akin to power calculations but for system identification, we determined the need for a minimum of five cycles [121]. Furthermore, the use of 16-day cycles (Figure 4) minimized the risk of possible aliasing with day of the week (which would have occurred with 14-day cycles).

A full discussion on the analyses and results are beyond the scope of this paper, but interested readers can find more information at [22,89]. In brief, our preliminary analyses on estimating and validating a dynamical model for each person were encouraging both for preliminary dynamical models and the selection of tailoring variables for each person [22,89]. These models produce dynamic daily predictions of steps relative to actual steps (see Figure 5, which visualizes this for one participant). Specifically, Figure 5 visualizes the dynamic interrelationship between the key variables that could be valuable for predicting steps. In this context, this included goals; available points, if points were provided (ie, goal attainment

the previous day); a person’s self-reported prediction on how busy and stressed they will be; their prediction on how typical their day will be; and if it is a weekend or weekday. The bottom portion illustrates the predicted steps (pink line) relative to actual steps (black) and suggested goals (dotted blue line). Light pink zones represent cycles that were used for estimation in this particular model, and blue represent validation cycles. Model fit for this participant was 46%, which, based on Cohen’s conventions for multiple regression, would represent a large effect in terms of percent variance explained.

Using percentage model fit as a benchmark, we conducted data-driven analyses to support optimization of the dynamical model that conceptually maps on roughly to reliability and validity. In this case, reliability and validity are estimated for our dynamical models for predicting human behavior (as indicated by model fit) and, by extension, the selection of tailoring variables. For every individual, we conducted an exhaustive search of potential variations of predictors (eg, only our manipulated inputs or up to four additional endogenous or exogenous variables as plausible tailoring variables) using an ARX approach. In line with the leave-one-out approach commonly used when cross-validating models such as PA estimation via accelerometers [126], we carried out estimation or validation using every cycle from our five-cycle system identification study as both estimation and validation data.

For selection of the model and, thus, the tailoring variables to use for each person, we chose to use multiple criteria with the first three reflective of issues of reliability [22,89] and the last more reflective of validity. We chose these criteria to increase the likelihood of finding individualized models that are reliable and, thus, are likely to remain true and appropriate outside of the current data and valid, thus predictive and useful within an eventual controller. We combined them into an approach that penalized models that did not perform as well on these dimensions. Different weights (w) were assigned to four characteristics that affect model consistency and reliability: (1) overall highest fit ($w=2$), with a higher penalty for lower fits; (2) cross-correlations between inputs ($w=2$), with higher penalty for inputs with high cross-correlation coefficients; (3) distance of the overall highest fit from the mean fit (mean % fit for all cycle combinations, for each input combination; $w=1$), with a higher penalty for larger differences; and (4) SD across models run for each participant (% fit for all cycle combinations, for each input combination; $w=1$), with higher penalty for higher variances. These weights were used to define and select models that were the best estimate in terms of reliability and validity.

We then turned to good enough validity. As these analyses are a variation of multiple linear regression, and the model fit estimate is analogous to r^2 , we chose Cohen’s conventions of explaining 3% of variance as a small effect, 13% as a medium effect, and 26% as a large effect [127]. Although there is no clear definition on good enough for individualization purposes, as, to the best of our knowledge, we are the first to do this, we chose to use the 13% medium effect as our a priori good enough marker for our best model selected for each participant. We also chose a minimum of 50% of our sample to meet this medium effect explained marker as good enough across.

Figure 4. System identification open loop experiment for Just Walk. These two signals were designed a priori using a pseudorandom signal design strategy. This strategy enabled specification of repeated 16-day cycles (delineated as different colors), which allows for robust data for estimation and validation of dynamical models.

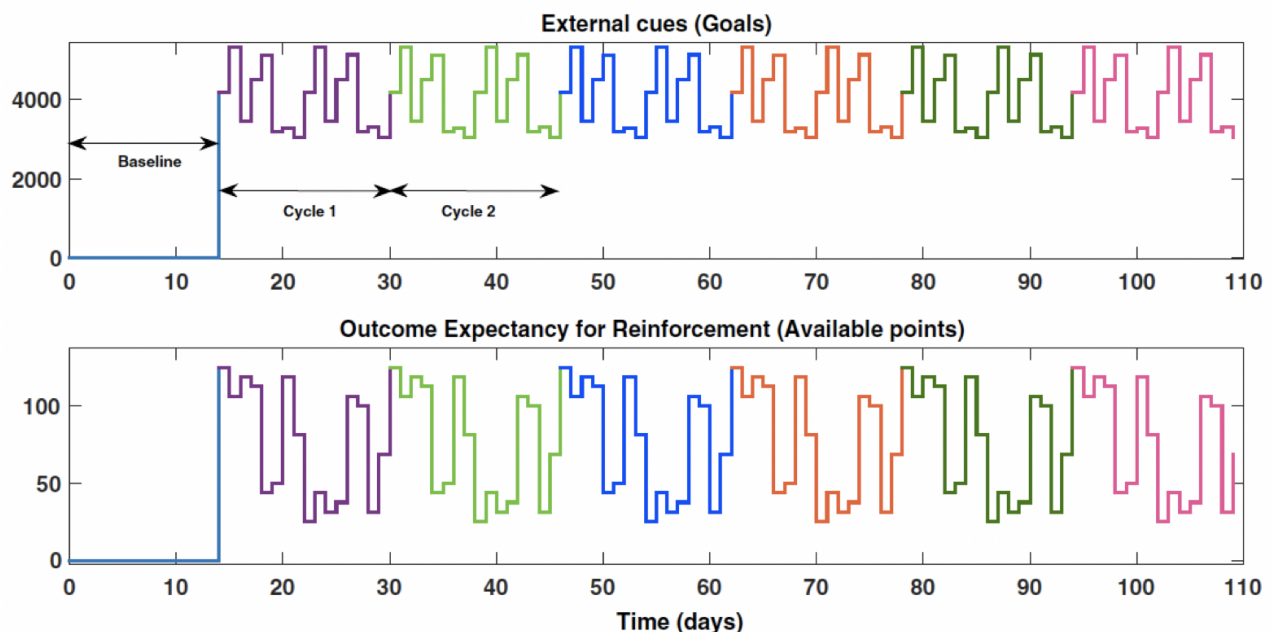
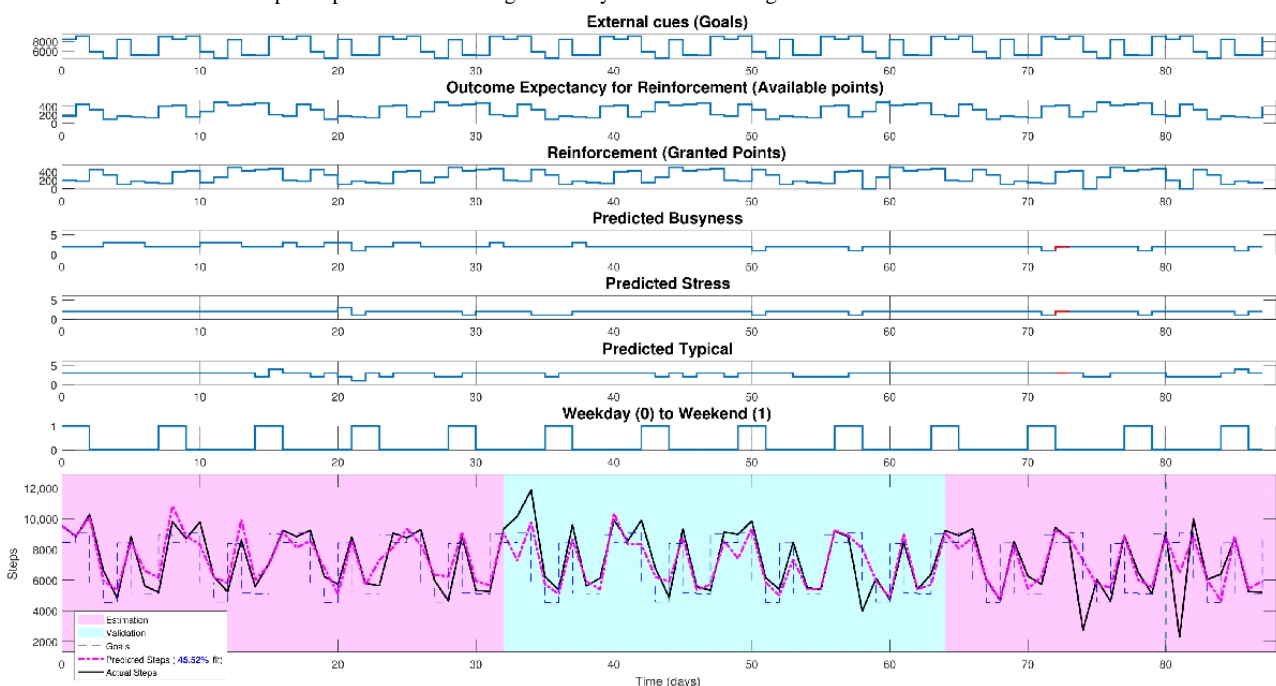


Figure 5. Visualization from one participant from Auto-Regressive Dynamical Modeling.



Note, however, that further validity testing related to individualization is possible and a core target of the more definitive optimization trial, the closed loop experiment (see below). Furthermore, we also fully acknowledge that our approach is only one of many (see Discussion). The overall average model fit (estimation and validation data) for all participants combined was 19.2% (SD 9.25). The range was 6.3% to 46%. Using Cohen's conventions, 20 out of 20 participants met the small effect threshold of explaining 3% of variance, 16 out of 20 met the 13% medium effect level, and 2 out of 20 met the 26% large effect level. On the basis of this, we achieved our good enough target of explained variance for

individuals, thus justifying subsequent steps. Although it is unclear what the minimal levels are needed for establishing robust individualization based solely on this, it does provide a preliminary indication of the ability to make distinctions between people in terms of tailoring variables. For example, using the medium effect as a minimal threshold, our approach produced meaningful individualized models for 80% of our sample [22]. From the perspective of pilot testing, we contend that this is likely an adequate target for accounting for individual differences compared with current best practice, though future work is needed to properly specify benchmarks for individualization (see Discussion).

In terms of tailoring variable selection, different tailoring variables were identified for different people [22]. In particular, the most common model included weekend or weekday as the only tailoring variable for 25% of our 20 participants. This model corresponds with the tailoring variable that would likely have been selected when using an aggregate mixed model across participants. Following this, perceived “typicality” of a day and weekend or weekday were the combined tailoring variables selected for 20% of our sample. The rest of participants had different tailoring variables selected. Put differently, if the aggregate model were used for tailoring variable selection, which would be the norm for most methods currently used related to optimizing adaptive interventions, it would have likely selected the inappropriate tailoring variable(s) for 75% of our sample with reliable models [22]. These results point both to the potential for control engineering approaches in individualized tailoring variable selection and for the need for this type of approach.

As of this writing, the team is conducting semiphysical modeling [128] to test our SCT structure (Figure 2). As stated before, the initial models we have already produced use an ARX approach that does not incorporate prior knowledge related to model structure or theorized dynamics [22,89]. As of this writing, we are using these models as our comparators for our SCT model. On the basis of theory, we should get improved model fits when we incorporate the elements of the model we specified, such as model structure, theorized dynamics, etc. If model fits do not improve, it is indicative that our theorized dynamical model structure provides no additional benefit beyond what we would have learned from the data alone for each individual. This is important to highlight as this is a second mechanism for supporting data-driven optimization related to individualization. This time though, the optimization is focused on optimizing the model structure and other theorized prior knowledge.

As this example illustrates, a great deal of valuable insights about human behavior and outcomes can be gleaned from system identification experiments. As this example also illustrates, this step can generate meaningful scientific insights as a mechanism for doing rigorous theory testing that is grounded in an individual first and then can be generalized if similar model structures are found, what we previously called data-driven case studies [124].

Step 4: Design the Controller, Including Optimization Criteria

The next step is to design the controller. This step is essential as it is the mechanism whereby prior insights can be translated into actionable dynamic decision rules (ie, the controller) for guiding an adaptive intervention. The key steps in this process include defining optimization criteria (eg, set-points), constraints of the controller (eg, clinical constraints of the intervention), and, for more complex controllers, alternative strategies the controller could use to maintain robustness to factors such as a person’s changing responsivity to an intervention. Creation of these is often supported via all of the prior work done (eg, dynamical model, intervention and outcome specification, and system identification experiment), as well as additional simulation studies specifically focused on the robustness of the

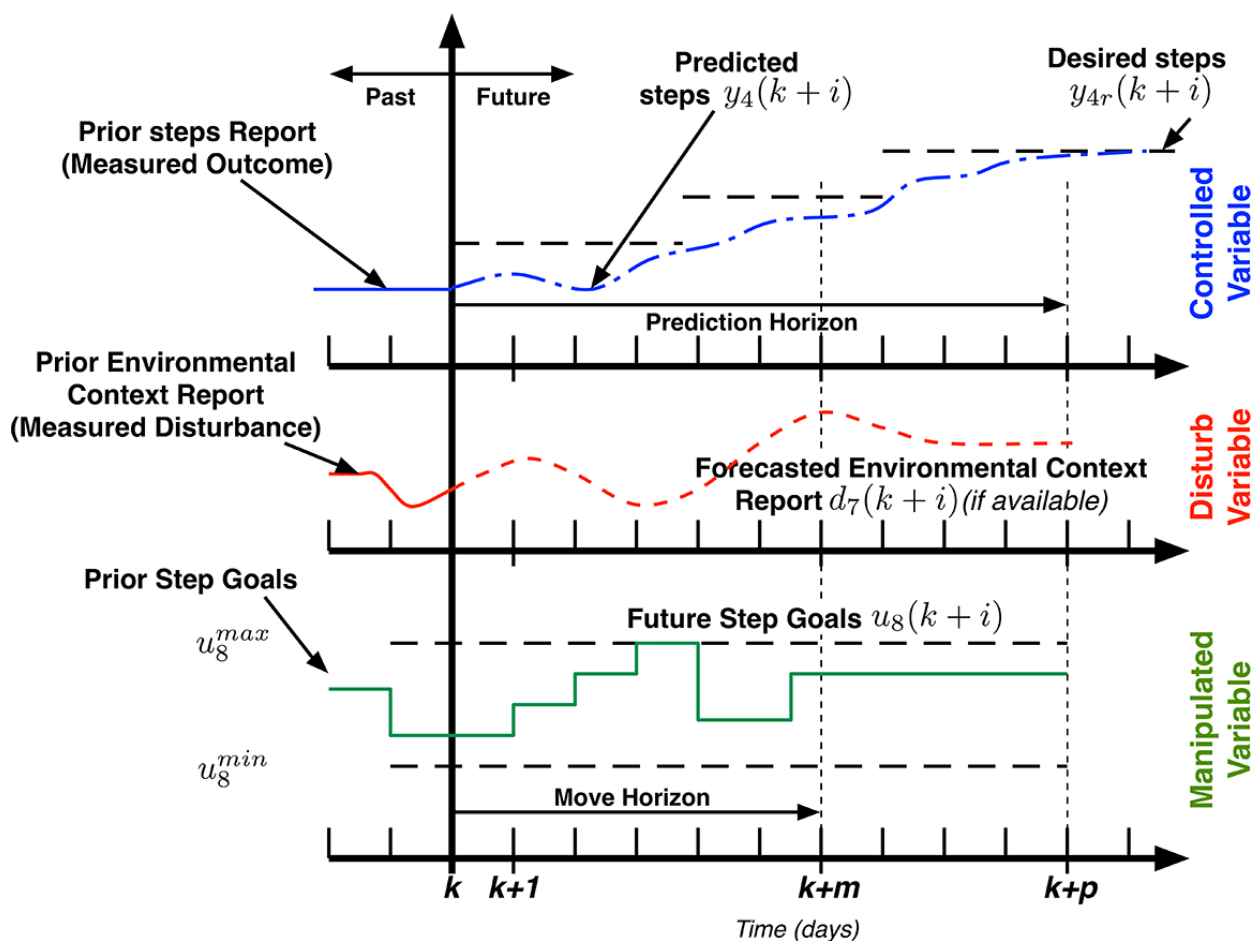
controller. These steps are done in any type of controller, including those that are not model-based.

In terms of controller design, the central focus of controller design is to define the targeted right balance between performance and robustness. Within a controller, strategies for supporting performance largely revolve around the quality of the previous steps. In particular, performance is improved when potent interventions and predictive models are available to be used by the controller. The prior work provides a foundation for anticipated performance of the intervention options and value of the dynamical model for making predictions. Strategies for maintaining robustness can be devised to help manage and mitigate these risks, which tends to be the larger focus of the controller design for this step. For interested readers, see our more detailed papers formulating our controller [50,129,130] and our strategies for facilitating robustness.

In terms of key tasks, the type of controller being targeted must be defined (eg, model-based or nonmodel based); several parameters for the controller must be defined, including the optimization criteria, constraints, and strategies for achieving robustness; and finally, simulation studies can be conducted to examine anticipated issues of robustness. The optimization criteria can be thought of as a definition of success that can be operationalized based on a measurable outcome variable. Constraints are the parameters that define what is feasible or appropriate within a given domain, such as what is safe, appropriate, or clinically viable. Finally, there are a wide range of strategies for supporting robustness. These assertions can be examined via simulation studies [129,130]. Specifically, control engineering includes methods for simulating plausible responses of controllers within different scenarios and contexts. This is valuable as it enables stress testing assumptions about the problem before recruiting participants.

Within the *Just Walk* example, the controller we chose to use is a hybrid model-predictive controller [131] that can be visualized in terms of its logic for decision making, as shown in Figure 6. As the broad goal of *Just Walk* is to help individuals meet and maintain national guidelines, we set our first optimization criterion when a person is in a state of initiating in more PA up to guidelines of 10,000 steps/day but then also included a less stringent secondary criterion of +3000 steps/day from their baseline step levels based on prior work on anticipated performance of adaptive PA interventions [132]. Prior work has illustrated that +3000 steps corresponds to approximately 30 min/day of moderate intensity activity and, thus, is still a meaningful behavioral target that has the likelihood of reducing disease risk [56,57]. We also included a second optimization criterion for those that meet guidelines and, thus, move to fostering maintenance. Once a person meets guidelines, the system reduces the total number of interactions, including goals suggested and points provided, with the target of reducing interventions to 0 except continued monitoring via the wearable device. We also established clinical constraints, including not changing suggested goals drastically (eg, by more than 4000 steps) from one day to the next. Finally, we have also formulated a variety of secondary strategies the system could take to maintain robustness in case known issues such as reduced adherence are observed.

Figure 6. Model-predictive controller “Receding Horizon” strategy. The model predictive controller visualized here is simplified to include only one controlled variable (desired daily steps), one input (ie, goals), and one disturbance (ie, environmental context). Controller moves (ie, goals) are calculated over a horizon, and only the first control move calculated is implemented. The entire procedure is repeated at the next assessment period and continues until the end of intervention.



In terms of how our controller works, the model-predictive controller forecasts changes in outcomes (ie, steps, intervention adherence) over time to determine an error projection that reflects current and expected deviations from the optimization criterion of 10,000 steps/day or +3000 steps/day from baseline. On the basis of this error projection, a real-time optimization algorithm chooses the sequence of future control actions (eg, adjusts step goal, points, and other factors) that minimizes the difference between the set-point (eg, 10,000 steps) and current steps.

The optimization problem is solved for each day considering a prediction time to obtain a predicted optimal step goal suggestion for each decision point. The first recommendation is provided, and the process repeats at each decision point. The model-predictive control strategy continually reevaluates the quality of the previous day's predictions on what was actually observed. The information can be incorporated into the model-predictive control algorithm, particularly if there are alternative strategies the controller might take based on changing observations for maintaining robustness.

We have conducted simulation studies to stress test the design of our controller. Figure 7 is a visualization of one of the simulations we ran for tests of robustness; in this case, the controller's responsivity to a person experiencing an external disturbance (eg, getting sick). This figure represents a simulation

study examining how our controller may respond, in this context, to a major unmeasured environment disturbance. As can be seen, the controller facilitates a gradual increase in steps over time using varying points. When the set-point levels have been reached, the controller switches to a maintenance phase that includes reduced suggestion of step goals (ie, last suggestion would be to maintain 10,000 steps) and reduced use of expected points for meeting the goals (ie, an expanding reinforcement schedule). As the simulation illustrates, the system would strive toward less interaction but be responsive to a person's steps falling below the set-point level to reactive initiate-phase suggestions (see day 112). For more information on the simulation work we have conducted for our controller see [129,130].

Step 5: Conduct a Control Optimization Trial

This is the key step for unpacking complex adaptive interventions via control systems engineering methods. This step can provide insights about *how*, *when*, *where*, and *for whom* each element functions to produce the desired effect and thus, is the essential strategy for unpacking a complex adaptive intervention and testing its elements. This step should thus, be done whenever the goal is to optimize an adaptive intervention via control systems engineering methods (as opposed to the other plausible adaptive intervention optimization trials). As

highlighted above, this is appropriate for the type of problem that has the attributes described in Textboxes 1 and 2.

The key tasks of this step include clear definition of the elements of the adaptive intervention, the design of subexperiments (eg, open loop system identification and closed loop experimentation) and data analysis plan to test the elements, and conducting the trial and the analyses.

As already highlighted, the key elements of an adaptive intervention include the decision points, tailoring variables (or, in this case, dynamical models), decision rules (or, in this case, the controllers), and, of course, the intervention options themselves and the meaningful proximal outcomes the intervention options target. In terms of decision points, these are often defined based on clinical intuition, such as the case in *Just Walk*, whereby our decision point was each morning. These can be tested via control engineering methods as they can be formulated, themselves, as decision rules for guiding just-in-time adaptive interventions, but that point is beyond the scope of this tutorial (and, arguably, MRT is likely more appropriate). As highlighted in step 3, system identification, particularly open loop experimentation, is a rigorous approach for optimizing the tailoring variables or dynamical models for each person and thus, a strategy for optimizing individualization. In terms of the decision rules or controllers, closed loop experimentation is the method to use to test them.

As highlighted in the introduction, it is common in control systems engineering to include multiple experiments provided sequentially, over time, to the same system (ie, person in this context). The key, from a design standpoint of the subexperiments, is to think through what is clinically appropriate or feasible and also what the logical progression is in terms of the likely changes that will occur within the target individual. In terms of the data analytic plan, as with system identification, there is a wealth of analytic strategies that are available, largely within MATLAB, for conducting the analyses. Much of the testing of controllers is actually built into the controllers themselves as, ultimately, they are mathematical equations seeking to minimize error while accounting for noise and other unknown issues. Controllers, thus, engage in self-testing relative to optimization criteria. The key advantage here of self-correction is also arguably a weakness, as this work hinges on the quality of the optimization criteria (a point we return to in the Discussion). A full description of the type of analyses that can be done and the many ways in which to design effective exploration or exploitation is beyond the scope of this introductory text, but interested readers should examine here [104]. Similarly, a full description of analyses for robustness testing is also beyond the scope, but readers can learn more here [43].

Figure 7. Controller simulation.

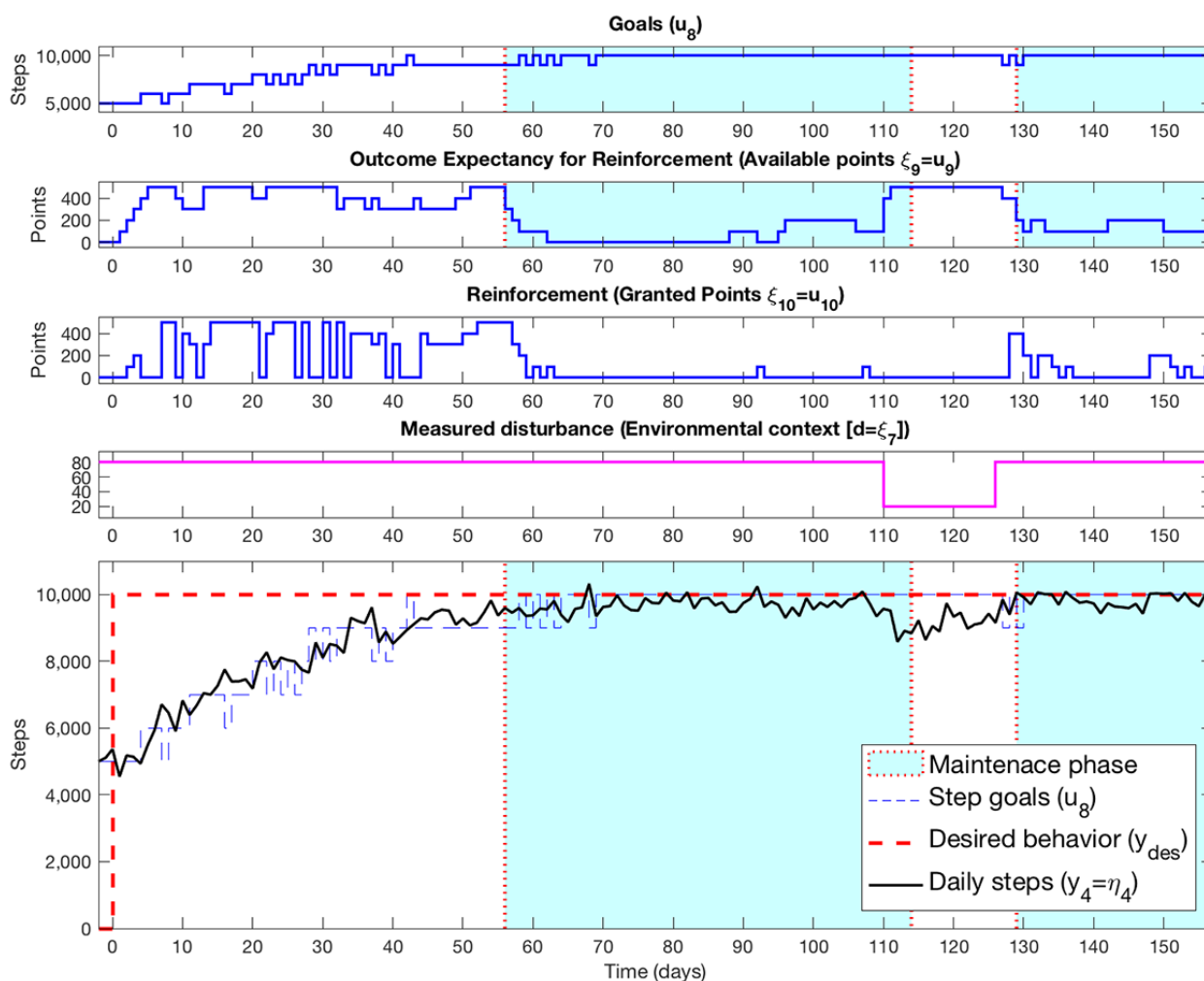
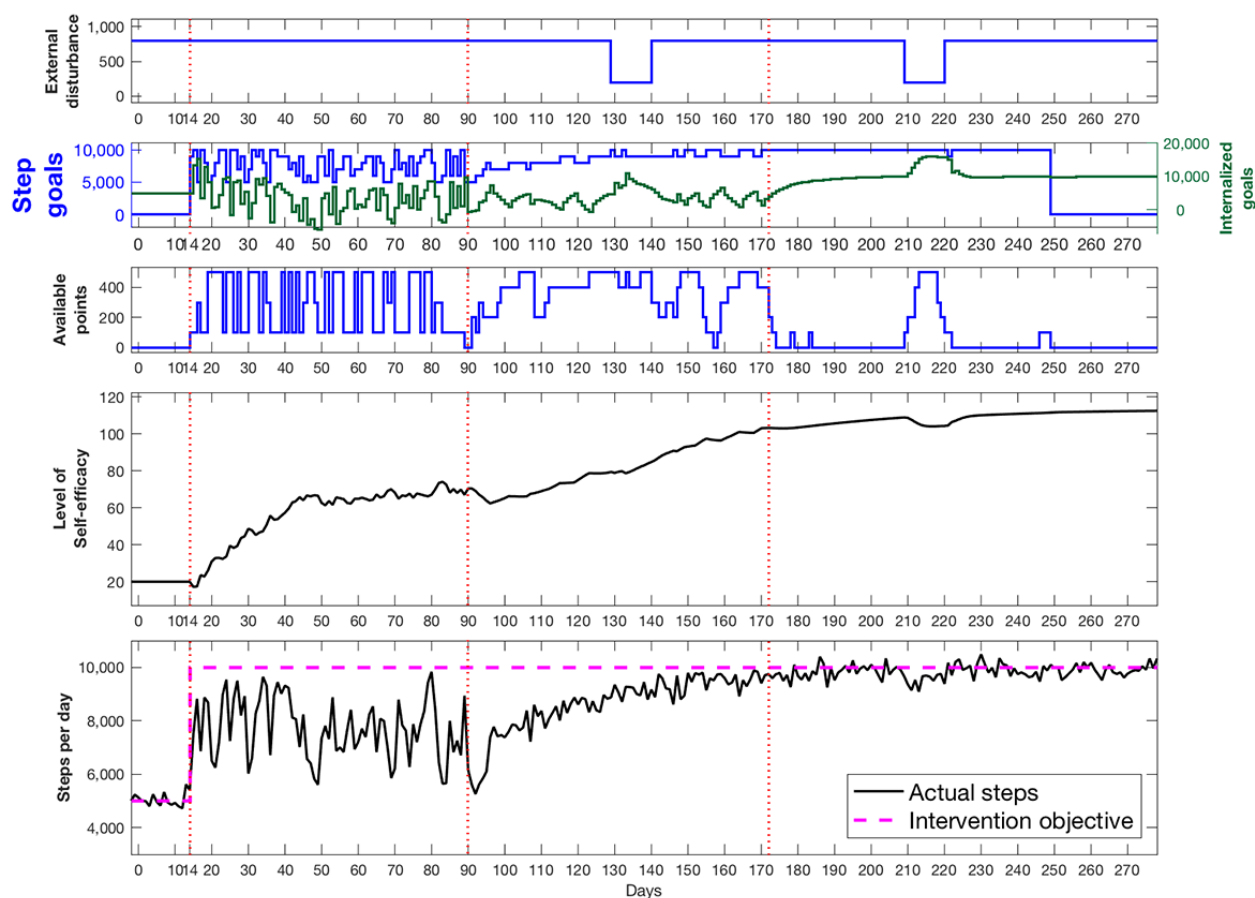


Figure 8. Control optimization trial for Just Walk.

Returning to our *Just Walk* example, we have designed a control optimization trial with four phases specifically designed to test key elements of our adaptive intervention (see Figure 8). The figure is a simulated time series of one participant taking part in our control optimization trial. Row 1 simulates “disturbances” such as getting sick to illustrate how the controller might react (eg, increase points or lower goals). There are four phases divided by the red vertical lines. Phase 1 is an initial measurement only, baseline period, which provides a grounding of the person’s current activity. Phase 2 is an “open-loop” system identification experiment, similar to the study in step 3, whereby step goals (row 2) and points (row 3) are systematically “excited” to enable generation of individualized dynamical models. This phase enables estimating or validating our dynamical model and individualized tailoring variable selection as per our prior study. In phase 3, the model-predictive controller uses those dynamical models to make intervention option decisions to foster initiation of PA towards PA guidelines (row 5) and increased self-efficacy (row 4). During phase 3, the model-predictive controller will strive for appropriate targets for our at-risk group (ie, 10,000 steps/day on average or, if a person does not achieve 10,000 steps/day during initiation, then 3000 steps/day above the person’s baseline median steps). Phase 4 focuses on testing the controller’s decision rules for maintenance (eg, see reduced points provided in row 3). Specifically, we will optimize our approach for providing as minimal support as possible while a person maintains set-point targets.

With this experiment completed, we will be able to systematically test and optimize core elements of our adaptive intervention. In particular, our open loop system identification portion enables data-driven optimization for individualized dynamical models and selecting individualized tailoring variables as described above. Unlike the above work, the final definition of success, which is a person maintaining targeted step levels, will be available and, thus, can be used to define percentage model fits that are, indeed, good enough for individualization purposes. Our closed loop subexperiments allow us to optimize our controller’s ability to achieve set-point targets for each individual for each state, including initiation, maintenance, and possible relapses. We can judge success or failure relative to our optimization criteria (eg, 10,000 steps/day).

Furthermore, we can also produce aggregate (also called nomothetic) information across the sample of participants. Specifically, another optimization check involves comparison of the percentage of our sample that achieves our maintenance targets relative to current best practice PA interventions that appear to produce maintenance targets for approximately 50% of their samples [82,84]. Using previous work as a referent, we can establish the plausibility that our approach is comparable with current best practices if 50% of participants meet our set-point target and exceed current best practices if a higher percentage of our sample achieves our set-point targets. Thus, the control optimization trial can enable both case-by-case (ie, idiographic) optimization for individualization (ie, meeting

minimal model fits) and adaptation (amount of time within the desired set-point range across the intervention) and nomothetic optimization (ie, percent of participants meeting target thresholds). This multi-criteria optimization fits with the multiple elements within an adaptive intervention. Furthermore, the study is highly efficient as these elements can all be systematically studied within a single study and, indeed for most of our criteria, on a case-by-case basis.

Discussion

Summary

Control systems engineering is a rich discipline that has strategies mHealth researchers and practitioners can use for optimizing elements of adaptive interventions. It is particularly well matched to problems that (1) are dynamic, (2) have useful dynamic interventions available, (3) have an outcome measure that can be measured with sufficient temporal density over an extended time period, and (4) have desirable states for the target outcomes that can be defined as the optimization criteria on a case-by-case basis. There are five suggested (though not necessarily all required) steps for optimizing an adaptive intervention via control engineering: (1) derive a preliminary dynamical model, (2) select intervention options, (3) conduct a system identification open loop experiment, (4) design the controller and optimization criteria, and (5) conduct a control systems optimization trial. This approach holds great promise for expanding the potential of adaptive interventions. This is because control engineering provides a wide range of approaches to systematically unpack and test or optimize the various elements of adaptive interventions both on a case-by-case or idiographic and an aggregate or nomothetic level.

Connections to Multiphase Optimization Strategy

These steps map on to the MOST framework [28]. Within the preparation phase of MOST, the four suggested steps include the following: (1) develop a conceptual model; (2) develop intervention components; (3) if necessary, pilot test the intervention components; and (4) define the optimization criteria. These steps map on to steps 1 to 4 of the process we delineate but with slight variations based on the requirements for control engineering. An essential difference is step 3, because system identification is valuable not only in preparation for an adaptive intervention (and thus mimics the purpose of step 3 of MOST) but also for theory testing. Thus, it should not necessarily be thought of as pilot testing for the intervention but instead as a valuable scientific pursuit in and of itself. Within MOST, the optimization phase involves conducting an optimization trial, such as a factorial design. One could view system identification experiments, thus, as a form of an optimization trial. That said, the control optimization trial (step 5 in our analogous process) is directly parallel to other optimization trials, as the goal of the trial is primarily on optimizing the intervention, whereas system identification is more focused on theory testing and, thus, not as clearly similar to the optimization trials. If there is interest in seeing if this controller performs better than current standard of care, then the final step of MOST, evaluation via an RCT, can occur. Specifically, if the controller meets the threshold of the

optimization criteria, the evaluation phase can then proceed whereby the control-driven intervention can be evaluated relative to a meaningful comparator (eg, current standard of care complex intervention [30]). If, however, the goal is to develop modules that are repurposable, self-contained intervention components (ie, components designed to function separately), then another plausible approach would be to modularize this work for other use cases, as delineated in agile science [103,106].

Beyond the steps, there is also synergy between MOST and control engineering principles. A central focus of MOST is efficiency, including the use of efficient experimental designs and grounding research in real-world constraints related to implementation with the long-term goal of facilitating more efficient and robust knowledge accumulation across studies. Continuous optimization is the second common principle that emphasizes the logic of a continual, iterative process related to further improving and refining behavioral interventions. Control engineering shares these principles of efficiency and continuous optimization. Overall, our work fits well with MOST and current trends in mHealth and the science of behavior change [133].

Added Considerations Within Control Systems Engineering

As highlighted already, control engineering practices include the principle of triangulation [134]. Unlike the concept of a definitive trial [134], the logic of triangulation (sometimes also called consilience [135]), involves the use of multiple methods and approaches to synergistically study a problem. The basic logic is that every method comes with inherent strengths and weaknesses. When different methods with different strengths and weaknesses point in a common direction, confidence in the assertion increases. Just like how neuropsychologists look for patterns across neurocognitive tests instead of relying on one test, control engineers use a wide range of methods that each have strengths and limitations for iteratively optimizing dynamical models and controllers. This is illustrated in our detailed discussion about a control optimization trial and the many ways in which it can be defined and operationalized via mixed use of open loop system identification experiments, closed loop tests, and robustness testing. If multiple methods and criteria point in a similar direction, then, one can have increased confidence that the overall system is working. Furthermore, if the different tests are not providing consistent results, then the discrepancies can often be used to better understand which elements of the complex intervention are likely inadequate, thus supporting optimization.

These methods are designed to understand and support better prediction and decision making for a given individual. Although we have highlighted the strengths of this approach, there are inherent weaknesses. For example, one potential trade-off exists related to the optimization criteria. If the optimization criteria that are chosen are not meaningful, then even if the controller achieves success (ie, optimization criteria are met), then nothing clinically meaningful has been achieved. This can be mitigated, of course, with optimization criteria that are grounded in clear evidence showing that they are clinically meaningful as is the case within *Just Walk*. Note that this problem is not unique to

control systems engineering. It exists within other methods including RCTs, but with RCTs, it manifests via the control condition chosen [136]. In brief, when a poor control condition is chosen, a statistically significant difference may be found (ie, success for this method), but that does not necessarily equate to a meaningful result. To put it more colloquially, one could compare a bad intervention and use a worse intervention as a control, run a trial, and find that bad is better than worse. Unfortunately, the end result is still a bad intervention. Regardless of methods, it is essential to have clarity on what success means, in terms of real-world utility, as is argued in agile science [103,106].

A second major trade-off of control engineering and, indeed, any idiographic approach, is the undervaluing of generalizability to other individuals and contexts. This establishes the need for other methods that are better at balancing this idiographic emphasis with more of a nomothetic emphasis, such as RCTs. With that said, generalization to other individuals and contexts can feasibly occur via a different pathway toward generalizability knowledge, namely, the generalizability concept of causal explanatory models [137]. Shadish et al [137], in their formulation on a theory of generalization, highlighted the concept of causal explanatory models, which are mechanistic models that not only define if there is causal effect (what they called a causal descriptive model and what is produced by an RCT) but *how* the effect occurs, mechanistically. Arguably, dynamical modeling, particularly when robust semiphenomenological models can be validated, move in the direction of causal explanatory models and, thus, can feasibly aid in improving mechanistic understanding of a phenomenon and, thus, produce generalizable knowledge.

Returning to the concept of triangulation, an RCT can balance out the weaknesses of control engineering methods. As illustrated in the introduction and optimization section, RCTs compromise on providing insights about how, when, where, and for whom a given intervention element works, in the pursuit of stronger internal validity at the intervention package level and also increased external validity in terms of statistical claims of generalizability to the population the sample is conceptually drawn from. As the control optimization trial is an inherently n-of-1 study design, it enables the possibility of it being embedded within an RCT as the intervention arm. This possibility enables a highly efficient way of conducting multiple tests within a single trial that is squarely grounded in the philosophical logic of triangulation, as one trial can test intervention elements and also compare the package to another package. Indeed, including a control condition as a comparison with a control systems optimization trial is, arguably, a highly efficiently rigorous approach to test an adaptive intervention [103].

We emphasize triangulation as we see this as well matched to the complexity of adaptive interventions and possibly behavioral interventions more generally, even outside of the domain of control systems engineering. It is the cornerstone of our key thesis that adaptive interventions are more likely to succeed if its elements can be iteratively improved via optimization. In particular, the fact that there are so many elements within an adaptive intervention (eg, intervention components, decision

points, tailoring variables, and decision rules) establishes the need for triangulation. This fits with discussions in psychology, such as the need for a pluralistic approach to causality [134].

Implications and Future Work

As articulated elsewhere, advancements in digital technologies are rapidly converging to enable a new era in the understanding of human behavior [13,18,138]. A central argument made elsewhere is that the time is right for health and behavioral sciences to reexamine their experimental and analytic strategies [13,18,138]. Although there is great opportunity for a variety of other methods, health and behavioral scientists should more carefully consider control systems engineering. Not only is the time right, from a technical standpoint, but very classic work in psychological science engaged within control theory; thus suggesting that this is really a return to classical roots in psychological science [139-142]. Conceptually, there are many reasons to believe that control systems engineering could be a foundational class of methods behavioral and health scientists could use to improve impact, particularly related to individualized mHealth interventions. Of course, this requires far more research and empirical work before any firm conclusions can be drawn on the potential.

In terms of limitations and future work, more work is needed to clearly evaluate the utility of this approach relative to other methods. For example, the current method used in MOST for optimizing a static intervention is a factorial trial, and SMART and MRT are proposed for adaptive interventions. One valuable test to be conducted is comparison of an optimized intervention to an intervention that was not optimized using these methods. This comparison can be made using an RCT. As the control optimization trial is an inherently n-of-1 method, it is possible to compare the control optimization trial, as a proxy of an optimized intervention, with a control condition that lacks control engineering features. A trial such as this would provide insights on the plausible added value optimization via control engineering may produce relative to more traditional approaches for intervention development whereby the elements are not optimized but, instead, the elements of the intervention are defined based on prior aggregate evidence, user-centered research, and theory.

Building on this point, future work should focus on providing greater clarity on when to use which method for optimizing static and adaptive interventions. As one possible formulation on this, SMART appears useful when the goal is the selection of a progression of decisions to make with relatively infrequent adaptation (eg, once every few months) and with well-specified if-then decision rules. As such, SMART might be particularly valuable within clinical practice. MRT appears particularly valuable for just-in-time adaptive interventions. We argue that control systems engineering methods are likely particularly valuable when the goal is to facilitate a more long-term trajectory of change, such as gradually increasing a target behavior whereby achievement of a desired state cannot happen immediately (eg from 6000 steps/day to 10,000 steps/day or a 5% reduction in weight) but, instead, requires slow progression and building up of skills. Similarly, control engineering methods can also be valuable for facilitating maintenance of a targeted

set-point by facilitating small adjustments and provision of support in sort of stepped-care framework. Although we think these general principles are correct conceptually, future empirical work is needed to explore the strengths and limitations of these approaches and the assertions made on when to use which method.

More work related to establishing meaning optimization criteria is needed. This work hinges on well-specified definitions of success that are clinically and practically meaningful but that is not necessarily always available for all elements of an adaptive intervention. For example, we established our model fit estimates as good enough for individualization based on Cohen's work [127]. We fully recognize that this is an extension and thus may not be appropriate. Future work is needed to think clearly through what good enough optimization is for elements and the adaptive interventions overall.

Finally, future work should further explore if and, if so, how to integrate the logic of triangulation more actively within the

development of mHealth interventions. As highlighted before, there is already research starting in this domain but, future research that provides scaffolding for health and behavioral scientists to work through this more complex approach to the design, optimization, and evaluation of interventions could be valuable. We have started this process through the articulation of agile science [103], but further work is needed. Finally, if control engineering does prove valuable, there will be a need for more interdisciplinary training between control engineers and psychologists.

Conclusions

In sum, mHealth is well poised to take advantage of control engineering methods for the optimization of adaptive interventions. The time is now for health and behavioral scientists to more closely examine control engineering methods. If the approach proves valuable for health problems, new partnerships should be forged between health and behavioral sciences and control systems engineers in the design, optimization, and evaluation of adaptive interventions.

Acknowledgments

This work was supported by a grant from the National Science Foundation (IIS: 1449751, PI: Hekler).

Authors' Contributions

EBH was the principal investigator for the collective research efforts and primarily drafted the manuscript. All other authors reviewed and provided support to sections of the paper and also engaged in the design, implementation, and evaluation of the *Just Walk* case study.

Conflicts of Interest

None declared.

References

1. Sheeran P, Klein WM, Rothman AJ. Health Behavior Change: Moving from Observation to Intervention. *Annu Rev Psychol* 2017 Jan 03;68:573-600. [doi: [10.1146/annurev-psych-010416-044007](https://doi.org/10.1146/annurev-psych-010416-044007)] [Medline: [27618942](https://pubmed.ncbi.nlm.nih.gov/27618942/)]
2. Rothman AJ, Gollwitzer PM, Grant AM, Neal DT, Sheeran P, Wood W. Hale and hearty policies: how psychological science can create and maintain healthy habits. *Perspect Psychol Sci* 2015 Nov;10(6):701-705. [doi: [10.1177/1745691615598515](https://doi.org/10.1177/1745691615598515)] [Medline: [26581721](https://pubmed.ncbi.nlm.nih.gov/26581721/)]
3. Mann T, Tomiyama AJ, Ward A. Promoting public health in the context of the "Obesity Epidemic": false starts and promising new directions. *Perspect Psychol Sci* 2015 Nov;10(6):706-710 [FREE Full text] [doi: [10.1177/1745691615586401](https://doi.org/10.1177/1745691615586401)] [Medline: [26581722](https://pubmed.ncbi.nlm.nih.gov/26581722/)]
4. van der Linden S, Maibach E, Leiserowitz A. Improving public engagement with climate change: five "best practice" insights from psychological science. *Perspect Psychol Sci* 2015 Nov;10(6):758-763. [doi: [10.1177/1745691615598516](https://doi.org/10.1177/1745691615598516)] [Medline: [26581732](https://pubmed.ncbi.nlm.nih.gov/26581732/)]
5. Chase JA. Interventions to increase physical activity among older adults: a meta-analysis. *Gerontologist* 2015 Aug;55(4):706-718 [FREE Full text] [doi: [10.1093/geront/gnu090](https://doi.org/10.1093/geront/gnu090)] [Medline: [25298530](https://pubmed.ncbi.nlm.nih.gov/25298530/)]
6. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. *Am J Public Health* 2011 Apr;101(4):751-758 [FREE Full text] [doi: [10.2105/AJPH.2010.194381](https://doi.org/10.2105/AJPH.2010.194381)] [Medline: [21330590](https://pubmed.ncbi.nlm.nih.gov/21330590/)]
7. Conn VS, Hafdahl AR, Moore SM, Nielsen PJ, Brown LM. Meta-analysis of interventions to increase physical activity among cardiac subjects. *Int J Cardiol* 2009 Apr 17;133(3):307-320 [FREE Full text] [doi: [10.1016/j.ijcard.2008.03.052](https://doi.org/10.1016/j.ijcard.2008.03.052)] [Medline: [18582959](https://pubmed.ncbi.nlm.nih.gov/18582959/)]
8. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. *JMIR Mhealth Uhealth* 2015 Feb 26;3(1):e20 [FREE Full text] [doi: [10.2196/mhealth.3335](https://doi.org/10.2196/mhealth.3335)] [Medline: [25803705](https://pubmed.ncbi.nlm.nih.gov/25803705/)]

9. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clin Trials* 2014 Aug;11(4):426-434 [[FREE Full text](#)] [doi: [10.1177/1740774514536795](https://doi.org/10.1177/1740774514536795)] [Medline: [24902922](#)]
10. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Transl Behav Med* 2014 Sep;4(3):260-274 [[FREE Full text](#)] [doi: [10.1007/s13142-014-0265-0](https://doi.org/10.1007/s13142-014-0265-0)] [Medline: [25264466](#)]
11. 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;34S:1209-1219 [[FREE Full text](#)] [doi: [10.1037/hea0000306](https://doi.org/10.1037/hea0000306)] [Medline: [26651462](#)]
12. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med* 2016 Sep 23:- [[FREE Full text](#)] [doi: [10.1007/s12160-016-9830-8](https://doi.org/10.1007/s12160-016-9830-8)] [Medline: [27663578](#)]
13. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. *Am J Prev Med* 2016 Nov;51(5):816-824. [doi: [10.1016/j.amepre.2016.05.001](https://doi.org/10.1016/j.amepre.2016.05.001)] [Medline: [27745681](#)]
14. 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](https://doi.org/10.1186/2001-1326-2-10)] [Medline: [23663660](#)]
15. Riley WT, Martin CA, Rivera DE, Hekler EB, Adams MA, Buman MP, et al. Development of a dynamic computational model of social cognitive theory. *Transl Behav Med* 2016 Dec;6(4):483-495 [[FREE Full text](#)] [doi: [10.1007/s13142-015-0356-6](https://doi.org/10.1007/s13142-015-0356-6)] [Medline: [27848208](#)]
16. Riley WT, Nilsen WJ, Manolio TA, Masys DR, Lauer M. News from the NIH: potential contributions of the behavioral and social sciences to the precision medicine initiative. *Transl Behav Med* 2015 Sep;5(3):243-246 [[FREE Full text](#)] [doi: [10.1007/s13142-015-0320-5](https://doi.org/10.1007/s13142-015-0320-5)] [Medline: [26327928](#)]
17. Riley WT, Rivera DE. Methodologies for optimizing behavioral interventions: introduction to special section. *Transl Behav Med* 2014 Sep;4(3):234-237 [[FREE Full text](#)] [doi: [10.1007/s13142-014-0281-0](https://doi.org/10.1007/s13142-014-0281-0)] [Medline: [25264463](#)]
18. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? *Transl Behav Med* 2011 Mar;1(1):53-71 [[FREE Full text](#)] [doi: [10.1007/s13142-011-0021-7](https://doi.org/10.1007/s13142-011-0021-7)] [Medline: [21796270](#)]
19. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Int J Nurs Stud* 2013 May;50(5):587-592. [doi: [10.1016/j.ijnurstu.2012.09.010](https://doi.org/10.1016/j.ijnurstu.2012.09.010)] [Medline: [23159157](#)]
20. Freigoun MT, Martín CA, Magann AB, Rivera DE, Phatak SS, Korinek E, et al. System identification of Just Walk: a behavioral mHealth intervention for promoting physical activity. In: IEEE. 2017 Presented at: American Control Conference; July 3, 2017; Seattle, WA p. 116-121.
21. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *J Behav Med* 2018 Feb;41(1):74-86. [doi: [10.1007/s10865-017-9878-3](https://doi.org/10.1007/s10865-017-9878-3)] [Medline: [28918547](#)]
22. Phatak SS, Freigoun MT, Martín CA, Rivera DE, Korinek EV, Adams MA, et al. Modeling individual differences: a case study of the application of system identification for personalizing a physical activity intervention. *J Biomed Inform* 2018;79:82-97. [doi: [10.1016/j.jbi.2018.01.010](https://doi.org/10.1016/j.jbi.2018.01.010)] [Medline: [29409750](#)]
23. Åström KJ. Automatic Control - The Hidden Technology. In: *Advances in Control*. London: Springer; 1999:1-28.
24. Ashour M, Bekiroglu K, Yang CH, Lagoa C, Conroy D, Smyth J. On the mathematical modeling of the effect of treatment on human physical activity. In: IEEE. 2016 Presented at: IEEE Conference on Control Applications (CCA); September 19-22, 2016; Buenos Aires. [doi: [10.1109/CCA.2016.7587951](https://doi.org/10.1109/CCA.2016.7587951)]
25. Karimi M, Rao RR. Human Body Composition Estimation and Model-Free Control Design for Weight Management. In: ASME. 2014 Presented at: ASME Proceedings | Biomedical and Bioengineering Applications; October 22-24 2014; San Antonio, TX URL: <http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?articleid=2086071> [doi: [10.1115/DSCC2014-6354](https://doi.org/10.1115/DSCC2014-6354)]
26. Karimi M, Rao RR. On the estimation and control of human body composition. *Trans Inst Meas Control* 2017 May 30:-. [doi: [10.1177/0142331217705946](https://doi.org/10.1177/0142331217705946)]
27. Pirolli P, Mohan S, Venkatakrishnan A, Nelson L, Silva M, Springer A. Implementation intention and reminder effects on behavior change in a mobile health system: a predictive cognitive model. *J Med Internet Res* 2017 Nov 30;19(11):e397. [doi: [10.2196/jmir.8217](https://doi.org/10.2196/jmir.8217)] [Medline: [29191800](#)]
28. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). New York: Springer; 2018.
29. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med* 2007 May;32(5 Suppl):S112-S118 [[FREE Full text](#)] [doi: [10.1016/j.amepre.2007.01.022](https://doi.org/10.1016/j.amepre.2007.01.022)] [Medline: [17466815](#)]

30. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. *Am J Prev Med* 2016 Nov;51(5):843-851 [[FREE Full text](#)] [doi: [10.1016/j.amepre.2016.06.008](https://doi.org/10.1016/j.amepre.2016.06.008)] [Medline: [27745684](#)]
31. Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment. *Transl Behav Med* 2014 Sep;4(3):238-251 [[FREE Full text](#)] [doi: [10.1007/s13142-013-0239-7](https://doi.org/10.1007/s13142-013-0239-7)] [Medline: [25264464](#)]
32. Chakraborty B, Collins LM, Strecher VJ, Murphy SA. Developing multicomponent interventions using fractional factorial designs. *Stat Med* 2009 Sep 20;28(21):2687-2708 [[FREE Full text](#)] [doi: [10.1002/sim.3643](https://doi.org/10.1002/sim.3643)] [Medline: [19575485](#)]
33. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. *Am J Prev Med* 2014 Oct;47(4):498-504 [[FREE Full text](#)] [doi: [10.1016/j.amepre.2014.06.021](https://doi.org/10.1016/j.amepre.2014.06.021)] [Medline: [25092122](#)]
34. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. *Annu Rev Clin Psychol* 2012;8:21-48 [[FREE Full text](#)] [doi: [10.1146/annurev-clinpsy-032511-143152](https://doi.org/10.1146/annurev-clinpsy-032511-143152)] [Medline: [22224838](#)]
35. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. *Health Psychol* 2015 Dec;34S:1220-1228 [[FREE Full text](#)] [doi: [10.1037/hea0000305](https://doi.org/10.1037/hea0000305)] [Medline: [26651463](#)]
36. Sutton RS, Barto A. Reinforcement Learning: An Introduction. Boston: MIT press; 1998.
37. Consolvo S, Bentley FR, Hekler EB, Phatak SS. Mobile User Research: A Practical Guide. San Rafael, CA: Morgan & Claypool Publishers; May 2017:-.
38. Almirall D, Nahum-Shani I, Wang L, Kasari C. Experimental designs for research on adaptive interventions: Singly and sequentially randomized trials. In: Optimization of behavioral, biobehavioral, and biomedical interventions: Advanced topics. New York: Springer; 2018.
39. Li Y, Ang KH, Chong GC. PID control system analysis and design. *IEEE Control Syst* 2006 Feb;26(1):32-41. [doi: [10.1109/MCS.2006.1580152](https://doi.org/10.1109/MCS.2006.1580152)]
40. Ang KH, Chong G, Li Y. PID control system analysis, design, and technology. *IEEE Trans Control Syst Technol* 2005 Jul;13(4):559-576. [doi: [10.1109/TCST.2005.847331](https://doi.org/10.1109/TCST.2005.847331)]
41. Qin SJ, Badgwell TA. A survey of industrial model predictive control technology. *Control Eng Pract* 2003 Jul;11(7):733-764. [doi: [10.1016/S0967-0661\(02\)00186-7](https://doi.org/10.1016/S0967-0661(02)00186-7)]
42. Eysenbach G. The law of attrition. *J Med Internet Res* 2005 Mar;7(1):e11 [[FREE Full text](#)] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](#)]
43. Morari M, Zafiriou E. Robust Process Control. Englewood Cliffs, NJ: Prentice Hall; 1989.
44. Floegel TA, Florez-Pregonero A, Hekler EB, Buman MP. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. *J Gerontol A Biol Sci Med Sci* 2017 Feb;72(2):229-236. [doi: [10.1093/gerona/glw098](https://doi.org/10.1093/gerona/glw098)] [Medline: [27257217](#)]
45. Hekler EB, Buman MP, Grieco L, Rosenberger M, Winter SJ, Haskell WL, et al. Validation of physical activity tracking via Android smartphones compared to ActiGraph accelerometer: laboratory-based and free-living validation studies. *JMIR Mhealth Uhealth* 2015 Apr 15;3(2):e36 [[FREE Full text](#)] [doi: [10.2196/mhealth.3505](https://doi.org/10.2196/mhealth.3505)] [Medline: [25881662](#)]
46. Ljung L. System Identification: Theory for the User (2nd Edition). Upper Saddle River, NJ: Prentice Hall; 1999.
47. Ljung L. System Identification Toolbox for Use with MATLAB: User's Guide. Natick, MA: The MathWorks Inc; 1995.
48. Ljung L. Experiments with Identification of Continuous Time Models. *IFAC Proc Volumes* 2009;42(10):1175-1180. [doi: [10.3182/20090706-3-FR-2004.00195](https://doi.org/10.3182/20090706-3-FR-2004.00195)]
49. Ljung L, Glad T. Modeling of Dynamic Systems. In: Prentice Hall. New Jersey: Prentice-Hall; 1994.
50. Martin CA, Rivera DE, Hekler EB. A control engineering approach for optimizing physical activity behavioral interventions. In: IEEE. 2016 Presented at: Ecuador Technical Chapters Meeting (ETCM); October 12-14, 2016; Guayaquil, Ecuador p. 12-14. [doi: [10.1109/ETCM.2016.7750851](https://doi.org/10.1109/ETCM.2016.7750851)]
51. Rivera DE, Martín CA, Timms KP, Deshpande S, Nandola N, Hekler E. Control Systems Engineering for Optimizing Behavioral mHealth Interventions. In: Mobile Health: Sensors, Analytic Methods, and Applications. Basel: Springer International Publishing; 2017:455-493.
52. Clague J, Bernstein L. Physical activity and cancer. *Curr Oncol Rep* 2012 Dec;14(6):550-558 [[FREE Full text](#)] [doi: [10.1007/s11912-012-0265-5](https://doi.org/10.1007/s11912-012-0265-5)] [Medline: [22945451](#)]
53. Bianchini F, Kaaks R, Vainio H. Weight control and physical activity in cancer prevention. *Obes Rev* 2002 Feb;3(1):5-8. [doi: [10.1046/j.1467-789X.2002.00046.x](https://doi.org/10.1046/j.1467-789X.2002.00046.x)]
54. Murtagh EM, Nichols L, Mohammed MA, Holder R, Nevill AM, Murphy MH. The effect of walking on risk factors for cardiovascular disease: an updated systematic review and meta-analysis of randomised control trials. *Prev Med* 2015 Mar;72:34-43. [doi: [10.1016/j.ypmed.2014.12.041](https://doi.org/10.1016/j.ypmed.2014.12.041)] [Medline: [25579505](#)]
55. Karstoft K, Winding K, Knudsen SH, Nielsen JS, Thomsen C, Pedersen BK, et al. The effects of free-living interval-walking training on glycemic control, body composition, and physical fitness in type 2 diabetic patients: a randomized, controlled trial. *Diabetes Care* 2013 Feb;36(2):228-236 [[FREE Full text](#)] [doi: [10.2337/dc12-0658](https://doi.org/10.2337/dc12-0658)] [Medline: [23002086](#)]
56. Tudor-Locke C, Hart TL, Washington TL. Expected values for pedometer-determined physical activity in older populations. *Int J Behav Nutr Phys Act* 2009 Aug 25;6:59 [[FREE Full text](#)] [doi: [10.1186/1479-5868-6-59](https://doi.org/10.1186/1479-5868-6-59)] [Medline: [19706192](#)]

57. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, et al. How many steps/day are enough? For older adults and special populations. *Int J Behav Nutr Phys Act* 2011 Jul 28;8:80 [FREE Full text] [doi: [10.1186/1479-5868-8-80](https://doi.org/10.1186/1479-5868-8-80)] [Medline: [21798044](https://pubmed.ncbi.nlm.nih.gov/21798044/)]
58. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *Lancet* 2014 Jul;384(9937):45-52. [doi: [10.1016/s0140-6736\(14\)60648-6](https://doi.org/10.1016/s0140-6736(14)60648-6)]
59. Tudor-Locke C, Craig C, Thyfault J, Spence J. A step-defined sedentary lifestyle index: <5000 steps/day. *Appl Physiol Nutr Metab* 2013 Feb;38(2):100-114 [FREE Full text] [doi: [10.1139/apnm-2012-0235](https://doi.org/10.1139/apnm-2012-0235)] [Medline: [23438219](https://pubmed.ncbi.nlm.nih.gov/23438219/)]
60. Cdc. BRFSS Prevalence & Trends Data URL: <https://www.cdc.gov/brfss/brfssprevalence/index.html> [accessed 2018-05-14] [WebCite Cache ID 6zPqUjydm]
61. Owen N, Bauman A, Brown W. Too much sitting: a novel and important predictor of chronic disease risk? *Br J Sports Med* 2009 Feb;43(2):81-83. [doi: [10.1136/bjsm.2008.055269](https://doi.org/10.1136/bjsm.2008.055269)] [Medline: [19050003](https://pubmed.ncbi.nlm.nih.gov/19050003/)]
62. King AC, Sallis JF. Why and how to improve physical activity promotion: lessons from behavioral science and related fields. *Prev Med* 2009 Oct;49(4):286-288 [FREE Full text] [doi: [10.1016/j.ypmed.2009.07.007](https://doi.org/10.1016/j.ypmed.2009.07.007)] [Medline: [19631233](https://pubmed.ncbi.nlm.nih.gov/19631233/)]
63. Haskell WL, Blair SN, Hill JO. Physical activity: health outcomes and importance for public health policy. *Prev Med* 2009 Oct;49(4):280-282. [doi: [10.1016/j.ypmed.2009.05.002](https://doi.org/10.1016/j.ypmed.2009.05.002)] [Medline: [19463850](https://pubmed.ncbi.nlm.nih.gov/19463850/)]
64. Bauman A, Bull F, Chey T, Craig CL, Ainsworth BE, Sallis JF, IPS Group. The international prevalence study on physical activity: results from 20 countries. *Int J Behav Nutr Phys Act* 2009 Mar 31;6:21 [FREE Full text] [doi: [10.1186/1479-5868-6-21](https://doi.org/10.1186/1479-5868-6-21)] [Medline: [19335883](https://pubmed.ncbi.nlm.nih.gov/19335883/)]
65. Woolf K, Reese CE, Mason MP, Beaird LC, Tudor-Locke C, Vaughan LA. Physical activity is associated with risk factors for chronic disease across adult women's life cycle. *J Am Diet Assoc* 2008 Jun;108(6):948-959. [doi: [10.1016/j.jada.2008.03.015](https://doi.org/10.1016/j.jada.2008.03.015)] [Medline: [18502225](https://pubmed.ncbi.nlm.nih.gov/18502225/)]
66. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008 Jan;40(1):181-188. [doi: [10.1249/mss.0b013e31815a51b3](https://doi.org/10.1249/mss.0b013e31815a51b3)] [Medline: [18091006](https://pubmed.ncbi.nlm.nih.gov/18091006/)]
67. Sugiyama T, Healy GN, Dunstan DW, Salmon J, Owen N. Joint associations of multiple leisure-time sedentary behaviours and physical activity with obesity in Australian adults. *Int J Behav Nutr Phys Act* 2008 Jul 1;5:35 [FREE Full text] [doi: [10.1186/1479-5868-5-35](https://doi.org/10.1186/1479-5868-5-35)] [Medline: [18590570](https://pubmed.ncbi.nlm.nih.gov/18590570/)]
68. Matthews CE. Physical activity in the United States measured by accelerometer: comment. *Med Sci Sports Exerc* 2008 Jun;40(6):1188; author reply 1189. [doi: [10.1249/MSS.0b013e31817057da](https://doi.org/10.1249/MSS.0b013e31817057da)] [Medline: [18496095](https://pubmed.ncbi.nlm.nih.gov/18496095/)]
69. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. *CMAJ* 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: [10.1503/cmaj.051351](https://doi.org/10.1503/cmaj.051351)] [Medline: [16534088](https://pubmed.ncbi.nlm.nih.gov/16534088/)]
70. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee report, 2008. Part A: Executive Summary. *Nutr Rev* 2009 Feb;67(2):114-120. [doi: [10.1111/j.1753-4887.2008.00136.x](https://doi.org/10.1111/j.1753-4887.2008.00136.x)] [Medline: [19178654](https://pubmed.ncbi.nlm.nih.gov/19178654/)]
71. Rauner A, Jekauc D, Mess F, Schmidt S, Woll A. Tracking physical activity in different settings from late childhood to early adulthood in Germany: the MoMo longitudinal study. *BMC Public Health* 2015 Apr 17;15:391 [FREE Full text] [doi: [10.1186/s12889-015-1731-4](https://doi.org/10.1186/s12889-015-1731-4)] [Medline: [25887314](https://pubmed.ncbi.nlm.nih.gov/25887314/)]
72. Hirvensalo M, Lintunen T. Life-course perspective for physical activity and sports participation. *Eur Rev Aging Phys Act* 2011 Jan 11;8(1):13-22. [doi: [10.1007/s11556-010-0076-3](https://doi.org/10.1007/s11556-010-0076-3)]
73. Kern ML, Reynolds CA, Friedman HS. Predictors of physical activity patterns across adulthood: a growth curve analysis. *Pers Soc Psychol Bull* 2010 Aug;36(8):1058-1072. [doi: [10.1177/0146167210374834](https://doi.org/10.1177/0146167210374834)] [Medline: [20573949](https://pubmed.ncbi.nlm.nih.gov/20573949/)]
74. Levin S, Jacobs DR, Ainsworth BE, Richardson MT, Leon AS. Intra-individual variation and estimates of usual physical activity. *Ann Epidemiol* 1999 Nov;9(8):481-488. [Medline: [10549881](https://pubmed.ncbi.nlm.nih.gov/10549881/)]
75. Bélanger M, Gray-Donald K, O'Loughlin J, Paradis G, Hanley J. Influence of weather conditions and season on physical activity in adolescents. *Ann Epidemiol* 2009 Mar;19(3):180-186. [doi: [10.1016/j.annepidem.2008.12.008](https://doi.org/10.1016/j.annepidem.2008.12.008)] [Medline: [19217000](https://pubmed.ncbi.nlm.nih.gov/19217000/)]
76. Brandon CA, Gill DP, Speechley M, Gilliland J, Jones GR. Physical activity levels of older community-dwelling adults are influenced by summer weather variables. *Appl Physiol Nutr Metab* 2009 Apr;34(2):182-190. [doi: [10.1139/H09-004](https://doi.org/10.1139/H09-004)] [Medline: [19370048](https://pubmed.ncbi.nlm.nih.gov/19370048/)]
77. Centers for Disease ControlPrevention (CDC). Monthly estimates of leisure-time physical inactivity--United States, 1994. *MMWR Morb Mortal Wkly Rep* 1997 May 9;46(18):393-397 [FREE Full text] [Medline: [9157273](https://pubmed.ncbi.nlm.nih.gov/9157273/)]
78. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR. Sources of variance in daily physical activity levels as measured by an accelerometer. *Med Sci Sports Exerc* 2002 Aug;34(8):1376-1381. [Medline: [12165695](https://pubmed.ncbi.nlm.nih.gov/12165695/)]
79. Pivarnik JM, Reeves MJ, Rafferty AP. Seasonal variation in adult leisure-time physical activity. *Med Sci Sports Exerc* 2003 Jun;35(6):1004-1008. [doi: [10.1249/01.MSS.0000069747.55950.B1](https://doi.org/10.1249/01.MSS.0000069747.55950.B1)] [Medline: [12783049](https://pubmed.ncbi.nlm.nih.gov/12783049/)]
80. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. *Public Health* 2007 Dec;121(12):909-922. [doi: [10.1016/j.puhe.2007.04.009](https://doi.org/10.1016/j.puhe.2007.04.009)] [Medline: [17920646](https://pubmed.ncbi.nlm.nih.gov/17920646/)]
81. Small M, Bailey-Davis L, Morgan N, Maggs J. Changes in eating and physical activity behaviors across seven semesters of college: living on or off campus matters. *Health Educ Behav* 2013;40(4):435-441. [Medline: [23232092](https://pubmed.ncbi.nlm.nih.gov/23232092/)]

82. King AC, Hekler EB, Castro CM, Buman MP, Marcus BH, Friedman RH, et al. Exercise advice by humans versus computers: maintenance effects at 18 months. *Health Psychol* 2014 Feb;33(2):192-196. [doi: [10.1037/a0030646](https://doi.org/10.1037/a0030646)] [Medline: [23421896](https://pubmed.ncbi.nlm.nih.gov/23421896/)]
83. Hekler EB, Buman MP, Pothakandiyil N, Rivera DE, Dzierzewski JM, Morgan AA, et al. Exploring behavioral markers of long-term physical activity maintenance: a case study of system identification modeling within a behavioral intervention. *Health Educ Behav* 2013 Oct;40(1 Suppl):51S-62S [FREE Full text] [doi: [10.1177/1090198113496787](https://doi.org/10.1177/1090198113496787)] [Medline: [24084400](https://pubmed.ncbi.nlm.nih.gov/24084400/)]
84. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behavior change following physical activity and dietary interventions. *Health Psychol* 2011 Jan;30(1):99-109. [doi: [10.1037/a0021974](https://doi.org/10.1037/a0021974)] [Medline: [21299298](https://pubmed.ncbi.nlm.nih.gov/21299298/)]
85. Eakin E, Reeves M, Winkler E, Lawler S, Owen N. Maintenance of physical activity and dietary change following a telephone-delivered intervention. *Health Psychol* 2010 Nov;29(6):566-573. [doi: [10.1037/a0021359](https://doi.org/10.1037/a0021359)] [Medline: [20954778](https://pubmed.ncbi.nlm.nih.gov/20954778/)]
86. D'Alonzo KT, Cortese LB. An investigation of habitual and incidental physical activity among Costa Rican and Costa Rican American teenage girls. *J Transcult Nurs* 2007 Jul;18(3):201-207. [doi: [10.1177/1043659607301296](https://doi.org/10.1177/1043659607301296)] [Medline: [17607056](https://pubmed.ncbi.nlm.nih.gov/17607056/)]
87. Marshall AL, Bauman AE, Patch C, Wilson J, Chen J. Can motivational signs prompt increases in incidental physical activity in an Australian health-care facility? *Health Educ Res* 2002 Dec;17(6):743-749. [Medline: [12507349](https://pubmed.ncbi.nlm.nih.gov/12507349/)]
88. Hekler EB, Castro CM, Buman MP, King AC. The CHOICE study: a "taste-test" of utilitarian vs. leisure walking among older adults. *Health Psychol* 2012 Jan;31(1):126-129 [FREE Full text] [doi: [10.1037/a0025567](https://doi.org/10.1037/a0025567)] [Medline: [21928901](https://pubmed.ncbi.nlm.nih.gov/21928901/)]
89. Freigoun MT, Martin CA, Magann AB, Rivera DE, Phatak SS, Korinek EV. System identification of Just Walk: A behavioral mHealth intervention for promoting physical activity. In: IEEE. 2017 Presented at: American Control Conference (ACC); May 24-26, 2017; Seattle, WA. [doi: [10.23919/ACC.2017.7962940](https://doi.org/10.23919/ACC.2017.7962940)]
90. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins J, Marteau TM. Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis. *Prev Med* 2015 Jun;75:75-85 [FREE Full text] [doi: [10.1016/j.ypmed.2015.03.001](https://doi.org/10.1016/j.ypmed.2015.03.001)] [Medline: [25843244](https://pubmed.ncbi.nlm.nih.gov/25843244/)]
91. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. *Am J Prev Med* 2013 Nov;45(5):658-667. [doi: [10.1016/j.amepre.2013.06.017](https://doi.org/10.1016/j.amepre.2013.06.017)] [Medline: [24139781](https://pubmed.ncbi.nlm.nih.gov/24139781/)]
92. Skinner BF. *Science and Human Behavior*. New York: Simon and Schuster; 1953.
93. Ferster CB. *Century psychology series*. In: *Schedules of reinforcement with Skinner*. New York: Appleton-Century-Crofts; 1970:37-46.
94. Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall; 1986.
95. Adams MA, Hurley JC, Todd M, Bhuiyan N, Jarrett CL, Tucker WJ, et al. Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. *BMC Public Health* 2017 Dec 29;17(1):286 [FREE Full text] [doi: [10.1186/s12889-017-4197-8](https://doi.org/10.1186/s12889-017-4197-8)] [Medline: [28356097](https://pubmed.ncbi.nlm.nih.gov/28356097/)]
96. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. *PLoS One* 2013 Dec;8(12):e82901 [FREE Full text] [doi: [10.1371/journal.pone.0082901](https://doi.org/10.1371/journal.pone.0082901)] [Medline: [24349392](https://pubmed.ncbi.nlm.nih.gov/24349392/)]
97. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. *Am J Prev Med* 2016 Nov;51(5):825-832 [FREE Full text] [doi: [10.1016/j.amepre.2016.06.013](https://doi.org/10.1016/j.amepre.2016.06.013)] [Medline: [27745682](https://pubmed.ncbi.nlm.nih.gov/27745682/)]
98. Martin CA, Desphande S, Hekler EB, Rivera DE. A system identification approach for improving behavioral interventions based on social cognitive theory. 2015 Presented at: American Control Conference (ACC); July 1-3, 2015; Chicago, IL USA. [doi: [10.1109/ACC.2015.7172261](https://doi.org/10.1109/ACC.2015.7172261)]
99. Guo P, Rivera DE, Downs DS, Savage JS. Semi-physical identification and state estimation of energy intake for interventions to manage gestational weight gain. 2016 Presented at: American Control Conference (ACC); August 1, 2016; Boston, MA, USA. [doi: [10.1109/ACC.2016.7525092](https://doi.org/10.1109/ACC.2016.7525092)]
100. Martin CA, Rivera DE, Riley WT, Hekler EB, Buman MP, Adams MA, et al. A Dynamical Systems Model of Social Cognitive Theory. In: IEEE. 2014 Presented at: American Control Conference (ACC); July 21, 2014; Portland, OR USA. [doi: [10.1109/ACC.2014.6859463](https://doi.org/10.1109/ACC.2014.6859463)]
101. Riley WT, Martin CA, Rivera DE, Hekler EB, Adams MA, Buman MP, et al. Development of a dynamic computational model of social cognitive theory. *Transl Behav Med* 2016 Dec;6(4):483-495 [FREE Full text] [doi: [10.1007/s13142-015-0356-6](https://doi.org/10.1007/s13142-015-0356-6)] [Medline: [27848208](https://pubmed.ncbi.nlm.nih.gov/27848208/)]
102. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. *Am J Prev Med* 2016 Nov;51(5):825-832 [FREE Full text] [doi: [10.1016/j.amepre.2016.06.013](https://doi.org/10.1016/j.amepre.2016.06.013)] [Medline: [27745682](https://pubmed.ncbi.nlm.nih.gov/27745682/)]
103. 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](https://doi.org/10.1007/s13142-016-0395-7)] [Medline: [27357001](https://pubmed.ncbi.nlm.nih.gov/27357001/)]
104. Oggunnaike B, Ray WH. *Process Dynamics, Modeling, and Control*. Oxford: Oxford University Press; 1994.

105. Dunton GF. Ecological momentary assessment in physical activity research. *Exerc Sport Sci Rev* 2017 Dec;45(1):48-54 [[FREE Full text](#)] [doi: [10.1249/JES.0000000000000092](https://doi.org/10.1249/JES.0000000000000092)] [Medline: [27741022](https://pubmed.ncbi.nlm.nih.gov/27741022/)]
106. Klasnja P, Hekler EB, Korinek EV, Harlow J, Mishra SR. Toward Usable Evidence: Optimizing Knowledge Accumulation in HCI Research on Health Behavior Change. New York, NY, USA: ACM; 2017 Presented at: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2011; Denver, Colorado, USA. [doi: [10.1145/3025453.3026013](https://doi.org/10.1145/3025453.3026013)]
107. Nilsen W, Ertin E, Hekler E, Kumar S, Lee I, Mangharam R. Modeling Opportunities in mHealth Cyber-Physical Systems. In: *Mobile Health: Sensors, Analytic Methods, and Applications*. Basel: Springer International Publishing; 2017.
108. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. *Transl Behav Med* 2015 Sep;5(3):335-346 [[FREE Full text](#)] [doi: [10.1007/s13142-015-0324-1](https://doi.org/10.1007/s13142-015-0324-1)] [Medline: [26327939](https://pubmed.ncbi.nlm.nih.gov/26327939/)]
109. Young M, Plotnikoff R, Collins C, Callister R, Morgan P. Social cognitive theory and physical activity: a systematic review and meta-analysis. *Obes Rev* 2014 Dec;15(12):983-995. [doi: [10.1111/obr.12225](https://doi.org/10.1111/obr.12225)] [Medline: [25428600](https://pubmed.ncbi.nlm.nih.gov/25428600/)]
110. 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 Apr;8(4):e62613 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0062613](https://doi.org/10.1371/journal.pone.0062613)] [Medline: [23638127](https://pubmed.ncbi.nlm.nih.gov/23638127/)]
111. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Effects of three motivationally targeted mobile device applications on initial physical activity and sedentary behavior change in midlife and older adults: a randomized trial. *PLoS One* 2016;11(6):e0156370 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0156370](https://doi.org/10.1371/journal.pone.0156370)] [Medline: [27352250](https://pubmed.ncbi.nlm.nih.gov/27352250/)]
112. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983 Jun;51(3):390-395. [Medline: [6863699](https://pubmed.ncbi.nlm.nih.gov/6863699/)]
113. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. *Health Psychol* 1994;13(1):39-46. [doi: [10.1037/0278-6133.13.1.39](https://doi.org/10.1037/0278-6133.13.1.39)]
114. Dong Y, Rivera DE, Thomas DM, Navarro-Barrientos JE, Downs DS, Savage JS, et al. A dynamical systems model for improving gestational weight gain behavioral interventions. *Proc Am Control Conf* 2012:4059-4064 [[FREE Full text](#)] [Medline: [24309837](https://pubmed.ncbi.nlm.nih.gov/24309837/)]
115. Dong Y, Rivera DE, Downs DS, Savage JS, Thomas DM, Collins LM. Hybrid model predictive control for optimizing gestational weight gain behavioral interventions. *Proc Am Control Conf* 2013:1970-1975 [[FREE Full text](#)] [Medline: [24336314](https://pubmed.ncbi.nlm.nih.gov/24336314/)]
116. Savage JS, Downs DS, Dong Y, Rivera DE. Control systems engineering for optimizing a prenatal weight gain intervention to regulate infant birth weight. *Am J Public Health* 2014 Jul;104(7):1247-1254. [doi: [10.2105/AJPH.2014.301959](https://doi.org/10.2105/AJPH.2014.301959)] [Medline: [24832411](https://pubmed.ncbi.nlm.nih.gov/24832411/)]
117. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *Am Psychol* 2002;57(9):705-717. [doi: [10.1037/0003-066X.57.9.705](https://doi.org/10.1037/0003-066X.57.9.705)]
118. Designkit. The Fieldguide to Human-Centered Design URL: <http://www.designkit.org/resources> [accessed 2018-05-15] [[WebCite Cache ID 6zR71QtQ9](#)]
119. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. *Am J Prev Med* 2009 May;36(5):452-457 [[FREE Full text](#)] [doi: [10.1016/j.amepre.2009.02.002](https://doi.org/10.1016/j.amepre.2009.02.002)] [Medline: [19362699](https://pubmed.ncbi.nlm.nih.gov/19362699/)]
120. Godfrey K. *Perturbation Signals for System Identification*. London, UK: Prentice Hall; 1993.
121. Martin CA, Rivera DE, Hekler EB. An identification test monitoring procedure for MIMO systems based on statistical uncertainty estimation. In: *IEEE. 2015 Presented at: Conference on Decision and Control (CDC); December 15-18, 2015; Osaka, Japan.* [doi: [10.1109/CDC.2015.7402627](https://doi.org/10.1109/CDC.2015.7402627)]
122. Lee H, Rivera DE, Mittelman HD. Constrained minimum crest factor multisine signals for “Plant-Friendly” identification of highly interactive systems. *IFAC Proc Volumes* 2003;36(16):915-920. [doi: [10.1016/S1474-6670\(17\)34877-2](https://doi.org/10.1016/S1474-6670(17)34877-2)]
123. Rivera DE, Lee H, Mittelman HD, Braun MW. Constrained multisine input signals for plant-friendly identification of chemical process systems. *J Process Control* 2009 Apr;19(4):623-635. [doi: [10.1016/j.jprocont.2008.08.006](https://doi.org/10.1016/j.jprocont.2008.08.006)]
124. Hekler EB, Klasnja P, Traver V, Hendriks M. Realizing effective behavioral management of health: the metamorphosis of behavioral science methods. *IEEE Pulse* 2013 Sep;4(5):29-34. [doi: [10.1109/MPUL.2013.2271681](https://doi.org/10.1109/MPUL.2013.2271681)]
125. Martin CA, Deshpande S, Hekler EB, Rivera DE. A system identification approach for improving behavioral interventions based on Social Cognitive Theory. 2015 Presented at: American Control Conference (ACC); July 1-3, 2015; Chicago, IL, USA. [doi: [10.1109/ACC.2015.7172261](https://doi.org/10.1109/ACC.2015.7172261)]
126. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. In: *Ferscha A, Mattern F, editors. Pervasive Computing. Pervasive 2004. Lecture Notes in Computer Science, vol 3001. Berlin, Heidelberg: Springer; 2004:1-17.*
127. Cohen J. A power primer. *Psychol Bull* 1992 Jul;112(1):155-159. [Medline: [19565683](https://pubmed.ncbi.nlm.nih.gov/19565683/)]
128. Lindskog P, Ljung L. Tools for semiphysical modelling. *Int J Adapt Control Signal Process* 1995 Nov;9(6):509-523. [doi: [10.1002/acs.4480090605](https://doi.org/10.1002/acs.4480090605)]
129. Martin CA, Rivera DE, Hekler EB. A decision framework for an adaptive behavioral intervention for physical activity using hybrid model predictive control. 2016 Presented at: American Control Conference (ACC); July 6-8, 2016; Boston, MA, USA. [doi: [10.1109/ACC.2016.7525468](https://doi.org/10.1109/ACC.2016.7525468)]

130. Martin CA, Rivera DE, Hekler EB. A decision framework for an adaptive behavioral intervention for physical activity using hybrid model predictive control. 2016 Presented at: American Control Conference (ACC); July 6-8, 2016; Boston, MA, USA. [doi: [10.1109/ACC.2016.7525468](https://doi.org/10.1109/ACC.2016.7525468)]
131. García CE, Prett D, Morari M. Model predictive control: theory and practice—a survey. *Automatica* 1989 May;25(3):335-348. [doi: [10.1016/0005-1098\(89\)90002-2](https://doi.org/10.1016/0005-1098(89)90002-2)]
132. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. *PLoS One* 2013 Dec;8(12):e82901 [FREE Full text] [doi: [10.1371/journal.pone.0082901](https://doi.org/10.1371/journal.pone.0082901)] [Medline: [24349392](https://pubmed.ncbi.nlm.nih.gov/24349392/)]
133. Riddle M, Science of Behavior Change Working Group. News from the NIH: using an experimental medicine approach to facilitate translational research. *Transl Behav Med* 2015 Dec;5(4):486-488 [FREE Full text] [doi: [10.1007/s13142-015-0333-0](https://doi.org/10.1007/s13142-015-0333-0)] [Medline: [26622921](https://pubmed.ncbi.nlm.nih.gov/26622921/)]
134. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol* 2016 Dec 1;45(6):1776-1786 [FREE Full text] [doi: [10.1093/ije/dyv341](https://doi.org/10.1093/ije/dyv341)] [Medline: [26800751](https://pubmed.ncbi.nlm.nih.gov/26800751/)]
135. Wilson E. *Consilience: The Unity of Knowledge*. New York: Vintage; 1999.
136. Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, et al. Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. *Transl Behav Med* 2014 Dec;4(4):407-423 [FREE Full text] [doi: [10.1007/s13142-014-0262-3](https://doi.org/10.1007/s13142-014-0262-3)] [Medline: [25584090](https://pubmed.ncbi.nlm.nih.gov/25584090/)]
137. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Independence KY USA: Wadsworth Publishing; 2002.
138. Miller G. The smartphone psychology manifesto. *Perspect Psychol Sci* 2012 May;7(3):221-237. [doi: [10.1177/1745691612441215](https://doi.org/10.1177/1745691612441215)] [Medline: [26168460](https://pubmed.ncbi.nlm.nih.gov/26168460/)]
139. Carver C, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. *Psychol Bull* 1982;92(1):111-135. [doi: [10.1037/0033-2909.92.1.111](https://doi.org/10.1037/0033-2909.92.1.111)]
140. Rioch DM. Plans and the structure of behavior. *Arch Gen Psychiatry* 1961 Dec 1;5(6):620-622. [doi: [10.1001/archpsyc.1961.01710180104014](https://doi.org/10.1001/archpsyc.1961.01710180104014)]
141. Flach JM, Bennett KB, Jagacinski RJ, Mulder M, van Paassen M. The closed-loop dynamics of cognitive work. In: *The Oxford Handbook of Cognitive Engineering*. Oxford: Oxford Press; 2013.
142. Norman DA, Stappers PJ. DesignX: complex sociotechnical systems. *She Ji: J Des Econ Innov* 2015;1(2):83-106. [doi: [10.1016/j.sheji.2016.01.002](https://doi.org/10.1016/j.sheji.2016.01.002)]

Abbreviations

ARX: Auto-regressive model with eXogenous Input
mHealth: mobile health
MOST: multiphase optimization strategy
MRT: microrandomization trial
PA: physical activity
PID: Proportional-Integral-Derivative
RCT: randomized controlled trial
RL: reinforcement learning
TTM: Transtheoretical Model
SCT: Social Cognitive Theory
SMART: sequential multiple assignment randomized trial

Edited by G Eysenbach; submitted 01.08.17; peer-reviewed by S Saeb, S Kelders, M Ainsworth; comments to author 23.11.17; revised version received 22.03.18; accepted 03.04.18; published 28.06.18.

Please cite as:

Hekler EB, Rivera DE, Martin CA, Phatak SS, Freigoun MT, Korinek E, Klasnja P, Adams MA, Buman MP
 Tutorial for Using Control Systems Engineering to Optimize Adaptive Mobile Health Interventions
J Med Internet Res 2018;20(6):e214
 URL: <http://www.jmir.org/2018/6/e214/>
 doi: [10.2196/jmir.8622](https://doi.org/10.2196/jmir.8622)
 PMID: [29954725](https://pubmed.ncbi.nlm.nih.gov/29954725/)

©Eric B Hekler, Daniel E Rivera, Cesar A Martin, Sayali S Phatak, Mohammad T Freigoun, Elizabeth Korinek, Predrag Klasnja, Marc A Adams, Matthew P Buman. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>),

28.06.2018. 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.

Review

Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety

Theresa Fleming^{1,2}, PhD; Lynda Bavin², PhD; Mathijs Lucassen³, PhD; Karolina Stasiak², PhD; Sarah Hopkins², PhD; Sally Merry², MD

¹Faculty of Health, Victoria University of Wellington, Wellington, New Zealand

²Department of Psychological Medicine, The University of Auckland, Auckland, New Zealand

³School of Health, Wellbeing and Social Care, The Open University, Milton Keynes, United Kingdom

Corresponding Author:

Theresa Fleming, PhD

Faculty of Health

Victoria University of Wellington

Kelburn Parade

Wellington,

New Zealand

Phone: 64 800 842 8888

Email: terry.fleming@vuw.ac.nz

Abstract

Background: Digital self-help interventions (including online or computerized programs and apps) for common mental health issues have been shown to be appealing, engaging, and efficacious in randomized controlled trials. They show potential for improving access to therapy and improving population mental health. However, their use in the real world, ie, as implemented (disseminated) outside of research settings, may differ from that reported in trials, and implementation data are seldom reported.

Objective: This study aimed to review peer-reviewed articles reporting user uptake and/or ongoing use, retention, or completion data (hereafter *usage data* or, for brevity, *engagement*) from implemented pure self-help (unguided) digital interventions for depression, anxiety, or the enhancement of mood.

Methods: We conducted a systematic search of the Scopus, Embase, MEDLINE, and PsychINFO databases for studies reporting user uptake and/or usage data from implemented digital self-help interventions for the treatment or prevention of depression or anxiety, or the enhancement of mood, from 2002 to 2017. Additionally, we screened the reference lists of included articles, citations of these articles, and the titles of articles published in *Internet Interventions*, *Journal of Medical Internet Research (JMIR)*, and *JMIR Mental Health* since their inception. We extracted data indicating the number of registrations or downloads and usage of interventions.

Results: After the removal of duplicates, 970 papers were identified, of which 10 met the inclusion criteria. Hand searching identified 1 additional article. The included articles reported on 7 publicly available interventions. There was little consistency in the measures reported. The number of registrants or downloads ranged widely, from 8 to over 40,000 per month. From 21% to 88% of users engaged in at least minimal use (eg, used the intervention at least once or completed one module or assessment), whereas 7-42% engaged in moderate use (completing between 40% and 60% of modular fixed-length programs or continuing to use apps after 4 weeks). Indications of completion or sustained use (completion of all modules or the last assessment or continuing to use apps after six weeks or more) varied from 0.5% to 28.6%.

Conclusions: Available data suggest that uptake and engagement vary widely among the handful of implemented digital self-help apps and programs that have reported this, and that usage may vary from that reported in trials. Implementation data should be routinely gathered and reported to facilitate improved uptake and engagement, arguably among the major challenges in digital health.

(*J Med Internet Res* 2018;20(6):e199) doi:[10.2196/jmir.9275](https://doi.org/10.2196/jmir.9275)

KEYWORDS

e-therapy; mobile applications; eHealth; depression; anxiety

Introduction**Background**

Digital interventions (including online or computerized programs or apps) have been shown to be efficacious for depression and anxiety [1-4] and they provide the opportunity to extend psychological therapy to people who might otherwise not receive it [3-5]. Adherence to digital interventions is important for therapeutic gains [3,6] and is generally satisfactory relative to face-to-face interventions [7]. Self-help interventions, provided without guidance or personal support, might reach users who are unable or unwilling to seek help and may be scaled up at lower cost than interventions involving assistance [4,5]. Adherence to self-help is generally lower than that to guided interventions [8], although approaches such as persuasive design and telepresence may enhance retention [5,9] and advantages such as scalability mean that self-help remains worthy of attention.

Interventions may have poorer outcomes when implemented in community or clinical settings than they do in research trials [10-12]. Proven interventions can fail in the real world because translation from research trials may involve changes in the conditions under which the original results were obtained [10,11]. For instance, trials may exclude participants with complex issues, and trial participants may have additional motivations to complete interventions, such as to please researchers or to help others. Trial participants might also benefit from assessment effects or face-to-face contacts that are independent of the intervention [10]. Digital technology is evolving rapidly; hence, digital interventions that are not updated or refined following trials are at risk of becoming dated and, consequently, less appealing by the time they are available outside of research settings [13]. For these reasons, it is important to examine the use of digital interventions in real-world settings. Examining self-help interventions in isolation provides the opportunity to highlight differences between individual interventions of this type.

Objectives

In this study, we aimed to systematically review peer-reviewed articles reporting user uptake (eg, number of users, registrations, or downloads) and/or ongoing use, adherence, retention, or completion data (hereafter *usage data* or, for brevity, *engagement*) from implemented digital self-help interventions for the prevention or treatment of anxiety or depression, or for the enhancement of mood. We note that aspects of engagement other than usage data (eg, emotional involvement) are important [14-16]. However usage data are widely reported, are important for efficacy [6,7], and are the focus here. We identified no prior systematic reviews on this topic.

Methods**Search Strategy**

Electronic searches were conducted of the Scopus, Embase, MEDLINE, and PsychINFO databases. The following search terms were used in Scopus, and the equivalent search was repeated on the Ovid Embase, MEDLINE, and PsychINFO databases:

(TITLE-ABS-KEY (implementation OR “real world” OR real-world OR naturalistic OR observational OR “open access” OR public OR “publicly available” OR “publically available” OR deployment OR community OR nationwide OR national OR regist* OR dissemination) AND TITLE-ABS-KEY (computerized PRE/5 therap*) OR etherap* OR e-therap* OR (online PRE/5 intervention*) OR (online PRE/5 treatment*) OR (internet PRE/5 intervention*) OR (website PRE/5 intervention*) OR (web-based PRE/5 intervention*) OR (web-based PRE/5 treatment*) OR “smartphone app*” OR “mobile app*” OR “smartphone intervention*” OR “smartphone program*” OR “mobile program*” OR “mobile intervention*” OR mhealth OR mtherapy) AND TITLE-ABS-KEY (depression OR anxiety OR mood OR “mental health” OR “psychological wellbeing”) AND TITLE-ABS-KEY (uptake OR adopt* OR regist* OR enrol* OR recruit* OR logon OR “logged on” OR usage OR adherence OR compliance OR complet* OR attrition OR “drop out” OR dropout OR drop-out)).

The search strategy was developed in partnership with a specialist research librarian. The search of the Ovid databases included “mp” (“multi-purpose”), thus incorporated all subject headings in which one or more word(s) matched the search term.

The following journals were hand-searched from their inception (all post 2002) up to and including their February 2017 issue: *Internet Interventions*, *Journal of Medical Internet Research (JMIR)*, and *JMIR Mental Health*. Finally, a hand search was conducted of the included studies’ reference lists, and the titles of articles that had cited the included papers.

Inclusion and Exclusion Criteria

Articles were included in the review if they:

- were digital (computerized or online programs or apps) self-help/unguided interventions explicitly described as being for the prevention or treatment of depression or anxiety, or for the enhancement or improvement of mood;
- reported data on user uptake (eg, number of users, registrations, or downloads) and/or usage, adherence, or attrition (eg, number or percentage of users beginning, completing, or partially completing the intervention, or using the intervention for a specified period of time);
- reported implementation (dissemination/observational) data; and

- were published in the peer-reviewed literature between January 1, 2002, and March 8, 2017.

Articles were excluded if they:

- were pilot, exploratory, or feasibility studies; randomized controlled trials (RCTs); or protocol papers;
- were studies in which users were subject to assessments for research purposes, over and above what would normally be embedded in the intervention (ie, interventions could be included if routinely administered assessments were embedded as part of the self-help tool, but were excluded if users were subject to face-to-face or additional assessments for research purposes);
- reported findings from supported digital interventions (ie, supported by a therapist or where other human support was provided) or interventions that utilized a moderator or that were blended (eg, an adjunct to face-to-face therapy); and/or
- were not available in English.

Study Selection

Two authors (TF and LB) independently screened all retrieved titles, and then read the abstracts of all potentially relevant articles. Articles identified by one or both screening author(s) as potentially relevant were reviewed in full text. For each article excluded at the full-text review, the main reason for exclusion was recorded.

Data Extraction

The characteristics of all the included articles were coded by 2 of the 3 authors (LB, KS, and SH) and checked by 1 author (ML). The authors utilized a data extraction template that was developed for this systematic review and piloted on 2 of the full-text articles. Any discrepancies were resolved by referring to the original article and via discussion. The following characteristics and data were extracted:

- article reference details and data collection period;
- intervention characteristics: name of the intervention, intervention type (eg, online program, computerized program such as CD-ROM, smartphone app), condition treated, therapeutic modality, intervention length, features of gamification and navigation, and whether previously trialled and reported in the peer-reviewed literature;
- number of persons registering for or downloading the intervention;
- registration rate (the percentage of visitors to the intervention's Website who then registered for the intervention), where data allowed;
- indicators of *at least minimal use*, such as number or percentage of users who began or used the intervention at least once or, where those data were not reported, number or percentage of users who completed at least one module or one assessment;
- indicators of *moderate use* (ie, more than *at least minimal use* but less than *completion or sustained use*), such as number or percentage of users completing a specified number or proportion of modules, or number of logins, or use for a specific period of time; and
- indicators of *completion or sustained use*, the number or percentage of users completing the intervention or, where

no end-point was specified, the number or percentage using it for at least 6 weeks. Where neither of these were specified, the number or percentage of users who completed a final assessment or assessment at 6 weeks or more was recorded.

Results

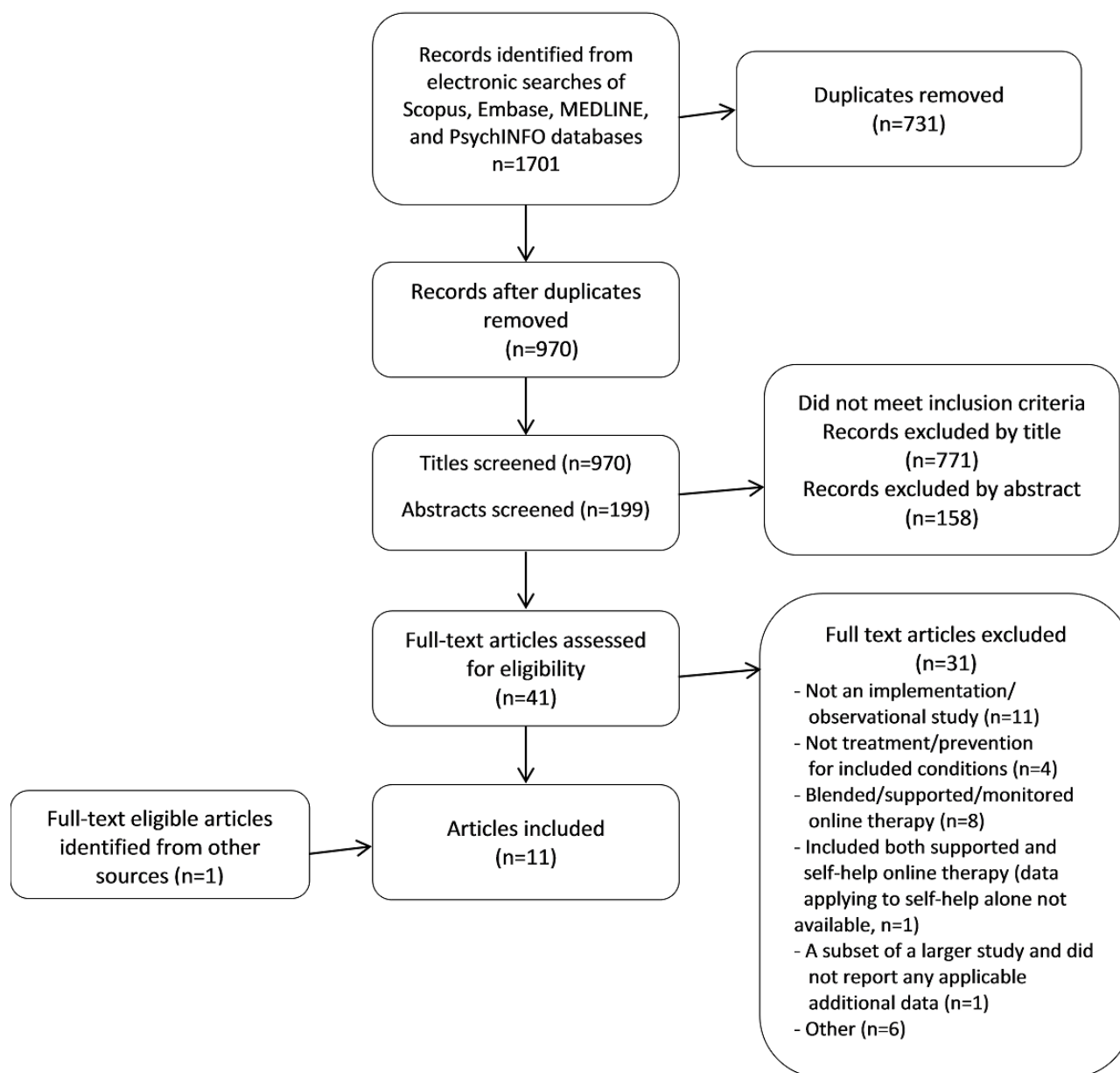
Study Selection

The database search yielded 1701 records, of which 970 remained after the removal of duplicates as shown in [Figure 1](#). The initial title screening excluded 771 records, and the abstract screening excluded a further 158 records. The full texts of the remaining 41 articles were reviewed, of which 10 passed the inclusion and exclusion criteria. Several papers required detailed consideration. A paper by Al-Asadi and colleagues [17] was excluded because it was not possible to separate data for those persons who selected therapist-assisted self-help from those who selected pure self-help. Another paper by Al-Asadi and colleagues [18] was included because results for those receiving pure self-help were provided. A paper by Menzies and colleagues [19] was included, despite being described as a trial, because the intervention was available online without referral, no researcher contact was involved, no assessments beyond those routinely included in the intervention were used, and there was no randomization. An intervention for posttraumatic stress disorder (PTSD) [20] was included as PTSD was classified as an anxiety disorder in the Diagnostic and Statistical Manual, Fourth Edition, Text Revision (DSM-IV-TR) [21], during most of our review period (ie, 2002-2017), although it was reclassified in the fifth edition of the DSM, issued in 2013 [22]. The hand-searching process identified one additional article [23] (Happify) for inclusion in the systematic review, bringing the total of eligible articles to 11.

Study and Program Characteristics

The 11 articles reported implementation data from 7 different interventions. Five articles reported on the original (ie, Mark I) or updated (ie, Mark II) version of the MoodGYM program [24-28], and the remaining 6 each described a unique intervention [18-20,23,29,30].

Study and program characteristics are summarized in [Table 1](#). All of the included interventions were available without referral and were free to the user, apart from one that offered purchases or subscription for some content [23]. All were described as based on cognitive behavioral therapy (CBT) or utilizing CBT among other therapeutic modalities (eg, positive psychology), apart from one that did not specify the modality used [18]. Four interventions were online programs of fixed length, using sequential navigation (where content is provided in a specific order), or a choice of sequential or open navigation. The remaining 3 interventions were available via smartphone as an app [20], a suite of apps [30], or could be accessed as an app or online [23]. These app-based interventions (hereafter *apps*) had no fixed length and used open navigation. Notably, the articles reporting on apps were all relatively recent (ie, from 2015 to 2016).

Figure 1. Flow diagram of article selection.

The efficacy of one program (MoodGYM) has been empirically supported through RCTs [25,31,32]. For 3 interventions there was some evidence: a small, exploratory pre-post trial found beneficial effects for CBTPsych [33]; a single-arm trial, which included coaching, showed significant therapeutic improvements for users of Intellicare apps [30]; and, for PTSD Coach, there was both a pilot RCT reporting modest, nonsignificant effects [34] and a subsequent RCT reporting significant therapeutic effects at post-treatment [35]. At the time of writing, we found no published studies examining the efficacy of 3 interventions: Anxiety Online, Happify, and HDep.

Uptake and Usage Data

User uptake (registrations or downloads) and usage data for each intervention are summarized in Table 2.

The number of registrations or downloads were reported in 8 papers. The remaining 3 studies reported related indicators as shown in Table 2. Registrations or downloads varied markedly from an average of 8 to at least 40,053 per month. Across the 7 interventions (and using the Mark II community user registration rate for the MoodGYM program), the median number of registrations or downloads was 401 per month, and the mean when excluding the lower and upper outliers was 2098 per month. There were 3 interventions with thousands of downloads or registrants per month, including 2 of the 3 apps, PTSD Coach, and Happify [20,23], and 1 of the 4 online programs, MoodGYM Mark II [27].

In Table 2, for MoodGYM, the number of assessments completed is likely to be similar to the number of modules completed, but not interchangeable. Assessments are completed at the beginning of a module and some users may not complete

modules that they commence. Furthermore, for the Mark I version, assessments were not compulsory (a module could be completed without doing the assessment associated with that module); therefore, users may have skipped assessments.

Just 1 study, which was of MoodGYM Mark II [27], reported both the number of website visitors and the number of registrations, allowing the calculation of a registration rate, which was 42.2%.

Available measures indicating *at least minimal use* were limited and varied widely as shown in Table 2, making direct comparisons challenging. Within these limitations, CBTPsych [19] had the highest percentage of registrants engaging in at least minimal use, although given the low number of registrations per month, this was very few individuals.

Next, we reviewed indicators of moderate use. Again, there was little consistency in available data. Of the online programs: up to 16% of MoodGYM users completed 2 or more modules or assessments (embedded within the modules) out of 5 [24-28]; 10% of HDep users completed module 4 of this 7-module program (although they could miss individual modules) [29]; and 39% of CBTPsych users completed at least 4 of the 7 modules [19]. The apps or blended interventions reported quite different measures. Over 40% of PTSD Coach users continued to use it a month after installation [20]. For Happify, 20.6% of those who had completed an initial assessment also completed a noncompulsory assessment 2 weeks later, and 7.2% completed an assessment at 4 weeks [23]. The study of Intellicare apps described those who used each app 10 or more times as “active users.” This group comprised 4.7% to 35.7% of users for each app [30].

Table 1. Intervention characteristics.

Publication	Intervention	Condition treated	Therapeutic modality	Intervention length	Gamification	Navigation	Evidence from prior trials
Christensen et al (2002) [24] Christensen et al (2004) [25]	MoodGYM Mark I (online program)	Depression/mood	CBT ^a	5 modules	Includes an interactive game	Sequential	Significant therapeutic effects in an RCT ^b
Christensen et al (2006) [26] Batterham et al (2008) [27] Neil et al (2009) [28]	MoodGYM Mark II (online program)	Depression/mood	CBT	5 modules	Includes an interactive game	Sequential	Significant therapeutic effects in RCTs
Al-Asadi et al (2014) [18]	Anxiety Online (suite of 5 online programs)	Anxiety	Not specified	12-week program	No apparent gamification ^c	Either	No
Lara et al (2014) [29]	HDep (online program)	Depression/mood	CBT	7 modules	No apparent gamification ^c	Either	No
Menzies et al (2016) [19]	CBTPsych (online program)	Anxiety (social anxiety for stutterers)	CBT	7 modules	No apparent gamification ^c	Sequential	Therapeutic effects in a small pre-post exploratory trial
Owen et al (2015) [20]	PTSD Coach (app)	Anxiety (PTSD)	CBT	No specified length	No apparent gamification ^c	Open	Pilot RCT: nonsignificant effects Subsequent RCT: significant therapeutic effects
Lattie et al (2016) [30]	Intellicare Apps (suite of apps)	Anxiety and depression	Mixed (includes CBT & positive psychology)	No specified length	Includes some gaming elements in one or more apps	Open	Significant effects in a single-arm trial with coaching provided through the program
Carpenter et al (2016) [23]	Happify (app & online program)	Anxiety and mood/depression	Mixed (includes CBT & positive psychology)	58 core activities	Includes some gaming elements and games	Open	No

^aCBT: cognitive behavioral therapy.

^bRCT: randomized controlled trial.

^cOn the basis of the study's description of the program (ie, as at data collection).

Table 2. Uptake and usage data.

Publication	Intervention (data collection period)	Registrations/downloads (time period in months)	Average registrations /downloads per month	At least minimal use	Moderate use	Completion or sustained use
Christensen et al (2002) [24] ^a	MoodGYM Mark I (Apr 2001 to Sep 2001)	2909 (6)	485	51.7% completed at least one depression assessment	16% completed at least two depression assessments	Not stated
Christensen et al (2004) [25]	MoodGYM Mark I (Apr 2001 to Sep 2003)	19,607 (30)	654	62% completed at least one depression assessment	15.6% completed two or more modules	0.5% completed a noncompulsory assessment at beginning of the last module
Christensen et al (2006) [26]	MoodGYM Mark II (Sep 2003 to Oct 2004)	38,791 (14)	2,770	69% completed at least one depression assessment	Less than 7% progressed beyond two modules	Not stated
Batterham et al (2008) [27] ^b	MoodGYM Mark II (Jan 2006 to Apr 2007)	82,159 (16)	5135	37% completed one or more modules	10% completed 2 or more modules	Not stated
Neil et al (2009) [28]	MoodGYM Mark II (Jan 2006 to Nov 2007—adolescents)	7207 (23)	313	40.6% completed one or more modules	11.1% completed 2 or more modules	2.8% completed all 5 modules
Al-Asadi et al (2014) [18]	Anxiety Online (Oct 2009 to Jan 2012)	9394 persons completed assessment ^c (28)	336	33.1% accepted and commenced self-help program ^d	Not stated	3.7% of those who completed the first assessment also completed the post-treatment assessment
Lara et al (2014) [29]	HDep (Mar 2009 to Apr 2013)	17,318 persons registered and entered site at least twice ^e (50)	346	71.4% completed the first module	10% of users did module 4 (users could miss modules, so may not have completed 4)	Not stated
Menzies et al (2016) [19]	CBTPsych (Aug 2011 to Mar 2014)	267 (32)	8	88% logged on at least once	39% completed 4 or more modules	19.5% completed all 7 modules
Owen et al (2015) [20]	PTSD Coach (Mar 2011 to June 2014)	153,834 (36)	4273	61.1% returned to use the app after the day it was installed	52.1% continued to use app 1 week after installation; 41.6% continued to use app 1 month after installation	No specific completion point; however, 28.6% continued to use the app after 3 months, 19.4% continued after 6 months, and 10.6% after 1 year
Lattie et al (2016) [30]	Intellicare Apps (Sep 2014 to Oct 2015)	5210 (13)	401	84.1% of downloaded apps were launched at least once; between 38.7% and 70.2% of users used apps for at least one day	Between 4.7% and 35.7% (depending on specific app) of active users used the app on 10 or more occasions; between 13.1% and 23.3% used the app for 28 or more days	Not stated

Publication	Intervention (data collection period)	Registrations/downloads (time period in months)	Average registrations /downloads per month	At least minimal use	Moderate use	Completion or sustained use
Carpenter et al (2016) [23]	Happify (Dec 2014 to May 2016)	Total downloads not reported. 720,952 persons completed an assessment (18) ^f	40,053	21.2% of those who had completed an assessment at registration completed at least one more assessment ^f	20.6% also completed an assessment at 2 weeks, 7.2% completed an assessment at 4 weeks ^f	3.5% completed an assessment at 6 weeks, 2.1% completed an assessment at 8 weeks

^aThe timeframe covered in this study is a subset of that reported by Christensen et al [25] for the same intervention, but the 2002 study reports some data that are not reported in the 2004 study.

^bThe timeframe reported in this study is a subset of that reported by Neil et al [28] for the same intervention. However, this study reports data for all registrants, whereas Neil et al's [28] study only reports data for adolescent users.

^cPersons (n=9394) completed an online assessment and were then offered an online self-help or therapist-assisted program. No other indications of registration are reported.

^dPersons (n=3107) selected and commenced an online self-help program; they did not formally withdraw and were not recorded as "in progress" at the time of the publication.

^eNo other registration data were reported.

^fAssessments were not compulsory.

Only 2 studies directly reported intervention completion rates. In these cases, 2.8% of MoodGYM users completed all 5 modules [28] and 19.5% of CBTPsych users completed the program [19]. Looking at other indicators of sustained use for the online programs, 0.5% of MoodGYM Mark I users completed a noncompulsory assessment in the final module [25]. In total, 3.7% of Anxiety Online users who had completed an initial assessment and were offered a self-help online intervention also completed a post-treatment assessment [18]; however, these users may or may not have been utilizing the treatment. There were no specific completion data reported for the apps. However, 19.4% of PTSD Coach users continued to use the app after 6 months [20] and 3.5% of Happify users completed a 6-week assessment [23], although again, these users may have completed assessments without engaging in other content.

Combining completion or sustained use data with the number of people beginning each intervention (uptake), as reported here, suggests that fewer than 40 persons per month completed final assessments or final modules for any of the online modular programs during the study periods. In contrast, over 800 users completed the Happify app assessments at 8 weeks [23], and over 1000 persons per month demonstrated continued use of PTSD Coach after 3 months [20].

Discussion

Summary of Evidence

For digital mental health interventions to have a population-level impact, significant numbers of people must receive beneficial doses. This requires both sufficient uptake and ongoing use of effective interventions. Despite over 10,000 digital mental health interventions being publicly available in 2017 [36], we identified only 11 peer-reviewed publications reporting uptake and/or

usage data from publicly available digital self-help interventions for depression, anxiety, or low mood. This is disappointing, given 3 considerations. First, people may use interventions differently in the real world, as compared with trial conditions. Second, digital interventions allow relative ease of data collection through automation. Third, comparisons of uptake and usage across interventions could inform improvements in the field. Where data have been reported, diverse measures were used, making direct comparisons challenging. Nevertheless, large differences are apparent. The widest-reaching intervention in our review had tens of thousands of new users per month, whereas the least used one had fewer than ten. Moreover, ongoing use ranged from less than 1% to over 28% of users completing interventions or demonstrating sustained use.

The findings suggest that people may use digital mental health interventions differently in real-world settings, as compared with trial conditions. Although this may be true for many interventions, the phenomenon is easily quantifiable in digital interventions through embedded routine data collection. Completion rates, as reported in included studies, are lower than the completion rates of 43% to 99% in a systematic review of adherence in controlled trials of online interventions for depression and anxiety [7]. Direct comparisons between research trials and implementation usage data for the same intervention also suggest reduced adherence in real-world settings. For example, only 0.5% of community users (ie, users of MoodGYM as publicly implemented and freely available online) of MoodGYM completed a noncompulsory final assessment, compared with 22.5% of participants in a trial evaluating the same program [31]. Similarly, in the community, adolescent users of MoodGYM completed an average of 3.1 exercises, compared with an average of 9.4 exercises among adolescents in a school-based trial [28].

Textbox 1. Recommended reporting of real-world implementation data.

The following data should be reported from implemented digital mental health interventions:

- total number of registrants or downloads over a specified time period
- the characteristics (such as demographic details) of registrants and users where available
- the number of modules/levels or activities that can be completed and that are completed by users
- number of times the intervention has been accessed and/or the amount of time the user logged on
- the number or percentage of persons completing a “therapeutic dose” of the intervention
- clinical change or effectiveness measures as well as number or percentage of users with clinically significant improvements and deteriorations.

Previous studies have highlighted differences in adherence between guided and unguided self-help interventions [8]. Given the limited, heterogeneous data, we did not conduct meta-analyses or test for differences between interventions. However, on the face of it, a long established program (MoodGYM released in 2001) [24] and 2 of the recently tested app-based interventions (PTSD Coach [20] and Happify [23]) appear particularly promising in terms of uptake. Two interventions reported high sustained use. The first was CBTPsych [19], the 7-module online CBT program for addressing social anxiety among stutterers. Very few people used this, but retention was high. Second, a very different intervention, PTSD coach (available as an app, with no fixed length and an open navigation structure), had high sustained use [20]. Over 10 times as many people engaged in sustained use of apps as completed any of the online modular programs.

An issue for consideration is that of what is a beneficial dose at a population level. The dose or amount of exposure to digital mental health interventions for clinically significant effects has been considered in previous research, with greater adherence generally associated with greater clinical gains [6,7,9]. Included studies show that large numbers of people accessed some mental health interventions for brief periods. Relatively brief use might have a significant population impact if this exposes large numbers of people to ideas such as depression being common, there being a range of ways to address it, and help being available. Future research should consider this.

Perhaps, the strongest implication from this study is that future research should report intervention uptake, ongoing use, and impact in real-world settings. Transparent reporting of key data, such as those as shown in [Textbox 1](#), would facilitate comparisons. Alongside these data, reporting of intervention characteristics, modes of delivery, and features of implementation such as marketing and methods of dissemination would provide opportunities for understanding which interventions, applied in which ways, engage and retain users.

Strengths and Limitations

This review examined data from peer-reviewed articles. We did not examine grey literature or request data from providers of

interventions or Internet service providers due to resource constraints. Further research should explore this. There is a risk of publication bias, given that interventions with poor results may not be reported. A meta-analysis was not conducted due to the small number of published studies and the heterogeneity of data. However, with increased data and more consistent reporting, a meta-analysis would be a valuable future addition to the literature. We set inclusion and exclusion criteria to focus on interventions addressing the very common issues of depression, anxiety, and low mood. We included an intervention targeting PTSD. PTSD was included as an anxiety disorder in the edition of the Diagnostic and Statistical Manual/DSM-IV-TR [21], which was in use at the start of our review period (2002-2017), but not in the fifth edition, issued in 2013 [22]. Others might have made different decisions; however, we have endeavored to be transparent in this. These limitations notwithstanding, this is the first systematic review of implementation data in this area, and it highlights valuable opportunities for development.

Conclusion

Digital self-help interventions targeting depression, anxiety, or the enhancement of mood have the potential to improve population-level mental health in a highly scalable manner. However, for these interventions to achieve meaningful impact, they need to have adequate uptake and adherence in real-world settings. Only a handful of interventions have reported this information in the peer-reviewed literature to date, and these utilized diverse measures. Nevertheless, the published studies of unguided self-help interventions for anxiety, depression, and mood demonstrate large differences in uptake and engagement between interventions. Organizations delivering these interventions should take advantage of the opportunity to gather and publish data. Much of the data collection on intervention usage can be automated, making such collection and subsequent reporting generally easy and low-cost. We have proposed key metrics that should be considered. Transparent, comparable, and timely publication of real-world data would allow between-program comparisons and hence facilitate improvements in user uptake and engagement, arguably 2 of the major challenges in the digital health world.

Acknowledgments

No specific funding was received for the development of this paper. The authors are funded by the University of Auckland (TF, LB, KS, SH, and SM), Victoria University of Wellington (TF), and The Open University (ML).

Conflicts of Interest

TF, KS, ML, and SM are codevelopers of SPARX (a form of computerized therapy for depression cited in this paper). The intellectual property for SPARX is owned by Uniservices at the University of Auckland. The codevelopers can benefit financially from its commercialization outside New Zealand. The remaining authors have no conflicts of interest to declare.

References

- 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](https://doi.org/10.1371/journal.pone.0013196)] [Medline: [20967242](https://pubmed.ncbi.nlm.nih.gov/20967242/)]
- 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;10(3):e0119895 [FREE Full text] [doi: [10.1371/journal.pone.0119895](https://doi.org/10.1371/journal.pone.0119895)] [Medline: [25786025](https://pubmed.ncbi.nlm.nih.gov/25786025/)]
- 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](https://doi.org/10.1016/j.cpr.2012.02.004)] [Medline: [22466510](https://pubmed.ncbi.nlm.nih.gov/22466510/)]
- Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. *J Med Internet Res* 2017 Mar 24;19(3):e90 [FREE Full text] [doi: [10.2196/jmir.7111](https://doi.org/10.2196/jmir.7111)] [Medline: [28341617](https://pubmed.ncbi.nlm.nih.gov/28341617/)]
- 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;7:65 [FREE Full text] [Medline: [27148094](https://pubmed.ncbi.nlm.nih.gov/27148094/)]
- 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](https://doi.org/10.2196/jmir.1772)] [Medline: [21821503](https://pubmed.ncbi.nlm.nih.gov/21821503/)]
- 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](https://doi.org/10.2196/jmir.1194)] [Medline: [19403466](https://pubmed.ncbi.nlm.nih.gov/19403466/)]
- Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. *Internet Interv* 2014 Oct;1(4):205-215 [FREE Full text] [doi: [10.1016/j.invent.2014.08.003](https://doi.org/10.1016/j.invent.2014.08.003)] [Medline: [25904163](https://pubmed.ncbi.nlm.nih.gov/25904163/)]
- Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. *Transl Behav Med* 2018 Feb 17:1-6. [doi: [10.1093/tbm/ibx037](https://doi.org/10.1093/tbm/ibx037)] [Medline: [29471424](https://pubmed.ncbi.nlm.nih.gov/29471424/)]
- Machin D, Day S, Green S, editors. *Textbook of Clinical Trials*. Chichester, West Sussex: Wiley-Blackwell; 2007.
- Marchand E, Stice E, Rohde P, Becker CB. Moving from efficacy to effectiveness trials in prevention research. *Behav Res Ther* 2011 Jan;49(1):32-41 [FREE Full text] [doi: [10.1016/j.brat.2010.10.008](https://doi.org/10.1016/j.brat.2010.10.008)] [Medline: [21092935](https://pubmed.ncbi.nlm.nih.gov/21092935/)]
- Staples LG, Fogliati VJ, Dear BF, Nielssen O, Titov N. Internet-delivered treatment for older adults with anxiety and depression: implementation of the Wellbeing Plus Course in routine clinical care and comparison with research trial outcomes. *BJPsych Open* 2016 Sep;2(5):307-313. [doi: [10.1192/bjpo.bp.116.003400](https://doi.org/10.1192/bjpo.bp.116.003400)] [Medline: [27703794](https://pubmed.ncbi.nlm.nih.gov/27703794/)]
- Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RC, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: A school-based program (Op Volle Kracht) and a computerized program (SPARX). *Behav Res Ther* 2016 May;80:33-42. [doi: [10.1016/j.brat.2016.03.005](https://doi.org/10.1016/j.brat.2016.03.005)] [Medline: [27019280](https://pubmed.ncbi.nlm.nih.gov/27019280/)]
- Cruz-Benito J, Therón R, García-Peñalvo FJ, Pizarro Lucas E. Discovering usage behaviors and engagement in an educational virtual world. *Comput Human Behav* 2015 Jun;47:18-25 [FREE Full text] [doi: [10.1016/j.chb.2014.11.028](https://doi.org/10.1016/j.chb.2014.11.028)] [Medline: [25904163](https://pubmed.ncbi.nlm.nih.gov/25904163/)]
- Curtis HR, Halverson LR, Graham CR. Measuring student engagement in technology-mediated learning: a review. *Comput Educ* 2015;90(1):36-53 [FREE Full text] [doi: [10.1016/j.compedu.2015.09.005](https://doi.org/10.1016/j.compedu.2015.09.005)]
- Smith BG, Gallicano TD. Terms of engagement: analyzing public engagement with organizations through social media. *Comput Human Behav* 2015 Dec;53(4):82-90 [FREE Full text] [doi: [10.1016/j.chb.2015.05.060](https://doi.org/10.1016/j.chb.2015.05.060)] [Medline: [25904163](https://pubmed.ncbi.nlm.nih.gov/25904163/)]
- Al-Asadi AM, Klein B, Meyer D. Pretreatment attrition and formal withdrawal during treatment and their predictors: an exploratory study of the anxiety online data. *J Med Internet Res* 2014;16(6):e152 [FREE Full text] [doi: [10.2196/jmir.2989](https://doi.org/10.2196/jmir.2989)] [Medline: [24938311](https://pubmed.ncbi.nlm.nih.gov/24938311/)]
- Al-Asadi AM, Klein B, Meyer D. Posttreatment attrition and its predictors, attrition bias, and treatment efficacy of the anxiety online programs. *J Med Internet Res* 2014;16(10):e232 [FREE Full text] [doi: [10.2196/jmir.3513](https://doi.org/10.2196/jmir.3513)] [Medline: [25316533](https://pubmed.ncbi.nlm.nih.gov/25316533/)]
- Menzies R, O'Brian S, Lowe R, Packman A, Onslow M. International Phase II clinical trial of CBTPsych: a standalone Internet social anxiety treatment for adults who stutter. *J Fluency Disord* 2016 Jun;48:35-43. [doi: [10.1016/j.jfludis.2016.06.002](https://doi.org/10.1016/j.jfludis.2016.06.002)] [Medline: [27498893](https://pubmed.ncbi.nlm.nih.gov/27498893/)]

20. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. *JMIR Ment Health* 2015;2(1):e7 [FREE Full text] [doi: [10.2196/mental.3935](https://doi.org/10.2196/mental.3935)] [Medline: [26543913](https://pubmed.ncbi.nlm.nih.gov/26543913/)]
21. American Psychiatric Association. Diagnostic statistical manual of mental disorders (4th edition, text revision). Washington, DC: American Psychiatric Association; 2000.
22. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th edition). Washington, DC: American Psychiatric Association; 2013.
23. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Seeing the “big” picture: big data methods for exploring relationships between usage, language, and outcome in internet intervention data. *J Med Internet Res* 2016 Aug 31;18(8):e241 [FREE Full text] [doi: [10.2196/jmir.5725](https://doi.org/10.2196/jmir.5725)] [Medline: [27580524](https://pubmed.ncbi.nlm.nih.gov/27580524/)]
24. 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](https://doi.org/10.2196/jmir.4.1.e3)] [Medline: [11956035](https://pubmed.ncbi.nlm.nih.gov/11956035/)]
25. 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](https://doi.org/10.2196/jmir.6.4.e46)] [Medline: [15631970](https://pubmed.ncbi.nlm.nih.gov/15631970/)]
26. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an internet-based cognitive behaviour therapy program. *Aust N Z J Psychiatry* 2006 Jan;40(1):59-62. [doi: [10.1111/j.1440-1614.2006.01743.x](https://doi.org/10.1111/j.1440-1614.2006.01743.x)] [Medline: [16403040](https://pubmed.ncbi.nlm.nih.gov/16403040/)]
27. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. *Patient Prefer Adherence* 2008;2:97-105 [FREE Full text] [Medline: [19920949](https://pubmed.ncbi.nlm.nih.gov/19920949/)]
28. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. *J Med Internet Res* 2009;11(1):e6 [FREE Full text] [doi: [10.2196/jmir.1050](https://doi.org/10.2196/jmir.1050)] [Medline: [19275982](https://pubmed.ncbi.nlm.nih.gov/19275982/)]
29. Lara MA, Tiburcio M, Aguilar AA, Sánchez-Solís A. A four-year experience with a web-based self-help intervention for depressive symptoms in Mexico. *Rev Panam Salud Publica* 2014;35(5-6):399-406 [FREE Full text] [Medline: [25211568](https://pubmed.ncbi.nlm.nih.gov/25211568/)]
30. 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](https://doi.org/10.1016/j.invent.2016.06.003)] [Medline: [27398319](https://pubmed.ncbi.nlm.nih.gov/27398319/)]
31. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. *Br Med J* 2004 Jan 31;328(7434):265 [FREE Full text] [doi: [10.1136/bmj.37945.566632.EE](https://doi.org/10.1136/bmj.37945.566632.EE)] [Medline: [14742346](https://pubmed.ncbi.nlm.nih.gov/14742346/)]
32. Mackinnon A, Griffiths KM, Christensen H. Comparative randomised trial of online cognitive-behavioural therapy and an information website for depression: 12-month outcomes. *Br J Psychiatry* 2008 Feb;192(2):130-134 [FREE Full text] [doi: [10.1192/bjp.bp.106.032078](https://doi.org/10.1192/bjp.bp.106.032078)] [Medline: [18245031](https://pubmed.ncbi.nlm.nih.gov/18245031/)]
33. Helgadóttir FD, Menzies RG, Onslow M, Packman A, O'Brian S. A standalone internet cognitive behavior therapy treatment for social anxiety in adults who stutter: CBTpsych. *J Fluency Disord* 2014 Sep;41:47-54. [doi: [10.1016/j.jfludis.2014.04.001](https://doi.org/10.1016/j.jfludis.2014.04.001)] [Medline: [25173456](https://pubmed.ncbi.nlm.nih.gov/25173456/)]
34. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD coach app: a pilot randomized controlled trial with community trauma survivors. *Psychol Trauma* 2016 Jan 25;8(3):384-392. [doi: [10.1037/tra0000092](https://doi.org/10.1037/tra0000092)] [Medline: [27046668](https://pubmed.ncbi.nlm.nih.gov/27046668/)]
35. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. *J Consult Clin Psychol* 2017 Mar;85(3):267-273. [doi: [10.1037/ccp0000163](https://doi.org/10.1037/ccp0000163)] [Medline: [28221061](https://pubmed.ncbi.nlm.nih.gov/28221061/)]
36. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. *JAMA Psychiatry* 2017 May 01;74(5):437-438. [doi: [10.1001/jamapsychiatry.2017.0262](https://doi.org/10.1001/jamapsychiatry.2017.0262)] [Medline: [28384700](https://pubmed.ncbi.nlm.nih.gov/28384700/)]

Abbreviations

CBT: cognitive behavioral therapy

JMIR: Journal of Medical Internet Research

RCTs: randomized controlled trials

Edited by G Eysenbach; submitted 26.10.17; peer-reviewed by C Cheek, L Donkin, A Werner-Seidler, D Mohr, D Erbe, E Lattie, C Stiles-Shields; comments to author 07.12.17; revised version received 28.02.18; accepted 17.03.18; published 06.06.18.

Please cite as:

Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S

Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety

J Med Internet Res 2018;20(6):e199

URL: <http://www.jmir.org/2018/6/e199/>

doi: [10.2196/jmir.9275](https://doi.org/10.2196/jmir.9275)

PMID:

©Theresa Fleming, Lynda Bavin, Mathijs Lucassen, Karolina Stasiak, Sarah Hopkins, Sally Merry. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.06.2018. 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.

Original Paper

Online Lifestyle Modification Intervention: Survey of Primary Care Providers' Attitudes and Views

Reem M Hanna¹, MD; Gary Fischer², MD; Molly B Conroy³, MPH, MD; Cindy Bryce⁴, PhD; Rachel Hess⁵, MS, MD; Kathleen McTigue², MPH, MD

¹Division of Hospital Medicine, University of Colorado School of Medicine, Aurora, CO, United States

²Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA, United States

³Division of General Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, United States

⁴Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States

⁵Department of Population Health Science, University of Utah Health Sciences, Salt Lake City, UT, United States

Corresponding Author:

Kathleen McTigue, MPH, MD
Division of General Internal Medicine
University of Pittsburgh
230 McKee Place
Suite 600
Pittsburgh, PA, 15213
United States
Phone: 1 412 692 4843
Fax: 1 412 692 4838
Email: kmm34@pitt.edu

Abstract

Background: Online tools are a convenient and effective method of delivering lifestyle interventions to obese adult primary care patients. A referral model allows physicians to efficiently direct their patients to the intervention during a primary care visit. However, little is known of physicians' perspectives and utilization of the referral model for an online lifestyle modification intervention.

Objective: The aim was to evaluate the response of primary care providers (PCPs) to a referral model for implementing a year-long online intervention for weight loss to obese adult patients.

Methods: The PCPs at six primary care clinics were asked to refer adult obese patients to a year-long online lifestyle intervention providing self-management support for weight loss. Following the 1-year intervention, all providers at the participating practices were surveyed regarding their views of the program. Respondents completed survey items assessing their attitudes regarding the 1-year intensive weight loss intervention and identifying resources they would find helpful for assisting patients with weight loss. Referring physicians were asked about their level of satisfaction with implementing the counseling services using standard electronic health record referral processes. Attitudes toward obesity counseling among referring and nonreferring providers were compared. Impressions of how smoothly the referral model of obesity treatment integrated with the clinical workflow were also quantified.

Results: Of the 67 providers who completed the surveys, nonreferring providers (n=17) were more likely to prefer counseling themselves ($P=.04$) and to report having sufficient time to do so ($P=.03$) than referring providers (n=50) were. Nonreferring providers were more likely to report that their patients lacked computer skills (76%, 13/17 vs 34%, 17/50) or had less access to the Internet (65%, 11/17 vs 32%, 16/50).

Conclusions: Understanding providers' views and barriers regarding the integration of online tools will facilitate widespread implementation of an online lifestyle modification intervention.

(*J Med Internet Res* 2018;20(6):e167) doi:[10.2196/jmir.8616](https://doi.org/10.2196/jmir.8616)

KEYWORDS

online intervention; obesity; health information technology; referral model

Introduction

To improve patients' long-term health and to decrease health care costs, health care providers must play a pivotal role in addressing obesity. The US Preventive Services Task Force recommends screening all adults for obesity. For patients with obesity—a body mass index (BMI) of 30 kg/m^2 or higher—clinicians should offer or refer patients to a comprehensive behavioral intervention [1,2]. Although primary care providers (PCPs) are aware of the importance of addressing obesity in their patient population, counseling in the primary care setting remains uncommon [3]. In one study, not even a third (29%) of obese adults had a recorded diagnosis of obesity, and only 18% had received counseling for weight reduction, whereas 21% had received advice on physical activity and 25% had received advice on diet [4].

Numerous barriers account for these shortcomings, including the absence of practice tools, lack of training in behavior modification strategies, prohibitive costs, and time constraints [3]. Guidelines recommend using a trained interventionist, such as a registered dietitian, psychologist, exercise specialist, health counselor, or other health professional who adheres to formal protocols in weight management, to provide a comprehensive lifestyle modification intervention that is high intensity, with at least 14 sessions over a 6-month timeframe [5]. Unfortunately, many physicians lack the training or clinical infrastructure to comply with these recommendations.

Health information technology is considered a promising avenue for addressing gaps in care and providing patient-centered, yet affordable, health care services [6]. Online tools are considered a convenient and effective option for delivering lifestyle interventions [7]. However, studies on providers' perspectives and use of online lifestyle interventions are limited. Providers may not recommend online tools for their patients because they are not familiar with them or may be unconvinced of their efficacy or safety [8]. To improve widespread implementation of Internet tools, such tools must be acceptable to providers and smoothly integrated into their practice. The tools must also be easily implemented in a wide variety of settings at an affordable cost and be sustainable over time [9]. Integration of a health information technology intervention into practice must also maintain workflow and minimize disruption [10].

To enhance the adoption of the clinical intervention, careful implementation is essential for executing the program and achieving accurate results. A referral model, in which the physician initiates a referral for an online weight-loss intervention during the primary care visit, is one approach for integrating evidence-based behavioral care into practice. Physician referrals have been associated with positive outcomes in weight management, such as improved patient attendance in consultation sessions and better completion rates of weight management programs relative to self-referred patients [11]. The goal of this study was to evaluate physician response to a referral model for implementing a year-long online intervention providing self-management support for weight loss to obese adult primary care patients.

Methods

Parent Study

As part of a randomized controlled trial, three online lifestyle interventions were implemented in six primary care clinics and compared among 373 patients over 1 year of follow-up [12]. The clinics represented a range of practice settings in Western Pennsylvania, USA, including academic, private, rural, and urban. All the practices were part of the University of Pittsburgh Medical Center health system, which is an integrated health enterprise that is headquartered in Pittsburgh, PA. The clinical sites serve a patient population that is primarily white or African American, which is consistent with the region's racial and ethnic distribution. A study investigator met with clinicians at each of the participating sites before the intervention to explain the structure and goals of the intervention and evaluation process. The PCPs were asked to refer patients who met the inclusion criteria: BMI of 30 kg/m^2 or higher, aged 21 to 75 years, and receiving primary care at one of the participating practices. Patients were excluded if they had a myocardial infarction within the last 3 months or their PCP felt that a low-fat diet and moderately intense unsupervised physical activity would be unsafe. Additional exclusion criteria included having a health condition that was likely to impact body weight assessment (eg, severe congestive heart failure or ascites) or influence body weight (eg, cancer requiring treatment in the past year except for nonmelanomatous skin cancer), regular use of prescription medication that is likely to influence body weight, current pregnancy or a plan for pregnancy during the study, an inability to learn adequately from English language audio-taped materials, a perceived lack of basic computer or Internet skills, or a perceived lack of high-speed Internet access. A PCP referral was required for enrollment and only one individual per household could be enrolled.

Physicians used the practice's standard electronic health record (EHR)-based referral and consultant feedback systems. To facilitate enrollment, the EHR system prompted clinicians to consider referring patients who met the inclusion criteria of obesity ($\text{BMI} \geq 30 \text{ kg/m}^2$) [13]. While the PCPs provided a referral, the online coach and automated curriculum provided the comprehensive counseling to each patient. The PCPs received feedback on the weight change of each participant over the course of the year via one-page automated graphic reports of weight and behavior change, printed on paper and annotated with adherence data. This approach is similar to the use of paper consultant feedback letters. Referring PCPs were recommended to provide patients enrolled in the intervention with feedback and encouragement at their routine appointment. Institutional review board approval was obtained for this study from the University of Pittsburgh (PRO09080118).

Primary Care Provider Attitude Survey

The PCPs at the six participating primary care practices were surveyed regarding their views of the program following the 1-year intervention. The survey was anonymous. All providers at the participating practices were contacted by email with survey links, and paper surveys were distributed at practice meetings and resident seminars. Nonresponders received

electronic reminder prompts on multiple occasions. Responders could only complete the survey once and were asked to indicate whether or not they referred patients to the intervention.

The survey included questions about the providers' attitudes toward obesity counseling using a five-point Likert scale anchored at strongly disagree (0) and strongly agree (4). Referring providers were asked whether the EHR referral approach to enrollment and the processes for providing feedback on their patients' progress with lifestyle change integrated smoothly with their clinical workflow. They were also asked whether they provided their referred patients with feedback on their progress toward lifestyle change. All respondents were also asked about their preferences regarding potential clinical resources for maintaining weight loss.

Data Analysis

The PCPs were divided into "referring" and "nonreferring" categories based on whether or not they referred at least one patient to the study. Cross-sectional analyses examined sample demographics, compared attitudes toward obesity counseling, and quantified impressions of how smoothly the referral model of obesity treatment integrated with clinical workflow between referring and nonreferring providers. Data were summarized with descriptive statistics. Chi-square tests were used to examine the relationship between referral categories and providers' attitudes or referring status. Descriptive statistics including means and standard deviation were used. All analyses used Stata 11.1.

Results

Of 185 providers in the six participating primary care practices, 67 (36.2%) completed the provider surveys (Table 1). Among respondents, 99% (66/67) were physicians and 46% (31/67) were females. The majority were white (72%, 48/67) or Asian (22%, 15/67); 4% (3/67) were Hispanic. Fifty PCPs (75%) had referred at least one patient to the program. When comparing referring providers with nonreferring providers, the two groups did not differ in sex or race/ethnicity, but did vary in training

status—71% (12/17) of the nonreferring providers and 12% (6/50) of referring providers were resident physicians ($P<.001$). Among referring providers, most agreed that the referral approach to enrollment (94%, 47/50) and the process of providing 1-year follow-up reports on lifestyle progress (80%, 40/50) integrated smoothly with their normal workflow (Table 2). However, only approximately half of referring providers (52%, 26/52) reported that they typically provided their patients with feedback regarding their efforts to change their lifestyle or body weight throughout the intervention.

Referring and nonreferring providers differed in their counseling preferences ($P=.04$) and perception of whether a clinical encounter provides sufficient time to counsel patients on lifestyle decisions ($P=.03$; Table 2). For example, compared with referring providers, nonreferring providers more often reported a preference for counseling on healthy eating and exercise patterns themselves rather than referring for counseling (somewhat agree: 8%, 4/50; strongly agree: 0%, 0/50) versus 24% (4/17) and 6% (1/17), respectively. Nonreferring providers were more likely to report sufficient time during clinic visits to counsel patients adequately on diet, physical activity, and obesity compared to the referring providers ($P=.03$). Overall, 29% (5/17) of nonreferring providers agreed (somewhat or strongly) that time was sufficient, whereas 8% (4/50) of referring providers agreed (somewhat or strongly). Both referring and nonreferring providers reported that they believed their patients would benefit from advice to lose weight through lifestyle changes even though 18% (12/67) of responders somewhat agreed and 1% (1/67) strongly agreed that their patients were generally not interested in receiving counseling for diet, physical activity, or weight loss. In addition, 19% (13/67) reported that obesity should be managed outside the clinical setting. Nonreferring providers more often reported (somewhat or strongly) that their patients were generally not interested in using Internet-based lifestyle counseling ($P=.01$). Referring and nonreferring providers also differed in whether they raised concerns about their patients' computer skills ($P<.001$) or Internet access ($P=.04$).

Table 1. Demographics of provider respondents (referring providers vs nonreferring providers).

Characteristic	Referring providers (n=50), n (%)	Nonreferring providers (n=17), n (%)	Total (N=67), n (%)	P value
Gender (female)	26 (52)	5 (29)	31 (46)	.11
Ethnicity (Latino)	2 (4)	1 (6)	3 (4)	.75
Race				.10
White	38 (76)	10 (59)	48 (72)	
Black	0 (0)	1 (6)	1 (1)	
Asian	11 (22)	4 (24)	15 (22)	
Other	1 (2)	2 (12)	3 (4)	
Professional status				<.001
Resident	6 (12)	12 (71)	18 (27)	
Fellow	1 (2)	2 (12)	3 (5)	
Attending	43 (86)	2 (12)	45 (67)	
Nurse practitioner	0 (0)	1 (6)	1 (1)	

Table 2. Attitudes toward obesity counseling among referring and nonreferring providers. N/A: not applicable.

Attitudes and category	Referring providers (n=50), n (%)	Nonreferring providers (n=17), n (%)	Total (N=67), n (%)	P value
Report that referral approach to enrollment integrated smoothly^a				N/A
Yes	47 (94)	N/A	N/A	
No	3 (6)	N/A	N/A	
Provided feedback to patients regarding their efforts throughout the intervention^a				N/A
Yes	26 (52)	N/A	N/A	
No	24 (48)	N/A	N/A	
Prefer to counsel on healthy eating and exercise patterns, without referring for additional counseling				.04
Strongly disagree	19 (39)	2 (12)	21 (32)	
Somewhat disagree	26 (53)	10 (59)	36 (55)	
Somewhat agree	4 (8)	4 (24)	8 (12)	
Strongly agree	0 (0)	1 (6)	1 (2)	
There is typically sufficient time during the appointment to counsel patients adequately on diet, physical activity, and obesity				.03
Strongly disagree	31 (63)	5 (29)	36 (55)	
Somewhat disagree	14 (29)	7 (41)	21 (32)	
Somewhat agree	3 (6)	5 (29)	8 (12)	
Strongly agree	1 (2)	0 (0)	1 (2)	
Most of his/her patients would not benefit from advice to lose weight through lifestyle modification				.47
Strongly disagree	10 (59)	31 (63)	41 (62)	
Somewhat disagree	5 (29)	8 (16)	13 (20)	
Somewhat agree	2 (12)	6 (12)	8 (12)	
Strongly agree	0 (0)	4 (8)	4 (6)	
Patients are generally not interested in receiving counseling for diet, physical activity, and weight loss				.45
Strongly disagree	18 (37)	3 (18)	21 (32)	
Somewhat disagree	22 (45)	10 (59)	32 (48)	
Somewhat agree	8 (16)	4 (24)	12 (18)	
Strongly agree	1 (2)	0 (0)	1 (1)	
Obesity should be managed outside the clinical setting				.53
Strongly disagree	18 (37)	3 (18)	21 (32)	
Somewhat disagree	22 (45)	10 (59)	32 (49)	
Somewhat agree	6 (12)	3 (18)	9 (14)	
Strongly agree	3 (6)	1 (6)	4 (6)	
Report patients are generally not interested in using Internet-based lifestyle counseling				.01
Strongly disagree	9 (18)	0 (0)	9 (14)	
Somewhat disagree	29 (59)	9 (53)	38 (58)	
Somewhat agree	11 (22)	5 (29)	16 (24)	
Strongly agree	0 (0)	3 (18)	3 (5)	
Report patients generally have minimal computer skills				<.001
Strongly disagree	7 (14)	1 (6)	8 (12)	
Somewhat disagree	25 (51)	3 (18)	28 (42)	
Somewhat agree	17 (35)	9 (53)	26 (39)	
Strongly agree	0 (0)	4 (24)	4 (6)	

Attitudes and category	Referring providers (n=50), n (%)	Nonreferring providers (n=17), n (%)	Total (N=67), n (%)	P value
Report patients likely lack Internet access				.04
Strongly disagree	8 (16)	1 (6)	9 (14)	
Somewhat disagree	25 (51)	5 (29)	30 (45)	
Somewhat agree	14 (29)	7 (41)	21 (32)	
Strongly agree	2 (4)	4 (24)	6 (9)	

^aOnly primary care providers that referred patients to the intervention were asked to comment on the noted questions in the survey.

Table 3. Physicians' preferred resources to be offered in clinic to assist patients with weight loss.

Resources	Referring providers (n=50), n (%)	Nonreferring providers (n=17), n (%)	Total (N=67), n (%)	P value
In-person visit with a health educator or coach				.21
Yes	36 (73)	15 (88)	51 (77)	
No	13 (27)	2 (12)	15 (23)	
Remainder phone calls from clinic staff				.21
Yes	36 (73)	15 (88)	51 (77)	
No	13 (27)	2 (12)	15 (23)	
Internet/email contact from health educator or coach				.02
Yes	40 (82)	9 (53)	49 (74)	
No	9 (18)	8 (47)	17 (26)	
Paper list of community resources for healthy lifestyles				.29
Yes	33 (67)	9 (53)	42 (64)	
No	16 (33)	8 (47)	24 (36)	
Website compiling information on community resources for healthy lifestyles				.03
Yes	35 (71)	7 (41)	42 (64)	
No	14 (29)	10 (59)	24 (36)	
Clinic-based walking group or exercises classes				.78
Yes	24 (49)	9 (53)	33 (50)	
No	25 (51)	8 (47)	33 (50)	
Periodic physician visits				.98
Yes	20 (41)	7 (41)	27 (41)	
No	29 (59)	10 (59)	39 (59)	

Nonreferring providers more often agreed that their patients were likely to have minimal computer skills (76%, 13/17 vs 34%, 17/50), and were likely to lack Internet access (65%, 11/17 vs 32%, 16/50).

Most respondents reported that they found it useful to refer patients to a variety of community counseling resources for promoting healthy lifestyles (Table 3). In-person counseling visits with a health educator or coach, reminder phone calls from clinic staff, and online support from a health educator or coach were the most commonly endorsed resources that could be offered to patients. However, a preference for Web-based support—individualized counseling or an online compilation of community resources for healthy lifestyles—was more common among referring (vs nonreferring) providers.

Discussion

Although primary care physicians are aware of their potentially pivotal role in addressing the prevalence of obesity, weight counseling in the primary care setting remains uncommon. In this study, we found that a standard, EHR-based clinical referral coupled with consultant feedback processes can integrate online self-management support tools with primary care workflow in a manner that is acceptable to PCPs. Providers who referred patients to these resources were more likely to report that they had limited time for counseling compared with other providers. Conversely, a preference for personally delivering healthy lifestyle counseling was less common among referring PCPs, as was a perception that their patients lacked technical interest, skills, or Internet access. Physicians endorsed a range of

resources that could assist patients with weight loss, including online options.

Survey responses offered insight into several potential avenues for improving self-management support for obesity. For example, although physicians received feedback of the weight change of each participant, only 53% of the referring providers discussed this with patients and offered feedback regarding their efforts. Fostering more consistent feedback is potentially important because even minimal physician involvement may enhance outcomes of lifestyle modification interventions [13] and continuity of care in the primary care setting provides an opportunity for long-term lifestyle support. Furthermore, education about the obesity literature may be useful. For example, 20% of the providers in our study reported that obesity should be managed outside the clinical setting. Yet, physicians' direct discussion of patients' weight status has been associated with a significant weight loss and could be considered a targetable intervention [14]. In addition, 20% felt that most of their patients were generally not interested in receiving counseling on lifestyle modification, a perception that differs from published data on patients' preferences [15].

Most of the nonreferring providers were resident physicians (71%). Nonreferring providers more frequently preferred to counsel patients themselves rather than refer for counseling, and tended to report having sufficient time during clinic visits to counsel patients adequately on lifestyle modification. This could reflect the fact that resident physicians are allotted longer clinic visits, which provide them with more time to counsel patients. By choosing to counsel patients themselves rather than referring them to an intervention, residents may be more likely to learn the skills to counsel patients on lifestyle modification. Although these skills are essential, such practice might not be sustainable after the completion of their training.

Concerns about Internet access and minimal computer skills were raised by providers, particularly among those who did not refer patients. From our data, we cannot determine whether these responses reflect real or perceived barriers for patients. Broadband access in the US has expanded considerably in recent years, with more than 80% of the population having access to high-speed Internet access [16,17]. Web use among the minority populations has expanded considerably, particularly among African Americans, between 2000 and 2014, with 85% of white, 81% of Hispanic, and 78% of black non-Hispanic respondents

reporting Internet use in 2015 [18]. Furthermore, Internet use has increased disproportionately in populations that have historically shown below-average Internet use, including senior citizens, low-income Americans, and rural adults [16,19]. Yet, concerns over a digital divide remain [18,19], and the topic warrants attention whenever online tools are considered for patient care.

Because of the limited response rate (37%), these data may not reflect the full spectrum of PCP opinion regarding the referral model for weight-loss management. However, because 75% of referring doctors responded, their point of view should be adequately represented. In addition, physician response rates for surveys are often low, usually approximately 10 percentage points lower than that of the general population [20]. As noted previously, resident physicians were among the providers who completed the surveys, so study findings might not be applicable to other centers that lack trainees. The study is cross-sectional and providers were surveyed after the intervention. Thus, providers' attitudes might have changed due to the exposure to the intervention. Confounders could also be contributing to physicians' perceptions and views. For example, physicians' BMI has been associated with likelihood of physician initiating a weight-loss conversation [21]. Another limitation of the study is its lack of generalizability to populations without Internet access. In addition, it is not possible to compare attitudes across different practice settings (eg, rural vs urban). Nevertheless, this study has multiple strengths. To our knowledge, this is one of the first studies to evaluate physicians' views and compliance with a referral model for integrating online self-management support into primary care practice. In addition, this study involved multiple sites of care that represented a wide range of physicians and a diverse patient population.

Obesity has been recognized as one of the driving forces behind rising health care costs [22]. Primary care providers should be at the forefront of tackling the obesity epidemic. Referring patients to an intensive behavioral counseling intervention has been associated with clinically relevant improvements in health [23]. Online lifestyle interventions can provide a convenient and effective method of weight-loss management. Yet, only by understanding PCPs' views and barriers regarding the integration of online tools with routine preventive health practice will widespread implementation of evidence-based online tools be achieved.

Acknowledgments

The authors would like to thank the participants of the study. The study was funded by the Agency for Healthcare Research and Quality, R18 HS018155: Online Counseling to Enable Lifestyle Focused Obesity Treatment in Primary Care. Preliminary findings of this study were presented at the SGIM National Meeting in May 2014.

Conflicts of Interest

KM, CB, MBC, and GF cited royalties paid to the University of Pittsburgh, Division of General Internal Medicine, from Canary Health for work on the development of an online lifestyle program as a potential conflict of interest. No potential conflicts of interest were reported by RMH or RH.

References

1. Moyer VA, US Preventive Services Task Force. Screening for and management of obesity in adults: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012 Sep 04;157(5):373-378. [doi: [10.7326/0003-4819-157-5-201209040-00475](https://doi.org/10.7326/0003-4819-157-5-201209040-00475)] [Medline: [22733087](#)]
2. US Preventive Services Task Force. Screening for obesity in adults: recommendations and rationale. *Ann Intern Med* 2003 Dec 02;139(11):930-932. [Medline: [14644896](#)]
3. Manson JE, Skerrett PJ, Greenland P, VanItallie TB. The escalating pandemics of obesity and sedentary lifestyle. A call to action for clinicians. *Arch Intern Med* 2004 Feb 9;164(3):249-258. [doi: [10.1001/archinte.164.3.249](https://doi.org/10.1001/archinte.164.3.249)] [Medline: [14769621](#)]
4. Bleich SN, Pickett-Blakely O, Cooper LA. Physician practice patterns of obesity diagnosis and weight-related counseling. *Patient Educ Couns* 2011 Jan;82(1):123-129 [FREE Full text] [doi: [10.1016/j.pec.2010.02.018](https://doi.org/10.1016/j.pec.2010.02.018)] [Medline: [20303691](#)]
5. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *J Am Coll Cardiol* 2014 Jul 1;63(25 Pt B):2985-3023 [FREE Full text] [doi: [10.1016/j.jacc.2013.11.004](https://doi.org/10.1016/j.jacc.2013.11.004)] [Medline: [24239920](#)]
6. Lyden JR, Zickmund SL, Bhargava TD, Bryce CL, Conroy MB, Fischer GS, et al. Implementing health information technology in a patient-centered manner: patient experiences with an online evidence-based lifestyle intervention. *J Healthc Qual* 2013;35(5):47-57. [doi: [10.1111/jhq.12026](https://doi.org/10.1111/jhq.12026)] [Medline: [24004039](#)]
7. McTigue KM, Conroy MB. Use of the internet in the treatment of obesity and prevention of type 2 diabetes in primary care. *Proc Nutr Soc* 2013 Feb;72(1):98-108. [doi: [10.1017/S0029665112002777](https://doi.org/10.1017/S0029665112002777)] [Medline: [23098133](#)]
8. Hwang KO, Stuckey HL, Chen MC, Kraschnewski JL, Forjuoh SN, Poger JM, et al. Primary care providers' perspectives on online weight-loss programs: a big wish list. *J Med Internet Res* 2012 Jan 19;14(1):e16 [FREE Full text] [doi: [10.2196/jmir.1955](https://doi.org/10.2196/jmir.1955)] [Medline: [22262726](#)]
9. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. *Annu Rev Public Health* 2007;28:413-433. [doi: [10.1146/annurev.publhealth.28.021406.144145](https://doi.org/10.1146/annurev.publhealth.28.021406.144145)] [Medline: [17150029](#)]
10. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. *Telemed J E Health* 2007 Oct;13(5):509-517. [doi: [10.1089/tmj.2006.0089](https://doi.org/10.1089/tmj.2006.0089)] [Medline: [17999613](#)]
11. Binks M, O'Neil MP. Referral sources to a weight management program: relation to outcome. *J Gen Intern Med* 2002 Aug;17(8):596-603 [FREE Full text] [Medline: [12213140](#)]
12. Scanlan B, Tudorascu DL, Karpov I, Hess R, Fischer G, Simkin-Silverman LR, et al. Association of adherence measures with physical activity outcomes in an online weight loss trial: results from the OCELOT study. 2014 Apr 23 Presented at: 37th Annual Meeting of Society of General Internal Medicine; Apr 23, 2014; San Diego, CA.
13. Rollman BL, Fischer GS, Zhu F, Belnap BH. Comparison of electronic physician prompts versus waitroom case-finding on clinical trial enrollment. *J Gen Intern Med* 2008 Apr;23(4):447-450 [FREE Full text] [doi: [10.1007/s11606-007-0449-0](https://doi.org/10.1007/s11606-007-0449-0)] [Medline: [18373143](#)]
14. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA* 2014 Feb 26;311(8):806-814. [doi: [10.1001/jama.2014.732](https://doi.org/10.1001/jama.2014.732)] [Medline: [24570244](#)]
15. Potter MB, Vu JD, Croughan-Minihane M. Weight management: what patients want from their primary care physicians. *J Fam Pract* 2001 Jun;50(6):513-518. [Medline: [11401737](#)]
16. Pew Research Center. 2013. Three technology revolutions URL: <http://www.pewinternet.org/three-technology-revolutions/> [accessed 2017-07-31] [WebCite Cache ID 6sME8HbMc]
17. Broadband Availability in America: With Rural Americans Looking for High-Speed Services, Adequate Broadband Speeds Remain Out of Reach for Many. Washington, DC: Federal Communications Commission; 2015 Jan 30. URL: https://apps.fcc.gov/edocs_public/attachmatch/DOC-331734A1.pdf [WebCite Cache ID 6sMGB2IYW]
18. Kriel C. SiliconANGLE. 2015. FCC redefines broadband speed as 25 Mbps/3 Mbps, says Americans are underserved URL: <https://siliconangle.com/blog/2015/01/30/fcc-redefines-broadband-speed-as-25-mbps3-mbps-says-americans-are-underserved/> [WebCite Cache ID 6sMEhNvrf]
19. Perrin A, Duggan M. Pew Research Center. 2015 Jun 26. Americans' Internet access: 2000-2015 URL: <http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/> [WebCite Cache ID 6sMFB5Gk4]
20. Cummings SM, Savitz LA, Konrad TR. Reported response rates to mailed physician questionnaires. *Health Serv Res* 2001 Feb;35(6):1347-1355 [FREE Full text] [Medline: [11221823](#)]
21. Bleich SN, Bennett WL, Gudzone KA, Cooper LA. Impact of physician BMI on obesity care and beliefs. *Obesity (Silver Spring)* 2012 May;20(5):999-1005 [FREE Full text] [doi: [10.1038/oby.2011.402](https://doi.org/10.1038/oby.2011.402)] [Medline: [22262162](#)]
22. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer- and service-specific estimates. *Health Aff (Millwood)* 2009;28(5):w822-w831 [FREE Full text] [doi: [10.1377/hlthaff.28.5.w822](https://doi.org/10.1377/hlthaff.28.5.w822)] [Medline: [19635784](#)]
23. Wadden TA, Butryn ML, Hong PS, Tsai AG. Behavioral treatment of obesity in patients encountered in primary care settings: a systematic review. *JAMA* 2014 Nov 05;312(17):1779-1791 [FREE Full text] [doi: [10.1001/jama.2014.14173](https://doi.org/10.1001/jama.2014.14173)] [Medline: [25369490](#)]

Abbreviations

BMI: body mass index

EHR: electronic health record

PCP: primary care provider

Edited by G Eysenbach; submitted 01.08.17; peer-reviewed by K Antypas, M Amini, E Da Silva; comments to author 26.10.17; revised version received 20.12.17; accepted 06.01.18; published 08.06.18.

Please cite as:

Hanna RM, Fischer G, Conroy MB, Bryce C, Hess R, McTigue K

Online Lifestyle Modification Intervention: Survey of Primary Care Providers' Attitudes and Views

J Med Internet Res 2018;20(6):e167

URL: <http://www.jmir.org/2018/6/e167/>

doi: [10.2196/jmir.8616](https://doi.org/10.2196/jmir.8616)

PMID: [29884605](https://pubmed.ncbi.nlm.nih.gov/29884605/)

©Reem M Hanna, Gary Fischer, Molly B Conroy, Cindy Bryce, Rachel Hess, Kathleen McTigue. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.06.2018. 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.

Review

Internet-Based Interventions Aimed at Supporting Family Caregivers of People With Dementia: Systematic Review

Jenny Hopwood¹, MB BChir; Nina Walker¹, BSc (Hons), MSc; Lorraine McDonagh², BA, PhD; Greta Rait¹, MSc, MD, MRCP; Kate Walters¹, MSc, PhD, MRCP; Stephen Iliffe¹, BSc, MBBS, FRCGP; Jamie Ross³, BSc, PhD; Nathan Davies^{1,4}, BSc (Hons), MSc, PhD

¹Centre for Ageing Population Studies, Research Department of Primary Care and Population Health, University College London, London, United Kingdom

²Research Department of Primary Care and Population Health, University College London, London, United Kingdom

³eHealth Unit, Research Department of Primary Care and Population Health, University College London, London, United Kingdom

⁴Centre for Dementia Palliative Care Research, Marie Curie Palliative Care Research Department, University College London, London, United Kingdom

Corresponding Author:

Nathan Davies, BSc (Hons), MSc, PhD

Centre for Ageing Population Studies

Research Department of Primary Care and Population Health

University College London

Rowland Hill Street

Royal Free Hospital

London, NW3 2PF

United Kingdom

Phone: 44 2077940500 ext 34141

Email: n.m.davies@ucl.ac.uk

Abstract

Background: Caring for someone with dementia is one of the most challenging caring roles. The need for support for family caregivers has been recognized for some time but is often still lacking. With an aging population, demand on health and social care services is growing, and the population is increasingly looking to the internet for information and support.

Objective: In this review, we aimed to (1) identify the key components of existing internet-based interventions designed to support family caregivers of people with dementia, (2) develop an understanding of which components are most valued by caregivers, and (3) consider the evidence of effectiveness of internet-based interventions designed to support family caregivers of people with dementia.

Methods: We conducted a systematic search of online databases in April 2018. We searched reference lists and tracked citations. All study designs were included. We adopted a narrative synthesis approach with thematic analysis and tabulation as tools.

Results: We identified 2325 studies, of which we included 40. The interventions varied in the number and types of components, duration and dose, and outcomes used to measure effectiveness. The interventions focused on (1) contact with health or social care providers, (2) peer interaction, (3) provision of information, (4) decision support, and (5) psychological support. The overall quality of the studies was low, making interpretation and generalizability of the effectiveness findings difficult. However, most studies suggested that interventions may be beneficial to family caregiver well-being, including positive impacts on depression, anxiety, and burden. Particular benefit came from psychological support provided online, where several small randomized controlled trials suggested improvements in caregiver mental health. Provision of information online was most beneficial when tailored specifically for the individual and used as part of a multicomponent intervention. Peer support provided in online groups was appreciated by most participants and showed positive effects on stress. Finally, online contact with a professional was appreciated by caregivers, who valued easy access to personalized practical advice and emotional support, leading to a reduction in burden and strain.

Conclusions: Although mixed, the results indicate a positive response for the use of internet-based interventions by caregivers. More high-quality studies are required to identify the effectiveness of internet interventions aimed at supporting family caregivers, with particular focus on meeting the needs of caregivers during the different stages of dementia.

KEYWORDS

dementia; caregivers; internet; review

Introduction

Caring for someone with dementia can have a significant impact on the well-being of the caregiver. It is perceived as one of the most stressful and difficult forms of caring, as caregivers can face many years of managing difficult symptoms and making complex decisions [1,2]. Studies report higher levels of depression, emotional distress, and physical strain in caregivers of people with dementia than in caregivers for older adults with physical impairments [1,3].

There are around 670,000 family members and friends providing most care for people with dementia in the United Kingdom. Together, these caregivers are estimated to provide 1.3 billion hours of care per year and save the UK economy £12 billion annually [4]. Without the help of such caregivers, the formal care system would be likely to collapse [5].

With the given emotional and physical impact on caregiver well-being, psychological and practical support for caregivers is essential. There have been several trials of face-to-face interventions to support informal caregivers of people with dementia [6]. Reviews of interventions that provide information and advice have found varied results [7,8], but evidence of benefit has been found for some face-to-face psychological interventions in alleviating caregiver symptoms of depression [9]. However, uptake of such interventions is poor. It is estimated that around 10% of informal caregivers access caregiver support services [10], with the difficulty of leaving the care recipient and stigma being important barriers to uptake [11,12]. Individualizing caregiver interventions is also difficult economically, especially given the financial constraints in health care and the growing demand nationally and internationally due to the aging population [9].

Use of internet-based interventions may be an option to close the support gap for informal caregivers, particularly for those finding it difficult to leave their home or requiring flexibility due to caring responsibilities. Internet-based support interventions have the benefit of being relatively low cost and, by bringing the intervention into the home, may also have a role in reducing the social isolation that can come with caring [13,14]. Previous systematic reviews have suggested that internet-based interventions for informal caregivers of people with dementia have the capacity to improve various aspects of caregiver well-being, including depression, burden, and stress [15-17]. For psychological interventions in general, it is suggested that those with multiple components are better suited to support caregivers of people with dementia [9]. However, no previous reviews have identified what components might be important for interventions delivered via the internet for this group. Previous reviews have also focused predominantly on quantitative effectiveness data, which have been lacking in quality, and a mixed-methods review is important to provide

richer data on how caregivers use and find benefit from internet-based interventions.

This review aimed to (1) identify the key components of existing internet-based interventions designed to support family caregivers of people with dementia, (2) develop an understanding of which components are most valued by caregivers, and (3) consider the evidence of effectiveness of internet-based interventions designed to support family caregivers of people with dementia.

Technology and digital health interventions is a fast-paced research field, and therefore previous reviews are now outdated and require updating. Previous reviews have also focused on the effectiveness of whole interventions, where there are limited data to draw such strong conclusions, and in doing so have neglected a thorough and clear description of the content of interventions and their acceptability by caregivers.

Methods

Design

We conducted a systematic review of randomized controlled trials (RCTs), quasi-experimental designs (pre-post studies), quantitative studies, and qualitative studies, following the guidelines from the Centre for Reviews and Dissemination [18].

Inclusion and Exclusion Criteria

We included articles if they met the following criteria: (1) the intervention was aimed at informal caregivers (defined as a family member or friend providing unpaid care) of people with dementia, (2) the intervention was a digital intervention delivered via the internet, and (3) the article considered a specific intervention and provided a description of this.

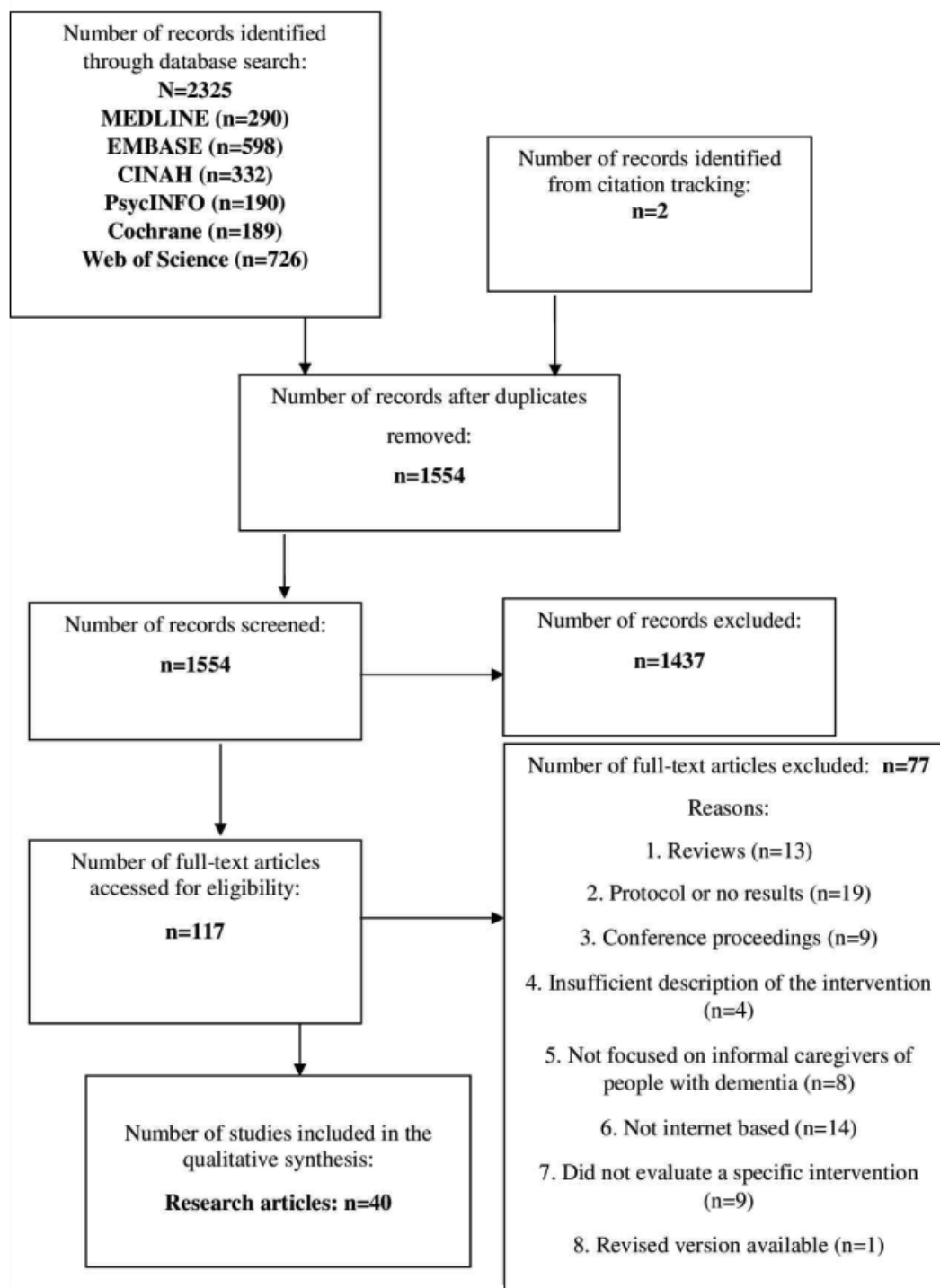
We excluded articles if (1) the intervention was telephone or telehealth based, (2) the interventions solely used Skype or another means of online calling, (3) the intervention had a large face-to-face component, (4) results or outcomes of the intervention were not reported, (5) the intervention was focused on the person with dementia, or (6) the study was not published in a peer-reviewed journal.

As our interest was in digital technologies that could be used by caregivers without input from health professionals, we excluded telephone-based support and those interventions with a large face-to-face component.

Search Strategy

We conducted a systematic literature search in CINAHL, the Cochrane Library, EMBASE, MEDLINE, PsycINFO, and Web of Science for articles published between January 1990 and April 2018. We selected 1990, as this was the period when the internet, including email, started to develop in commercial and public settings.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart describing the search process for articles on digital interventions for caregivers of people with dementia.



We tracked citations using Google Scholar, and hand searched reference lists for any additional relevant articles, in addition to hand searches of relevant journals. We identified literature reviews on the topic and checked them to ensure that our search identified relevant articles. Search terms and index terms (Medical Subject Headings) were identified from the initial scope of the literature. We added synonyms or abbreviations that we felt were appropriate to the search terms. [Multimedia Appendix 1](#) shows an example search strategy from MEDLINE. We contacted experts in the field. We also included gray

literature, including reviews of websites, in the scoping work but not in the review.

Selection Procedure

Article titles and abstracts were screened and excluded if they did not meet the inclusion criteria by 2 reviewers (JH and ND). We rapidly appraised non-English language articles, using their English abstracts, to ensure that we did not exclude any important articles. Articles considered relevant or where insufficient information was supplied in the abstract and title

were read in full by 2 reviewers (JH and ND). Two reviewers enhanced the validity and reliability of the selection procedure [18]. Any disagreement between reviewers or uncertainty about inclusion of articles would have been decided by a third reviewer, although this was not required. Figure 1 shows a Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart detailing the selection process.

Quality Appraisal

We appraised the literature for quality using the Critical Appraisal Skills Programme tools, using different tools for varying study design, qualitative studies [19], and RCTs [20], and an adapted version of the Critical Appraisal Skills Programme toolkit for quantitative designs [21]. We excluded no studies based on the results of their quality appraisal; rather, we used the study appraisal to develop discussion of the included studies.

Data Extraction and Synthesis

We developed a standardized data extraction tool to examine the included studies. We extracted data on study design, intervention details and components, duration of the intervention, participant characteristics, outcome measures, key findings, and the conclusions drawn by study authors. As the review included both quantitative and qualitative studies, and quantitative designs were heterogeneous, we could not pool quantitative data to conduct a meta-analysis. Therefore, we performed a narrative synthesis, using tabulation to organize the studies and a thematic analysis to categorize and group the studies. Two reviewers independently coded all studies (JH and ND). The 2 reviewers met to discuss each of their coding frames, discuss any disagreements, and develop a refined coding frame. Using the refined coding frame, all studies were coded by 1 reviewer (JH), and a selection of articles (50%) were blindly coded by a second reviewer (ND) and checked for agreement. Any disagreement was discussed and a third researcher would have been consulted if agreement could not be reached, although this was not required.

Results

Description of Studies

We included 40 articles [22-61] addressing 31 different interventions. There were 9 RCTs [22-30] (Multimedia Appendix 2), 7 quasi-experimental studies [31-37] (Multimedia Appendix 3), 4 qualitative studies [38-41] (Multimedia Appendix 4), and 20 studies with mixed or other methods [42-61] (Multimedia Appendix 5). All statistics in the multimedia appendices are reported as per the original articles.

All interventions aimed to address the needs of informal caregivers of people with dementia. One intervention also provided support for professional caregivers [26] and 2 provided support for people with dementia [41,49]. Some interventions limited their inclusion population to address specific caregiver needs, including 3 bilingual websites that addressed the needs of caregivers from minority ethnic groups [46,56,61]. Some restricted their intervention to informal caregivers experiencing stress [23,36,47], burden [24,29,57], depression [24,57], or anxiety [24,57].

Most interventions aimed to address the needs of caregivers providing care to people with all stages and types of dementia. One limited the intervention to spousal caregivers of people with mild cognitive impairment or “mild” dementia [44], 1 was limited to caregivers of people who were housebound with dementia [25], and 2 interventions also included people caring for those with other forms of neurodegenerative disease [36,39,40,52].

Although all interventions were primarily internet based, some had supplementary telephone-delivered components; for example, 1 provided a telephone number on their website for caregivers to contact a health care professional [32], and 1 included monthly telephone calls with caregivers [45].

A large number of different outcome measures were used in the studies. Outcomes included data on the usability of the interventions, as well as impacts on well-being, quality of life, burden, competence, physical health, and mental health. A wide range of validated and nonvalidated rating scales were used to assess the impact of the interventions. Qualitative results mainly focused on usability of the interventions and included observation of use [39-41,44,49], free-text surveys [38,41,49,51,55,61], interviews [38-41,46,47,47,53,54,59,60], and focus groups [42,45,56].

Quality Appraisal

The articles were of variable quality. Sample sizes ranged from 5 to 299, with many studies being pilot or feasibility studies. A problem with possible selection, performance, detection, or attrition bias was identified in many of the studies; many studies had small convenience samples, high attrition rates, and poor descriptions of randomization in trials and of data collection; and in some cases data collection was completed by nonblinded members of the study team.

Themes

We identified 5 themes as key components of the interventions: peer support; contact with a health or social care provider; provision of information; decision support; and psychological support.

Peer Support

Peer support was a key component of the interventions in 25 of the studies [22,25,26,29,31-33,36,38-42,44,45,47-50,52-56,58]. Peer support provided by fellow caregivers online was delivered either in private or in public, where all individuals using the intervention could see interactions. Common uses of peer support included supportive messages, information seeking, discussing the emotional impact of caring, and developing support networks outside of the Web-based intervention.

Private peer support was provided via private email or an online messenger service [22,36,39,40,45,48,49,52,53,55]. For example, the Digital Alzheimer Center allowed users to find others in their area caring for someone with the same diagnosis and then to communicate via private messaging [49]. However, in the few studies that quantified use of private messaging, use varied from very infrequent [49] to being one of the most-used functions [40].

One intervention, Inlife, provided the opportunity for the primary caregiver to develop their own networks of online support with friends, family, or significant others [48]. This allowed them to develop care books providing an overview of contact and practical information regarding the care of the individual, transfer care tasks among individuals, and provide help and assistance to one another.

Some interventions provided peer support in small groups [42]. For example, O'Connor and colleagues developed a virtual reality support group within an avatar environment [55] where groups of 3 to 4 anonymous caregivers communicated via online text. The groups were driven by the caregivers, allowing for exchange of ideas about communication, caring, and information about dementia, with some direction provided by a psychologist. Outcomes evaluated included loneliness, depression, burden, and perceived stress, but the study was underpowered to demonstrate any effects. A similar approach using videoconferencing software was used in another intervention where groups of caregivers met weekly online, initially with a facilitator, then as a peer group alone [39]. More than 90% of caregivers found this a positive experience, and there was a significant decline in stress in the experimental group. Use of the internet to deliver the intervention was felt to be as helpful as meeting people face-to-face by 61% of participants [52]. When this videoconference support group was compared with an internet-based chat group [53], both groups had a significant improvement in self-efficacy, but the video group showed a significantly greater improvement in mental health status. However, this was a pre-post study design with a duration of 6 months.

Public peer support usually consisted of forums [22,29,36,39,40,44,45,47,49,50,52-54,56,58] but also included chat rooms [32,41], shared blogs [25], links to peer groups on social networking sites [26], and video messages [31]. However, use of these tools was variable. Some studies reported that forums were not well used and were negatively reviewed by participants in qualitative reports [29,44]. In 1 study, this was thought to be due to the forum having an unclear purpose, the anonymity of participants, and a perceived high threshold for starting conversations [44]. In another study, 76% of participants visited the forum fewer than 12 times over the 12-week study [54]. However, some studies reported positive views, good rates of use, and a good impact on caregiver outcomes. For example, a 12-month RCT from Bass and colleagues analyzed the impact of the communication function and demonstrated a reduction in physical and emotional strain associated with use of the communication functions for caregivers who were initially under the most strain [22]. McKechnie and colleagues found a statistically significant improvement in the quality of the relationship with the care recipient but found no impact on depression or anxiety [54]. However, this was a smaller pre-post study with a short intervention period of only 12 weeks. Qualitative data suggested that participants found many benefits from peer interaction, including feeling understood through shared experience, finding reward in helping others, having reduced isolation, and being able to access information that would be difficult to find elsewhere [39,42,45,52,54]. However, in the 1 study where social isolation was measured using a

validated scale, peer interaction did not demonstrate a significant benefit [45].

Most interventions that provided private peer interaction also provided the option for public interaction [22,36,39,45,49,52,53]. When comparing private peer interaction with public peer interaction, Brennan and colleagues found that the public forum was used with increased frequency and duration compared with the private mail function [45], a contrast to findings from the Digital Alzheimer Center [49]. In qualitative feedback, participants found it difficult to recall the email addresses of others when using this private mail function so preferred to interact publicly.

The studies suggest that functions that have the potential for visual contact or group interaction may be more promising than simple chat-based functions in improving mental health status.

Contact With Professionals

Of the studies, 11 included components to allow caregivers to have direct contact with and ask questions of either a health or a social care professional [22,25,32,41,45,46,49,56,58-60]. Professionals included nurses [22,45], occupational therapists [46], or social workers [27,46,58]. In some interventions, the role of the professional was not clear; rather, the caregiver was described as having contact with a "medical professional" or "expert" [32,41,49,56,59,60], or a multidisciplinary team [25].

Most interventions required caregivers to contact health professionals themselves [22,25,32,45,46,49,56,58-60]. For example, the eHealthMonitor dementia portal [59,60] provided alerts for health professionals when caregivers entered a question; professionals could then respond online or arrange an appointment via telephone. Only 1 intervention adopted a proactive approach where health professionals contacted caregivers who self-assessed as having severe stress [27]. The intervention as a whole led to a significant decrease in hardship and grief compared with the control group, but there was no significant change in burden, depressive symptoms, or desire for nursing home placement.

On the whole, evaluation data from the studies showed that interaction with professionals was a positive experience for caregivers [45,46,58-60]. Professionals provided personalized practical advice for caregivers at home on caring and dementia, as well as emotional support, and caregivers reported feeling less isolated as a result. However, opinions about seeking this support electronically did vary [46,56], with some caregivers enjoying writing emails, while others felt confused about how much information to include.

Provision of Information

Most interventions provided information for caregivers about dementia, practical aspects of caregiving, or available local and national services. For some, this was the only function of the intervention [30,37,51,61], but for most information provision was part of a multicomponent intervention [22-29,32-36,38-40,42-50,52,53,56-60]. Some RCTs of multicomponent interventions that included the provision of information did demonstrate positive impacts on depression [24], anxiety [24], perceived stress [28,52], and attitudes toward dementia [26].

However, as information was part of a broader intervention, it was difficult to know the impact of this component. One intervention that was analyzed in an RCT that attempted to assess this was ComputerLink [22], which provided information on dementia, caregiving, and local services as part of a multicomponent intervention. Use of the information provision parts of the intervention was associated with reduced strain for caregivers living alone with care recipients and for spousal caregivers. However, other multiple-component interventions evaluated with qualitative methods found that caregivers found other components, such as interaction with professionals, more beneficial than information [40,46], with caregivers expressing frustration when required to review information that did not meet their specific needs [38,56]. When information was individualized, it was considered by caregivers as one of the most useful functionalities of the intervention [37,59]. This suggests that information does appear to be an important part of interventions, but the information should be tailored to the individual caregiver situation and not be the sole focus of the intervention.

Decision-Making Support

Some of the interventions recognized that decision making is a difficult process for caregivers and included decision aids [22,36,41,45,59,60]. However, most studies did not explain in detail how the intervention provided support with decision making; for example, Lorig and colleagues included decision-making assistance in their online workshops and chat forums [36] but lacked further description of how this was achieved. The only well-described decision aid intervention was ComputerLink [45], which included a tool based on multiattribute utility theory [62], where caregivers were led through a series of questions prioritizing important factors in the decision-making process. Use of the decision-making tool significantly improved caregiver confidence prior to having face-to-face discussions when compared with the control group. However, in some studies the decision-making tools were poorly used [41,45] and not appreciated by caregivers [41]. Instead, participants gained decision support from other components of the interventions, such as discussion with peers or professionals [41,45].

Psychological Support

Many interventions included components of psychological support [23,24,26-29,32-36,39,44,46,47,55], which were self-guided or professionally guided. Few used standardized forms of psychological interventions or therapy, but therapeutic relaxation techniques were commonly used.

Self-guided psychological support most often consisted of modules that caregivers worked through, and several were tested in RCTs. For example, Beauchamp and colleagues delivered a modular intervention that provided videos on cognitive and behavioral strategies to cope with difficult emotions [23]. In an RCT of the intervention, the experimental group had significantly greater improvements in stress, self-efficacy, intention to get support, strain, gain, depression, and anxiety. Similar results were found in other RCTs of similar psychological interventions, with reductions found in caregiver stress in 1 intervention [28] and improvements in attitudes

toward dementia, distress, empathy, and perspective in another [26]. However, the durations of these RCTs were short, ranging from 1 to 4 months.

Some interventions provided self-directed modules to work through, but caregivers were supported by a professional coach, who was most often a psychologist [24,33,38,44,57]. Caregivers were required to complete assignments, homework, reflective diaries, or regular assessments of their well-being. An RCT of 1 such intervention showed a reduction in symptoms of anxiety and depression with moderate and small effect sizes, respectively [24].

Some studies provided professionally delivered psychological therapies online, either via individual interaction with a therapist using email [35,46] or online interaction with a small group of caregivers [32,36,39,40,52,53,55]. In the ADCarer.com intervention [35], the professional (a psychologist, social worker, or counsellor) would respond to online messages from the caregiver within 48 hours using cognitive behavioral therapy techniques. In a pre-post assessment, the multicomponent intervention did lead to a significant reduction in caregiver distress. Interactive groups were delivered either using videoconferencing software [32,39,40,52,53] or an avatar-based format [55] and allowed small groups of caregivers to interact, guided by a professional. Improvements were found in caregiver mental health and quality-of-life outcomes, but with these interventions as with many others, it is difficult to tease out the specifics of components, as in both cases the virtual support group offered peer support as well as psychological support.

Overall, studies assessing psychological support suggested a positive effect on a variety of factors, including improving caregiver distress, depression, anxiety, and strain. However, some stressed the importance of cultural appropriateness. Kajiyama and colleagues used the popularity of Spanish-language telenovela (a type of television serial drama or soap opera produced mainly in Latin America) to appeal to Hispanic and Latino family caregivers [34].

Discussion

Principal Findings

Unlike previous reviews in this area, this review explored the key components of internet-based interventions to support family caregivers of people with dementia. We identified a broad variety of interventions, which focused on providing peer support, engaging with health and social care professionals, and providing information, decision support, and psychological support. Although effectiveness was not a focus of this review, some multiple-component interventions showed promise in reducing stress, anxiety, and depressive symptoms for family caregivers and in increasing self-efficacy [44,57]. However, as with previous reviews [15-17], the limited number of high-quality RCTs, as well as the multiple-component nature of many interventions, makes it difficult to report which aspects of the interventions were effective.

Peer support was a key component of many of the interventions discussed. Caring for someone with dementia has often been described as not only a lonely role but also one in which there

is a great deal of uncertainty. The peer support components of the interventions identified in this review aimed to target these feelings and were described positively by many participants, but no significant effect for peer interaction and social isolation was found [45]. However, qualitative data in this review suggest that peer support offered a form of socialization. Previous evidence is mixed on whether use of the internet reduces or enhances loneliness [63,64], but this review suggests that internet-based peer interaction may have a benefit for family caregivers. However, it is evident that the way that peer support is delivered is important, with opportunities for group interaction or videoconferencing being more beneficial than public-facing forums and private messaging functions.

The qualitative data suggest that interactions with health professionals are viewed positively; however, it is unclear whether this positivity was linked to the provision of contact online or whether caregivers may prefer this interaction face-to-face. The mix of professionals providing support in the studies suggests there is a lack of consensus on who is best to deliver professional support. This may reflect ambiguity caregivers feel about who is the most appropriate person to talk to when they need advice.

The provision of information was often at the core of interventions, and this supports findings from previous research where most caregivers preferred to receive information online rather than in paper format [65].

This review demonstrated that interventions that focused solely on decision making were, in general, not favored by family caregivers. However, decision-making tools were viewed more positively when they were used alongside other components, such as peer support. Decision making is often left to family caregivers when the person with dementia no longer has capacity, making this a difficult and challenging time for family caregivers. However, results from this review suggest that face-to-face meetings may be required to make decisions, and internet resources are only used as a method of preparation for discussions. This adds to our understanding of barriers to making decisions, which include a lack of information, poor communication, difficult dynamics and conflict within families, and limited emotional and practical support [66-70].

Interventions including online psychological support showed some of the most promising findings, with individual studies reporting significant reductions in caregiver stress, strain, depressive symptoms, and anxiety, in addition to increases in self-efficacy [23,26,28]. Although studies of both professionally guided and self-guided interventions indicated a positive outcome for participants, including caregiver mental health outcomes, they were quasi-experimental (pre-post studies), feasibility studies, and small RCTs, suggesting these conclusions should be made with caution.

Qualitative evaluations of the interventions demonstrated positive views from most caregivers toward internet-based support interventions, although it is clear that not all would benefit from such interventions. It may be that the internet is most beneficial for those who are classified as most vulnerable (ie, more stressed) [22].

Implications for the Development of Future Internet-Based Interventions for Caregivers

In developing an internet intervention for family caregivers, several issues need to be addressed. Questions of privacy and security were highlighted [59,60], reflected in the contrast of public versus private messaging approaches and password-protected websites. The details discussed by many on the websites are very personal and emotional topics. Sillence and colleagues discussed a series of factors that influence the mistrust and trust of health websites [71]. The design of the site contributed to most of the reasons for rejecting and mistrusting a website, including complex and busy layout, corporate look, and irrelevant content. However, the reasons for selecting and trusting a website were more focused on the content of the website, including unbiased information and personalized content.

Another issue is complexity. Some caregivers found functions such as private messaging, decision aids, and login screens complex, which affected their use. Using familiar-sounding language [71] and a strong iterative approach, in which the intervention undergoes multiple cycles of development and optimization [72], with future interventions are two ways to help overcome this challenge. Tailoring can reduce the quantity of information and resources caregivers must review, and caregivers may be more motivated to use an intervention they feel is applicable to their circumstances. This review found that where interventions were not personalized, caregivers found this frustrating and their needs were not met [47]. Finally, there is the question of internet literacy and access to the internet: the digital divide [73]. There appears to still be a gap between those who use or can use the internet and those who don't, with a study in 2015 highlighting that almost all adults over 70 years of age had difficulty using the intervention [41]. Many of the studies included in this review consisted of participants who were predominantly younger caregivers, whereas many people caring for someone with dementia are more likely to be older. Reducing the complexity of interventions, supporting access with potential support from health professionals, and highlighting the benefits of such interventions to understand their potential value may aid in bridging the divide. For most of the studies, this digital divide was ignored, as a requirement for participation was computer literacy [61], and observational studies assessed the usability of the interventions with caregivers who had already received training in using the website.

Implications for Policy, Clinical Practice, and Further Research

This review demonstrated the need for high-quality research to evaluate the effectiveness of internet-based interventions for caregivers of people with dementia, in particular larger phase 3 trials. Importantly within these studies, it would be useful to describe the interventions in more detail and to understand which aspects of the interventions are used more than others and provide the most benefit. Future research should also focus on which aspects of the interventions are most beneficial for different groups—for example, adult children compared with spouses—and how the interventions can best be delivered to address issues such as the digital divide. Future research would

also benefit from including theoretical considerations of how interventions are thought to provide support to caregivers.

This review identified a gap in the development of interventions targeting specific stages of the dementia trajectory. Many of the interventions in this review were broad and generic to the entire dementia trajectory. However, the needs of family caregivers vary at different stages of the disease and transition points; for example, around the end of life of the person with dementia, caregivers face specific challenges around decision making and management of difficult symptoms. Future interventions and research should address these different stages when developing digital interventions to support family caregivers [49].

Strengths and Limitations

Similarly to previous reviews in this area [15-17], comparison between studies was difficult, as the interventions used were complex and varied, with wide-ranging study designs and outcome measures. The review was also limited by the quality of some of the studies and the methods employed. There were relatively few RCTs from which to derive effectiveness data. Many of the studies were feasibility and pilot studies, so we were unable to draw definitive conclusions surrounding effectiveness and acceptability. For many of the studies, there were high levels of dropouts and for some interventions participants made limited use of some of the components of the interventions, therefore making it difficult to draw conclusions [46,49]. Few studies provided information on the effectiveness of individual components of the interventions, and some studies explored only usefulness and usability with reference to the

design and layout of the interventions, which on the whole were not well described. This is helpful only to an extent because, to develop or build on existing interventions, there needs to be an understanding of which elements have a positive effect on family caregivers and so should be included in new interventions.

Our literature search was limited by including only peer-reviewed publications, and there may have been several other interventions that were being practically used and applied but not published via academic routes. However, the search of the academic literature was thorough and we used a rigorous search strategy, updated before publication.

This review has built on previous literature by identifying the core components of interventions for family caregivers, which will be useful for future intervention development. As our inclusion criteria were much more comprehensive, this review provides a larger evidence base than previous reviews. Unlike previous reviews, we have particularly considered how caregivers are supported with decision making through Web-based interventions and we included data from many qualitative studies, providing richer information on how the interventions were perceived and valued by caregivers.

Conclusions

The evidence base for internet-based interventions for caregivers of people with dementia remains limited. Although this review recognizes that for some caregivers, a face-to-face intervention may be preferred, our findings highlight the promising potential of digital interventions to support caregivers, which warrants further development and testing.

Acknowledgments

This work was supported by the UK National Institute for Health Research School for Primary Care Research (NIHR SPCR) FR11 and Capacity Award 6 programs. This paper presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE search terms and strategy.

[PDF File (Adobe PDF File), 24KB - [jmir_v20i6e216_app1.pdf](#)]

Multimedia Appendix 2

Characteristics, components, outcomes, and key findings of randomized controlled trials.

[PDF File (Adobe PDF File), 47KB - [jmir_v20i6e216_app2.pdf](#)]

Multimedia Appendix 3

Characteristics, components, outcomes, and key findings of quasi-experimental (pre-post) studies.

[PDF File (Adobe PDF File), 42KB - [jmir_v20i6e216_app3.pdf](#)]

Multimedia Appendix 4

Characteristics, components, outcomes, and key findings of qualitative studies.

[PDF File (Adobe PDF File), 27KB - [jmir_v20i6e216_app4.pdf](#)]

Multimedia Appendix 5

Characteristics, components, outcomes, and key findings of mixed-methods and other methods studies.

[PDF File (Adobe PDF File), 65KB - [jmir_v20i6e216_app5.pdf](#)]

References

1. Cuijpers P. Depressive disorders in caregivers of dementia patients: a systematic review. *Aging Ment Health* 2005 Jul;9(4):325-330. [doi: [10.1080/13607860500090078](#)] [Medline: [16019288](#)]
2. Kneebone II, Martin PR. Coping and caregivers of people with dementia. *Br J Health Psychol* 2003 Feb;8(Pt 1):1-17. [doi: [10.1348/135910703762879174](#)] [Medline: [12643813](#)]
3. Ory MG, Hoffman RR, Yee JL, Tennstedt S, Schulz R. Prevalence and impact of caregiving: a detailed comparison between dementia and nondementia caregivers. *Gerontologist* 1999 Apr;39(2):177-185. [Medline: [10224714](#)]
4. Alzheimer's Society. Dementia UK update. London, UK: Alzheimer's Society; 2014 Nov. URL: https://www.alzheimers.org.uk/sites/default/files/migrate/downloads/dementia_uk_update.pdf [accessed 2018-05-24] [WebCite Cache ID 6zel2XXKV]
5. Alzheimer's Society. Carer support. 2013. URL: <https://www.alzheimers.org.uk/about-us/policy-and-influencing/what-we-think/carers-support> [accessed 2018-02-14] [WebCite Cache ID 6xE0DR7ht]
6. Livingston G, Barber J, Rapaport P, Knapp M, Griffin M, King D, et al. Clinical effectiveness of a manual based coping strategy programme (START, STRategies for RelaTives) in promoting the mental health of carers of family members with dementia: pragmatic randomised controlled trial. *BMJ* 2013 Oct 25;347:f6276 [FREE Full text] [Medline: [24162942](#)]
7. Corbett A, Stevens J, Aarsland D, Day S, Moniz-Cook E, Woods R, et al. Systematic review of services providing information and/or advice to people with dementia and/or their caregivers. *Int J Geriatr Psychiatry* 2012 Jun;27(6):628-636. [doi: [10.1002/gps.2762](#)] [Medline: [22038644](#)]
8. Thompson CA, Spilsbury K, Hall J, Birks Y, Barnes C, Adamson J. Systematic review of information and support interventions for caregivers of people with dementia. *BMC Geriatr* 2007;7:18 [FREE Full text] [doi: [10.1186/1471-2318-7-18](#)] [Medline: [17662119](#)]
9. Selwood A, Johnston K, Katona C, Lyketsos C, Livingston G. Systematic review of the effect of psychological interventions on family caregivers of people with dementia. *J Affect Disord* 2007 Aug;101(1-3):75-89. [doi: [10.1016/j.jad.2006.10.025](#)] [Medline: [17173977](#)]
10. Beesley L. Informal care in England. London, UK: The King's Fund; 2006. URL: <https://www.kingsfund.org.uk/sites/default/files/informal-care-england-wanless-background-paper-lucinda-beesley2006.pdf> [accessed 2018-05-29] [WebCite Cache ID 6zelDHWCU]
11. Brodaty H, Thomson C, Thompson C, Fine M. Why caregivers of people with dementia and memory loss don't use services. *Int J Geriatr Psychiatry* 2005 Jun;20(6):537-546. [doi: [10.1002/gps.1322](#)] [Medline: [15920707](#)]
12. Serafini JD, Damianakis T, Marziali E. Clinical practice standards and ethical issues applied to a virtual group intervention for spousal caregivers of people with Alzheimer's. *Soc Work Health Care* 2007;44(3):225-243. [doi: [10.1300/J010v44n03_07](#)] [Medline: [17548277](#)]
13. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. *Cyberpsychol Behav Soc Netw* 2015 May;18(5):268-272. [doi: [10.1089/cyber.2014.0549](#)] [Medline: [25919967](#)]
14. Milligan C, Passey D. Ageing and the Use of the Internet - Current Engagement and Future Needs: State-of-the-Art Report. Lancaster, UK: University of Lancaster; 2011.
15. Boots LMM, de Vugt ME, van Knippenberg RJM, Kempen GIJM, Verhey FRJ. A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. *Int J Geriatr Psychiatry* 2014 Apr;29(4):331-344. [doi: [10.1002/gps.4016](#)] [Medline: [23963684](#)]
16. McKechnie V, Barker C, Stott J. Effectiveness of computer-mediated interventions for informal carers of people with dementia-a systematic review. *Int Psychogeriatr* 2014 Oct;26(10):1619-1637 [FREE Full text] [doi: [10.1017/S1041610214001045](#)] [Medline: [24989249](#)]
17. Powell J, Chiu T, Eysenbach G. A systematic review of networked technologies supporting carers of people with dementia. *J Telemed Telecare* 2008;14(3):154-156. [doi: [10.1258/jtt.2008.003018](#)] [Medline: [18430288](#)]
18. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York, UK: Centre for Reviews and Dissemination, University of York; 2009. URL: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf [accessed 2018-05-24] [WebCite Cache ID 6zelYyI2G]
19. Critical Appraisal Skills Programme. CASP qualitative checklist. 2017. URL: http://docs.wixstatic.com/ugd/dded87_25658615020e427da194a325e7773d42.pdf [accessed 2018-02-14] [WebCite Cache ID 6xDz3wBsA]

20. Critical Appraisal Skills Programme. CASP Randomised Controlled Trial Checklist. 2017. URL: http://docs.wixstatic.com/ugd/dded87_4239299b39f647ca9961f30510f52920.pdf [accessed 2018-02-14] [WebCite Cache ID 6xDz89ZeG]
21. Critical Appraisal Skills Programme. Critical Appraisal Skills Programme 2017. 2017. URL: <http://www.casp-uk.net/casp-tools-checklists> [accessed 2018-03-01] [WebCite Cache ID 6xb7GYWHW]
22. Bass DM, McClendon MJ, Brennan PF, McCarthy C. The buffering effect of a computer support network on caregiver strain. *J Aging Health* 1998 Feb;10(1):20-43. [doi: [10.1177/089826439801000102](https://doi.org/10.1177/089826439801000102)] [Medline: [10182416](https://pubmed.ncbi.nlm.nih.gov/10182416/)]
23. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. *Gerontologist* 2005 Dec;45(6):793-801. [Medline: [16326661](https://pubmed.ncbi.nlm.nih.gov/16326661/)]
24. Blom MM, Zarit SH, Groot ZRBM, Cuijpers P, Pot AM. Effectiveness of an Internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. *PLoS One* 2015;10(2):e0116622 [FREE Full text] [doi: [10.1371/journal.pone.0116622](https://doi.org/10.1371/journal.pone.0116622)] [Medline: [25679228](https://pubmed.ncbi.nlm.nih.gov/25679228/)]
25. Fowler CN, Kott K, Wicks MN, Rutledge C. An interprofessional virtual healthcare neighborhood: effect on self-efficacy and sleep among caregivers of older adults with dementia. *J Gerontol Nurs* 2016 Nov 01;42(11):39-47. [doi: [10.3928/00989134-20160901-02](https://doi.org/10.3928/00989134-20160901-02)] [Medline: [27598270](https://pubmed.ncbi.nlm.nih.gov/27598270/)]
26. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J, et al. Web-based STAR e-learning course increases empathy and understanding in dementia caregivers: results from a randomized controlled trial in the Netherlands and the United Kingdom. *J Med Internet Res* 2015;17(10):e241 [FREE Full text] [doi: [10.2196/jmir.4025](https://doi.org/10.2196/jmir.4025)] [Medline: [26519106](https://pubmed.ncbi.nlm.nih.gov/26519106/)]
27. Hicken BL, Daniel C, Luptak M, Grant M, Kilian S, Rupper RW. Supporting Caregivers of Rural Veterans Electronically (SCORE). *J Rural Health* 2017 Jun;33(3):305-313. [doi: [10.1111/jrh.12195](https://doi.org/10.1111/jrh.12195)] [Medline: [27437642](https://pubmed.ncbi.nlm.nih.gov/27437642/)]
28. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian TY, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. *Aging Ment Health* 2013;17(5):544-554 [FREE Full text] [doi: [10.1080/13607863.2013.775641](https://doi.org/10.1080/13607863.2013.775641)] [Medline: [23461355](https://pubmed.ncbi.nlm.nih.gov/23461355/)]
29. Núñez-Naveira L, Alonso-Búa B, de Labra C, Gregersen R, Maibom K, Mojs E, et al. UnderstAID, an ICT platform to help informal caregivers of people with dementia: a pilot randomized controlled study. *Biomed Res Int* 2016;2016:5726465 [FREE Full text] [doi: [10.1155/2016/5726465](https://doi.org/10.1155/2016/5726465)] [Medline: [28116300](https://pubmed.ncbi.nlm.nih.gov/28116300/)]
30. Van Mierlo LD, Meiland FJM, Van de Ven PM, Van Hout HPJ, Dröes R. Evaluation of DEM-DISC, customized e-advice on health and social support services for informal carers and case managers of people with dementia; a cluster randomized trial. *Int Psychogeriatr* 2015 Aug;27(8):1365-1378. [doi: [10.1017/S1041610215000423](https://doi.org/10.1017/S1041610215000423)] [Medline: [25872457](https://pubmed.ncbi.nlm.nih.gov/25872457/)]
31. Davis B, Shehab M, Shenk D, Nies M. E-mobile pilot for community-based dementia caregivers identifies desire for security. *Gerontechnology* 2015;13(3):332-336. [doi: [10.4017/gt.2015.13.3.003.00](https://doi.org/10.4017/gt.2015.13.3.003.00)]
32. Glueckauf RL, Ketterson TU, Loomis JS, Dages P. Online support and education for dementia caregivers: overview, utilization, and initial program evaluation. *Telemed J E Health* 2004;10(2):223-232. [Medline: [15319052](https://pubmed.ncbi.nlm.nih.gov/15319052/)]
33. Griffiths PC, Kovaleva M, Higgins M, Langston AH, Hepburn K. Tele-Savvy: an online program for dementia caregivers. *Am J Alzheimers Dis Other Dement* 2018 Jan 01;1533317518755331. [doi: [10.1177/1533317518755331](https://doi.org/10.1177/1533317518755331)] [Medline: [29544342](https://pubmed.ncbi.nlm.nih.gov/29544342/)]
34. Kajiyama B, Fernandez G, Carter EA, Humber MB, Thompson LW. Helping Hispanic dementia caregivers cope with stress using technology-based resources. *Clin Gerontol* 2018;41(3):209-216. [doi: [10.1080/07317115.2017.1377797](https://doi.org/10.1080/07317115.2017.1377797)] [Medline: [29236621](https://pubmed.ncbi.nlm.nih.gov/29236621/)]
35. Kwok T, Au A, Wong B, Ip I, Mak V, Ho F. Effectiveness of online cognitive behavioral therapy on family caregivers of people with dementia. *Clin Interv Aging* 2014;9:631-636 [FREE Full text] [doi: [10.2147/CIA.S56337](https://doi.org/10.2147/CIA.S56337)] [Medline: [24748781](https://pubmed.ncbi.nlm.nih.gov/24748781/)]
36. Lorig K, Thompson-Gallagher D, Traylor L, Ritter P, Laurent D, Plant K. Building better caregivers: a pilot online support workshop for family caregivers of cognitively impaired adults. *J Appl Gerontol* 2013;31(3):423-437. [doi: [10.1177/0733464810389806](https://doi.org/10.1177/0733464810389806)]
37. van der Roest HG, Meiland FJM, Jonker C, Dröes R. User evaluation of the DEMentia-specific Digital Interactive Social Chart (DEM-DISC). A pilot study among informal carers on its impact, user friendliness and, usefulness. *Aging Ment Health* 2010 May;14(4):461-470. [doi: [10.1080/13607860903311741](https://doi.org/10.1080/13607860903311741)] [Medline: [20455122](https://pubmed.ncbi.nlm.nih.gov/20455122/)]
38. Kovaleva M, Blevins L, Griffiths PC, Hepburn K. An online program for caregivers of persons living with dementia: lessons learned. *J Appl Gerontol* 2017 Apr 01;733464817705958. [doi: [10.1177/0733464817705958](https://doi.org/10.1177/0733464817705958)] [Medline: [28460557](https://pubmed.ncbi.nlm.nih.gov/28460557/)]
39. Marziali E, Donahue P, Crossin G. Caring for others: internet health care support intervention for family caregivers of persons with Alzheimer's, stroke, or Parkinson's disease. *Fam Soc* 2018 May 03;86(3):375-383. [doi: [10.1606/1044-3894.3435](https://doi.org/10.1606/1044-3894.3435)]
40. Marziali E, Damianakis T, Donahue P. Internet-based clinical services: virtual support groups for family caregivers. *J Technol Hum Serv* 2006 Jul 18;24(2-3):39-54. [doi: [10.1300/J017v24n02_03](https://doi.org/10.1300/J017v24n02_03)]
41. Span M, Smits C, Jukema J, Groen-van de Ven L, Janssen R, Vernooij-Dassen M, et al. An interactive web tool for facilitating shared decision-making in dementia-care networks: a field study. *Front Aging Neurosci* 2015;7:128 [FREE Full text] [doi: [10.3389/fnagi.2015.00128](https://doi.org/10.3389/fnagi.2015.00128)] [Medline: [26217221](https://pubmed.ncbi.nlm.nih.gov/26217221/)]
42. Austrom MG, Geros KN, Hemmerlein K, McGuire SM, Gao S, Brown SA, et al. Use of a multiparty web based videoconference support group for family caregivers: innovative practice. *Dementia (London)* 2015 Sep;14(5):682-690 [FREE Full text] [doi: [10.1177/1471301214544338](https://doi.org/10.1177/1471301214544338)] [Medline: [25062788](https://pubmed.ncbi.nlm.nih.gov/25062788/)]

43. Boots LM, de Vugt ME, Smeets CM, Kempen GI, Verhey FR. Implementation of the blended care self-management program for caregivers of people with early-stage dementia (Partner in Balance): process evaluation of a randomized controlled trial. *J Med Internet Res* 2017 Dec 19;19(12):e423 [FREE Full text] [doi: [10.2196/jmir.7666](https://doi.org/10.2196/jmir.7666)] [Medline: [29258980](#)]
44. Boots LM, de Vugt ME, Withagen HE, Kempen GI, Verhey FR. Development and initial evaluation of the internet-based self-management intervention 'Partner in Balance' for spousal caregivers of people with early-stage dementia. *JMIR Res Protoc* 2016 Mar 01;5(1):e33 [FREE Full text] [doi: [10.2196/resprot.5142](https://doi.org/10.2196/resprot.5142)] [Medline: [26932438](#)]
45. Brennan PF, Moore SM, Smyth KA. The effects of a special computer network on caregivers of persons with Alzheimer's disease. *Nurs Res* 1995;44(3):166-172. [Medline: [7761293](#)]
46. Chiu T, Marziali E, Colantonio A, Carswell A, Gruneir M, Tang M, et al. Internet-based caregiver support for Chinese Canadians taking care of a family member with alzheimer disease and related dementia. *Can J Aging* 2009 Dec;28(4):323-336. [doi: [10.1017/S0714980809990158](https://doi.org/10.1017/S0714980809990158)] [Medline: [19925698](#)]
47. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. *J Med Internet Res* 2015;17(5):e117 [FREE Full text] [doi: [10.2196/jmir.3717](https://doi.org/10.2196/jmir.3717)] [Medline: [25967983](#)]
48. Dam AEH, van Boxtel MPJ, Rozendaal N, Verhey FRJ, de Vugt ME. Development and feasibility of Inlife: a pilot study of an online social support intervention for informal caregivers of people with dementia. *PLoS One* 2017;12(9):e0183386 [FREE Full text] [doi: [10.1371/journal.pone.0183386](https://doi.org/10.1371/journal.pone.0183386)] [Medline: [28886056](#)]
49. Hattink B, Dros 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](https://doi.org/10.2196/resprot.5040)] [Medline: [27444209](#)]
50. Jajor J, Rosolek M, Skorupska E, Krawczyk-Wasielewska A, Lisiński P, Mojs E. UnderstAID - a platform that helps informal caregivers to understand and aid their demented relatives - assessment of informal caregivers - a pilot study. *J Med Sci* 2016;84(4):229-234 [FREE Full text]
51. Lewis ML, Hobday JV, Hepburn KW. Internet-based program for dementia caregivers. *Am J Alzheimers Dis Other Demen* 2010 Dec;25(8):674-679. [doi: [10.1177/1533317510385812](https://doi.org/10.1177/1533317510385812)] [Medline: [21131674](#)]
52. Marziali E, Donahue P. Caring for others: internet video-conferencing group intervention for family caregivers of older adults with neurodegenerative disease. *Gerontologist* 2006 Jun;46(3):398-403 [FREE Full text] [Medline: [16731880](#)]
53. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 Internet-based intervention programs. *Am J Alzheimers Dis Other Demen* 2011 Feb;26(1):36-43. [doi: [10.1177/1533317510387586](https://doi.org/10.1177/1533317510387586)] [Medline: [21282276](#)]
54. McKechnie V, Barker C, Stott J. The effectiveness of an Internet support forum for carers of people with dementia: a pre-post cohort study. *J Med Internet Res* 2014;16(2):e68 [FREE Full text] [doi: [10.2196/jmir.3166](https://doi.org/10.2196/jmir.3166)] [Medline: [24583789](#)]
55. O'Connor M, Arizmendi BJ, Kaszniak AW. Virtually supportive: a feasibility pilot study of an online support group for dementia caregivers in a 3D virtual environment. *J Aging Stud* 2014 Aug;30:87-93 [FREE Full text] [doi: [10.1016/j.jaging.2014.03.001](https://doi.org/10.1016/j.jaging.2014.03.001)] [Medline: [24984911](#)]
56. Pagán-Ortiz ME, Cortés DE, Rudloff N, Weitzman P, Levkoff S. Use of an online community to provide support to caregivers of people with dementia. *J Gerontol Soc Work* 2014;57(6-7):694-709 [FREE Full text] [doi: [10.1080/01634372.2014.901998](https://doi.org/10.1080/01634372.2014.901998)] [Medline: [24689359](#)]
57. Pot AM, Blom MM, Willemse BM. Acceptability of a guided self-help Internet intervention for family caregivers: mastery over dementia. *Int Psychogeriatr* 2015 Aug;27(8):1343-1354. [doi: [10.1017/S1041610215000034](https://doi.org/10.1017/S1041610215000034)] [Medline: [25648589](#)]
58. Rentz M, Von Hoene A. Online coaching for caregivers: using technology to provide support and information. *Alzheimer's Care Today* 2010;11(3):206-209. [doi: [10.1097/ACQ.0b013e3181ebc878](https://doi.org/10.1097/ACQ.0b013e3181ebc878)]
59. Schaller S, Marinova-Schmidt V, Gobin J, Criegee-Rieck M, Griebel L, Engel S, et al. Tailored e-Health services for the dementia care setting: a pilot study of 'eHealthMonitor'. *BMC Med Inform Decis Mak* 2015 Jul 28;15:58 [FREE Full text] [doi: [10.1186/s12911-015-0182-2](https://doi.org/10.1186/s12911-015-0182-2)] [Medline: [26215731](#)]
60. Schaller S, Marinova-Schmidt V, Setzer M, Kondylakis H, Griebel L, Sedlmayr M, et al. Usefulness of a tailored ehealth service for informal caregivers and professionals in the dementia treatment and care setting: the eHealthMonitor Dementia Portal. *JMIR Res Protoc* 2016 Apr 05;5(2):e47 [FREE Full text] [doi: [10.2196/resprot.4354](https://doi.org/10.2196/resprot.4354)] [Medline: [27050401](#)]
61. Weitzman P, Neal L, Chen H, Levkoff SE. Designing a culturally attuned bilingual educational website for US Latino dementia caregivers. *Ageing Int* 2008 Feb 28;32(1):15-24. [doi: [10.1007/s12126-008-9000-9](https://doi.org/10.1007/s12126-008-9000-9)]
62. von Winterfeldt D, Edwards W. *Decision Analysis and Behavioral Research*. Cambridge, UK: Cambridge University Press; 1986.
63. Sum S, Mathews RM, Hughes I, Campbell A. Internet use and loneliness in older adults. *Cyberpsychol Behav* 2008 Apr;11(2):208-211. [doi: [10.1089/cpb.2007.0010](https://doi.org/10.1089/cpb.2007.0010)] [Medline: [18422415](#)]
64. Tsai H, Tsai Y. Changes in depressive symptoms, social support, and loneliness over 1 year after a minimum 3-month videoconference program for older nursing home residents. *J Med Internet Res* 2011;13(4):e93 [FREE Full text] [doi: [10.2196/jmir.1678](https://doi.org/10.2196/jmir.1678)] [Medline: [22086660](#)]
65. Chang E, Easterbrook S, Hancock K, Johnson A, Davidson P. Evaluation of an information booklet for caregivers of people with dementia: an Australian perspective. *Nurs Health Sci* 2010 Mar;12(1):45-51. [doi: [10.1111/j.1442-2018.2009.00486.x](https://doi.org/10.1111/j.1442-2018.2009.00486.x)] [Medline: [20487325](#)]

66. Davies N, Maio L, Rait G, Iliffe S. Quality end-of-life care for dementia: what have family carers told us so far? A narrative synthesis. *Palliat Med* 2014 Jul;28(7):919-930 [[FREE Full text](#)] [doi: [10.1177/0269216314526766](https://doi.org/10.1177/0269216314526766)] [Medline: [24625567](#)]
67. Hirschman KB, Kapo JM, Karlawish JHT. Why doesn't a family member of a person with advanced dementia use a substituted judgment when making a decision for that person? *Am J Geriatr Psychiatry* 2006 Aug;14(8):659-667. [doi: [10.1097/01.JGP.0000203179.94036.69](https://doi.org/10.1097/01.JGP.0000203179.94036.69)] [Medline: [16861370](#)]
68. Caron CD, Griffith J, Arcand M. End-of-life decision making in dementia. *Dementia* 2016 Jul 26;4(1):113-136. [doi: [10.1177/1471301205049193](https://doi.org/10.1177/1471301205049193)]
69. Mezey M, Kluger M, Maislin G, Mittelman M. Life-sustaining treatment decisions by spouses of patients with Alzheimer's disease. *J Am Geriatr Soc* 1996 Feb;44(2):144-150. [Medline: [8576503](#)]
70. Davies S, Nolan M. 'Making the best of things': relatives' experiences of decisions about care-home entry. *Ageing Soc* 2003 Jun 25;23(04):429-450. [doi: [10.1017/S0144686X03001259](https://doi.org/10.1017/S0144686X03001259)]
71. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? *Soc Sci Med* 2007 May;64(9):1853-1862. [doi: [10.1016/j.socscimed.2007.01.012](https://doi.org/10.1016/j.socscimed.2007.01.012)] [Medline: [17328998](#)]
72. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. *Am J Prev Med* 2016 Nov;51(5):843-851. [doi: [10.1016/j.amepre.2016.06.008](https://doi.org/10.1016/j.amepre.2016.06.008)] [Medline: [27745684](#)]
73. Bernhardt JM. Health education and the digital divide: building bridges and filling chasms. *Health Educ Res* 2000 Oct;15(5):527-531. [Medline: [11184212](#)]

Abbreviations

RCT: randomized controlled trial

Edited by G Eysenbach; submitted 01.12.17; peer-reviewed by K Harrison-Dening, C Callahan, Y Hu, R Pankomera; comments to author 11.01.18; revised version received 02.03.18; accepted 08.05.18; published 12.06.18.

Please cite as:

Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, Ross J, Davies N
Internet-Based Interventions Aimed at Supporting Family Caregivers of People With Dementia: Systematic Review
J Med Internet Res 2018;20(6):e216
URL: <http://www.jmir.org/2018/6/e216/>
doi: [10.2196/jmir.9548](https://doi.org/10.2196/jmir.9548)
PMID: [29895512](https://pubmed.ncbi.nlm.nih.gov/29895512/)

©Jenny Hopwood, Nina Walker, Lorraine McDonagh, Greta Rait, Kate Walters, Stephen Iliffe, Jamie Ross, Nathan Davies. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 12.06.2018. 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.

Original Paper

Effect Evaluation of a Web-Based Coaching Intervention to Support Implementation of Sex Education Among Secondary School Teachers: Randomized Controlled Trial

Lisette Schutte^{1,2}, PhD; Fraukje E.F. Mevissen¹, PhD; Suzanne Meijer², MPH; Theo Paulussen³, PhD; Pepijn van Empelen³, PhD; Gerjo Kok¹, PhD

¹Department of Work and Social Psychology, Maastricht University, Maastricht, Netherlands

²Department of Youth, STI AIDS Netherlands, Amsterdam, Netherlands

³Expertise Group Child Health, Netherlands Organisation for Applied Scientific Research, Leiden, Netherlands

Corresponding Author:

Lisette Schutte, PhD

Department of Work and Social Psychology

Maastricht University

PO Box 616

6200 MD

Maastricht, 6200 MD

Netherlands

Phone: 31 433881908

Fax: 31 433671032

Email: lisette.schutte@maastrichtuniversity.nl

Abstract

Background: The quality of implementation is important to ensure the effectiveness of behavioral change interventions in practice. Implementing such programs with completeness and adherence is not an automatic process and may require additional support. In school settings, the support teachers receive during implementation is often limited and appears to fall short when attempting to preserve completeness and adherence in program delivery. With the aim to improve completeness and adherence of teachers' delivery of a sexual health promoting intervention ("Long Live Love" [LLL]) in secondary education, a Web-based e-coach was developed ("lesgevenindiefde.nl" or "teachinglove.nl"). The effectiveness of the e-coach, as part of a broader implementation strategy, in influencing teachers' implementation was evaluated.

Objective: This study aimed to report on the effect evaluation to determine the effect of the Web-based e-coach on teacher implementation of a school-based sex education program called LLL and on its determinants.

Methods: A cluster randomized controlled trial (e-coaching vs waiting list control) was conducted with a baseline assessment (T0) and follow-up (T1) 2 weeks after completing the LLL program. A total of 43 schools with 83 teachers participated in the study. In the follow-up, 38 schools participated, 23 in the e-coaching condition with 41 teachers and 15 in the control condition with 26 teachers. Multilevel regression analysis was used to evaluate the effect of the e-coaching website on implementation behavior, namely, completeness and adherence to LLL implementation, and on its determinants.

Results: The e-coaching intervention was not found to have an effect on teachers' implementation behavior; teachers assigned to the experimental e-coaching website did not score higher on completeness ($P=.60$) or adherence ($P=.67$) as compared with teachers in the control condition. When comparing the 30 teachers who made actual use of the e-coaching website with the 37 teachers who did not, no significant differences were found either ($P\geq.54$). In addition, there was no effect of e-coaching on the determinants of teacher implementation behavior ($t_{67-75}\leq 0.69$; $P\geq.22$).

Conclusions: E-coaching was not found to be effective in enhancing completeness of and adherence to LLL by teachers. The lack of effect may be attributed to the intervention content, the limited use, or the study design itself. The e-coaching intervention may not have adequately addressed adherence and completeness of LLL to bring about behavioral change. Furthermore, the e-coaching intervention was not or insufficiently used by teachers. A possible biased sample of motivated, able teachers may have agreed to participate in the study, and a possible "ceiling effect" may have been present because of the high implementation

grade. This, however, does not imply that Web-based coaching in itself is an ineffective strategy to promote adherence and completeness of program implementation. A process evaluation is required as follow-up.

Trial Registration: International Standard Randomised Controlled Trial Number ISRCTN11754581; <http://www.isrctn.com/ISRCTN11754581> (Archived by WebCite at <http://www.webcitation.org/70C5TU0Oh>)

(*J Med Internet Res* 2018;20(6):e96) doi:[10.2196/jmir.7053](https://doi.org/10.2196/jmir.7053)

KEYWORDS

sex education; randomized controlled trial; coaching; secondary schools

Introduction

Background

Implementation is important to ensure the effectiveness of an intervention in practice. An intervention that is implemented completely and according to its guidelines is more likely to be successful in changing the target groups' determinants and behavior than programs that are not implemented fully [1-3]. In school-based sexual health promotion, teachers are the key players for the implementation of these programs. Their implementation is, however, often suboptimal; programs are not being implemented completely or with sufficient fidelity to produce measurable outcomes [4-8]. Several terms are used interchangeably to describe the fidelity of implementing an intervention. Fidelity has different dimensions or aspects, namely, the degree to which an intervention is conducted: (1) competently (competence) and (2) according to protocol (adherence). Adherence refers to the extent to which a program is implemented conforming to the guidelines. Competence relates to skillfulness in the delivery of the program, and thus how well it is implemented [3]. There is a need for greater attention to the quality of implementation and its related determinants, namely, teachers' beliefs about the innovation and characteristics of the innovation, organizational factors, and characteristics of the implementation-enhancing intervention [9,10,6].

Supporting Teachers During Implementation

Teachers appear to be in need of support in every phase of the implementation process to enable them to put the innovation into practice [3,11,12]. Supporting teachers in the implementation phase has, however, been insufficiently considered, as most work has been invested to promote teachers' awareness and adoption of new interventions [4,13,9,14]. This applies in particular to school-based sex education programs, which address the sensitive subject of sexuality. Providing support before implementation in the form of training often equips teachers with skills for correct implementation, but this is not enough [15]. It remains important to provide teachers with more personal assistance and ongoing support and consultation during the process of putting an innovation into practice [6,16,17,14]. Currently, this support is limited to providing practical support in the form of teacher manuals with practical information on the content of the lessons and on how to deliver such lessons. However, more in-depth coaching focusing on determinants of implementation such as self-efficacy and social support to enhance completeness and fidelity is lacking [18,19,11,20-26].

Web-Based Coaching to Improve Teachers' Implementation of School-Based Sex Education

To stimulate the correct implementation, with completeness and adherence, of the (revised) school-based sex education program "Long Live Love" (LLL) [27,6], an e-coaching intervention ("lesgeveinindiefde.nl" or "teachinglove.nl") was systematically developed [28]. The e-coach aimed to improve teachers' implementation behavior through self-reflection and skills development. Through e-coaching, we aimed at making teachers aware of the importance of completeness and adherence in relation to program effectiveness and increase their awareness regarding their own (suboptimal) implementation behavior. In addition, the e-coach provided tools to help teachers improve their implementation by giving support on how to deal adequately with potentially difficult classroom and reproductive health (SRH) lessons, such as creating a safe atmosphere in the classroom for students to openly discuss relationships and sexuality, handling personal questions addressed to teachers by students, and intervening on negative remarks or behavior toward homosexuality. The content of the e-coach was guided by theories on implementation behavior [10,29] and based on a needs assessment among the target group [28]. Intervention objectives were psychosocial determinants, such as awareness, teachers' personal benefit, social support, (anticipated) student responses, and self-efficacy. The e-coach could be used by teachers before and during deliverance of the LLL program. For a more detailed description of the e-coach, refer to the study by Schutte et al [28].

The e-coach was part of a broader strategy for implementation, aimed at promoting each phase of the LLL implementation process. The municipal health services (MHS) were involved in the development and delivery of the implementation strategy. This strategy included instruction protocols used by the MHS to promote adoption and continuation of LLL by teachers in schools, a teacher training delivered by the MHS, and a teacher manual to enhance and facilitate implementation. The MHS training was aimed at introducing the revised LLL program to teachers and motivating them to use the program and use it as intended by enhancing teachers' knowledge, attitudes, and skills. The training was provided before implementation of LLL and was followed by e-coaching. An effect and process evaluation for the pilot implementation of the coaching website was conducted. This occurred simultaneously with the pilot implementation of the revised school-based LLL intervention for students [30]. The aim of this study was to determine the effect of e-coaching on (determinants of) teachers'

implementation behavior. The process evaluation is described elsewhere (unpublished data [31]).

Methods

Design

A clustered randomized controlled trial (e-coaching vs waiting-list control) was conducted, with a baseline assessment (T0) and follow-up (T1) 2 weeks after completing the LLL program. Teachers were not informed about the existence of these 2 groups.

Recruitment and Procedure

From all the secondary schools in the Netherlands, 19.0% (115/610) of the schools were randomly selected after stratification according to region and education level (preparatory applied education, higher general continued education, and preparatory scholarly education). Teachers within these schools were invited by email and telephone to use the revised LLL program and to participate in a survey study on their experience and implementation of SRH education and LLL. Only teachers who taught SRH were contacted. In the Netherlands, teachers are the primary decision makers in the use of SRH programs [32]. The schools with teachers who accepted the invitation (n=45) were randomly assigned to either the control (n=20) or the intervention (e-coach) group (n=25).

Teachers in the intervention and control group who consented to participation first received the baseline survey (T0) by post. The T0 survey focused on determinants of SRH and LLL implementation and took approximately 30 min to fill out. Teachers had 2 weeks to complete and return the survey. Nonresponders got a reminder by email and eventually by telephone 3 days after the deadline, and were given another 2 weeks to return the survey.

At the same time, teachers in both groups were offered a training from the MHS in their region before implementing the revised LLL program. The training was offered to but not taken for personal reasons by 2 schools (4 teachers in the control group and 1 teacher in the intervention group). Finally, 39 of the participating teachers in the survey (24 from the intervention group and 15 from the control group) from 19 schools received training from 14 different MHS, as indicated at T1. The remaining teachers from 19 schools (11 teachers from the control group and 17 teachers from the intervention group) did not receive training, either because they refused the training as they felt there was no need or because the MHS in their region was not offering the training. Separate trainings were delivered to teachers in the e-coach intervention group versus the control group, with teachers in the intervention group receiving additional information during the training about the e-coaching website and being stimulated to use it during the implementation of LLL.

Teachers in both groups then received the LLL program (a package including a student magazine, a student DVD, and a teacher manual) by post mail, which they could implement within (approximately) 2 months following the baseline

measurement for teachers (T0). Additionally, teachers in the intervention group were given access to the e-coaching website with a personal user name and password, and an edition of the LLL teacher manual, which contained references to the website. The teachers in the control group were not exposed to or informed about the website, until after the end of the e-coach evaluation. They received the regular LLL teacher manual without any references to the e-coach. Halfway during the pilot implementation, an email reminded teachers in the intervention group to use the e-coaching website. In addition, 1 week before the expected completion of the LLL program, all teachers were reminded by email and telephone about the upcoming posttest questionnaire (T1). Within 2 weeks after completing the implementation of the LLL program, the T1 survey was sent to all teachers. Reminders were sent by email and eventually by telephone to nonresponders. All procedures in the study were approved by the authorized Ethical Review Committee of Psychology & Neuroscience at Maastricht University. Registration of this trial was not required in the Netherlands as it is a nonmedical paper and is uncommon for psychological research such as this.

Measurements

The survey used for the effect evaluation focused on determinants targeted by the e-coach and was based on the theoretical framework explaining teachers' adoption and implementation of SRH developed by Paulussen et al [33], which is a combination of the theory of planned behavior [34], social cognitive theory [35], and diffusion of innovations theory [36]. Further description of and foundation for this framework can be found in a study conducted by Schutte et al [6]. At baseline (T0), we measured background characteristics of the teachers, including (SRH and LLL) teaching experience and their LLL curriculum-related beliefs and student response. At posttest (T1), we measured the same determinants but also included measures on completeness of and adherence to LLL implementation. In addition, subjective evaluations of the e-coach and the MHS training were included (this will be further discussed in the process evaluation) [31].

Demographic variables (T0) included gender, age, teaching subject, educational level of students, years of teaching experience, years of teaching SRH, perceived expertise in teaching SRH, perceived need for support in providing SRH, attitude toward teaching SRH, attitude toward reflecting on own SRH teaching methods, past experience with previous versions of LLL, and sexual morality.

Table 1 provides an overview of all *outcome measures*. For measuring *curriculum-related beliefs* (T0 and T1), *teacher benefits*, *subjective norm*, *social support*, and *self-efficacy* were assessed together with (*anticipated*) *student responses*.

Teacher's implementation behavior (T1) was measured based on rates of completeness of and adherence to LLL implementation. *Completeness* was expressed by the proportion of the 19 core learning activities of the LLL program being implemented ($\sum \text{implemented activities} / 19 \times 100$). In this study, adherence was measured as one aspect or dimension of fidelity.

Table 1. Measures, number of items, reliability, example items, and answer scale.

Measurements	Items	Cronbach alpha	Exemplary items (response scales)
Demographic variables			
Gender	1		What is your gender? (0=female, 1=male)
Age	1		What is your age?
Teaching subject	1		What subject do you teach? (1=biology, 2=health care, 3=citizenship, 4=other)
Years of teaching experience	1		How many years have you been working in education?
Years teaching SRH ^a	1		How many years have you been teaching SRH?
Perceived expertise teaching SRH	1		How experienced are you in teaching SRH? (1=very inexperienced, 7=very experienced)
Perceived need for support in providing SRH	1		Do you need support in providing SRH lessons? (1=no, certainly not; 5=yes, certainly)
Attitude toward teaching SRH	6	.82	Indicate what you think about teaching SRH: Teaching SRH is... (important, necessary, fun, difficult, comfortable, competent; 1=not at all, 7=yes totally)
Attitude toward reflecting on own SRH teaching methods	1		Indicate what you think about reflecting on your own SRH teaching methods: (important, useful, good; 1=not at all, 7=yes, totally)
Use of previous LLL ^b	1		Have you used the previous LLL in the past for SRH lessons? (0=no, 1=yes)
Years teaching LLL	1		For how many years have you been using LLL?
Sexual morality	5	.62	Young people who have just met should not have sex (1=strongly disagree, 5=strongly agree)
Curriculum-related beliefs			
Teacher benefits	6	.72	I gained insight in the sexuality experience of youngsters (1=strongly disagree, 5=strongly agree)
Subjective norms	6	.81	Do you think that the following people appreciate you using LLL to provide sexual education? (principal, governing body, external consultants or health education experts, students, colleagues teaching the same and colleagues teaching a different subject, parents; 1=no, certainly not; 5=yes, certainly)
Social support	6	.75	Do you expect support from the following people when implementing LLL? (governing body, colleagues teaching the same and different subjects, and the parent association; 1=no, certainly not, 5=yes, certainly)
Self-efficacy	12	.76	I am able to create a safe atmosphere in the classroom where students feel safe to openly talk about sex and relationships (1=no, certainly not; 5=yes, certainly)
Interactive context			
Student response	3	.63	Indicate how students generally respond to LLL (interested, shy, positively; 1=not at all, 7=yes totally)
E-coaching			
Used at all	1		Did you visit the “Lesgeven in de Liefde” website for teachers during your use of the new LLL program (0=no, never; 1=yes)
Implementation behavior			
Completeness—calculated as percentage of the program (ie, learning activities) being implemented. (ie, Σ learning activities/19 \times 100)	1		Did you cover this (learning activity)? (1=yes, 0=no)

Measurements	Items	Cronbach alpha	Exemplary items (response scales)
Adherence	1		How did you implement the new LLL program? (1=I reviewed the program and only selected a few ideas for my SRH lessons, 2=I reviewed the program and selected many ideas for my SRH lessons, 3=I used the program as a guideline for my lessons and delivered some lesson suggestions according to the teacher manual, 4=I followed the guidelines of the program as closely as possible and delivered most lesson suggestions according to the teacher manual, and 5=I delivered all lesson suggestions for the LLL program exactly according to the teacher manual)

^aSRH: sexual and reproductive health.

^bLLL: Long Live Love.

Adherence was measured as the extent to which the LLL program was implemented according to the guidelines as prescribed in the teacher manual, with scores ranging from 1 (“I reviewed the program and only delivered a few lesson suggestions according to the teacher manual”) to 5 (“I delivered all lesson suggestions for the LLL program exactly according to the teacher manual”) [37]. All measures, including number of items, response scales, reliability, and exemplary items, are presented in Table 1.

Analyses

Data were analyzed using IBM SPSS Statistics 24. Given the nested structure of the design and the data (partly repeated; measurements nested within teachers nested within schools), multilevel regression analyses were used to evaluate the effects of e-coaching on teachers’ implementation of LLL and its determinants. Unstandardized regression coefficients are reported, along with the standard error of beta. An additional advantage of the mixed regression is that it takes into account all participants, including those with only one measure. An unstructured covariance matrix for the repeated measures was selected. Two levels were defined in the multilevel analysis: school and teacher. The pseudo R^2 was calculated at school level and at teacher level for outcome variables measured only at posttest (completeness and adherence) and for outcome variables measured at pre- and posttest.

The model included the predictors group (1 for intervention group [e-coach] and 0 for control group) for the outcomes of implementation behavior (completeness and adherence), and group, time of measurement (baseline and posttest), and the interaction time \times group for the determinants. The mixed model was estimated with the restricted maximum likelihood method to obtain unbiased variance estimates. The intraclass correlation coefficient was between .18 and .43, confirming that multilevel analysis was required.

Results

Participants Flow

Of the 115 schools approached, a total of 45 schools, including 112 teachers, agreed to participate in the pilot implementation of the revised LLL and the evaluation of their experience with implementing LLL. Teachers’ nonwillingness to participate was predominantly because of sexual education already having been provided in the school and lack of time. The schools were randomly assigned to either the waiting-list control group (20

schools including 46 teachers) or the e-coach intervention group (25 schools including 66 teachers). In addition, 2 schools (one from each condition, including 6 teachers) withdrew before the start of the pilot implementation of LLL because of internal organizational changes leaving 43 schools (106 teachers) at baseline. On average, there were 2 to 3 teachers per school. At baseline (T0), the survey was completed by 83 teachers ($n=50$ in the intervention group and $n=33$ in the control group) from 43 schools. Nonresponse was mainly because of lack of time. Follow-up measurement (T1) was completed by 67 teachers (80% of those completing T0; $n=41$ in the intervention group) from 38 schools. In addition, dropout at T1 ($n=16$) was mainly caused by lack of time. See Figure 1 for school allocation and participant flow. A dropout analysis, accounting for teachers’ background characteristics, indicated no significant differences between teachers who dropped out versus those who did not dropout ($t_{29,147} \leq -0.36$; $P \geq .07$). Moreover, there was no significant difference in dropout between teachers in the intervention and control groups ($\chi^2_1=0.5$; B (regression weight) = $-.32$; $P=.48$).

Participants

Of the 83 teachers participating in the baseline questionnaire, 53 were female (64%) and 58 (70%) were biology teachers. The other teachers either taught the subject care ($n=19$; 22.89%) or citizenship ($n=6$; 7.23%). The mean age was 43 years. Years of teaching experience ranged from 1 to 39 years, whereas years of experience teaching sexual education ranged from 0 to 35. Teachers generally felt fairly experienced in teaching SRH, had a positive attitude toward teaching SRH, had a positive attitude toward reflecting on their own SRH teaching methods, had a positive attitude toward teaching SRH, and a permissive sexual morality. Teachers expressed a limited need for support in providing SRH lessons. One-third of the teachers had experience with the previous LLL program, ranging from 1 to 10 years. No differences could be observed at baseline between the intervention and control groups (see Table 2).

Effects of E-Coaching on (Determinants of) Implementation of Long Live Love

Overall, teachers reported completing on average 73% of the LLL program (range 37%-98%), and 43% of teachers reported implementing the program largely in accordance with the guidelines in the teacher manual (mean 3.46, SD 0.75). Only 6% ($n=4$) implemented LLL exactly conforming to the guidelines in the teacher manual. No significant difference was

found in completeness (regression weight=-2.12 [SE=3.99]; 95% CI -10.26 to 6.02; $P=.60$) or adherence (regression weight=0.09 [SE=.21]; 95% CI -0.33 to 0.51; $P=.67$) between teachers in the control group as compared with teachers in the e-coaching group, with small effect sizes (.17 and .14, respectively); see Table 3). In addition, no significant time \times group interaction effect was found for the determinants of implementation behavior ($t_{67-75} \leq 0.69$; $P \geq .22$; see Table 4). All pseudo R^2 values were smaller than .19 across all outcomes at

both the school and teacher level. On the basis of the survey, it turned out that of the 41 teachers in the intervention group, 30 actually visited the website (75%). When comparing the 30 teachers who made actual use of the e-coaching website with the 37 teachers who did not, still no significant differences were found in completeness (regression weight=-2.21 [SE=3.62]; 95% CI for B -5.03 to 9.46; $P=.54$) or adherence of LLL (regression weight=0.06 [SE=0.19]; 95% CI for B -0.32 to 0.44; $P=.74$). No significant differences were found between determinants either ($t_{28-38} \leq 0.08$; $P \geq .29$).

Figure 1. School allocation and participant flow.

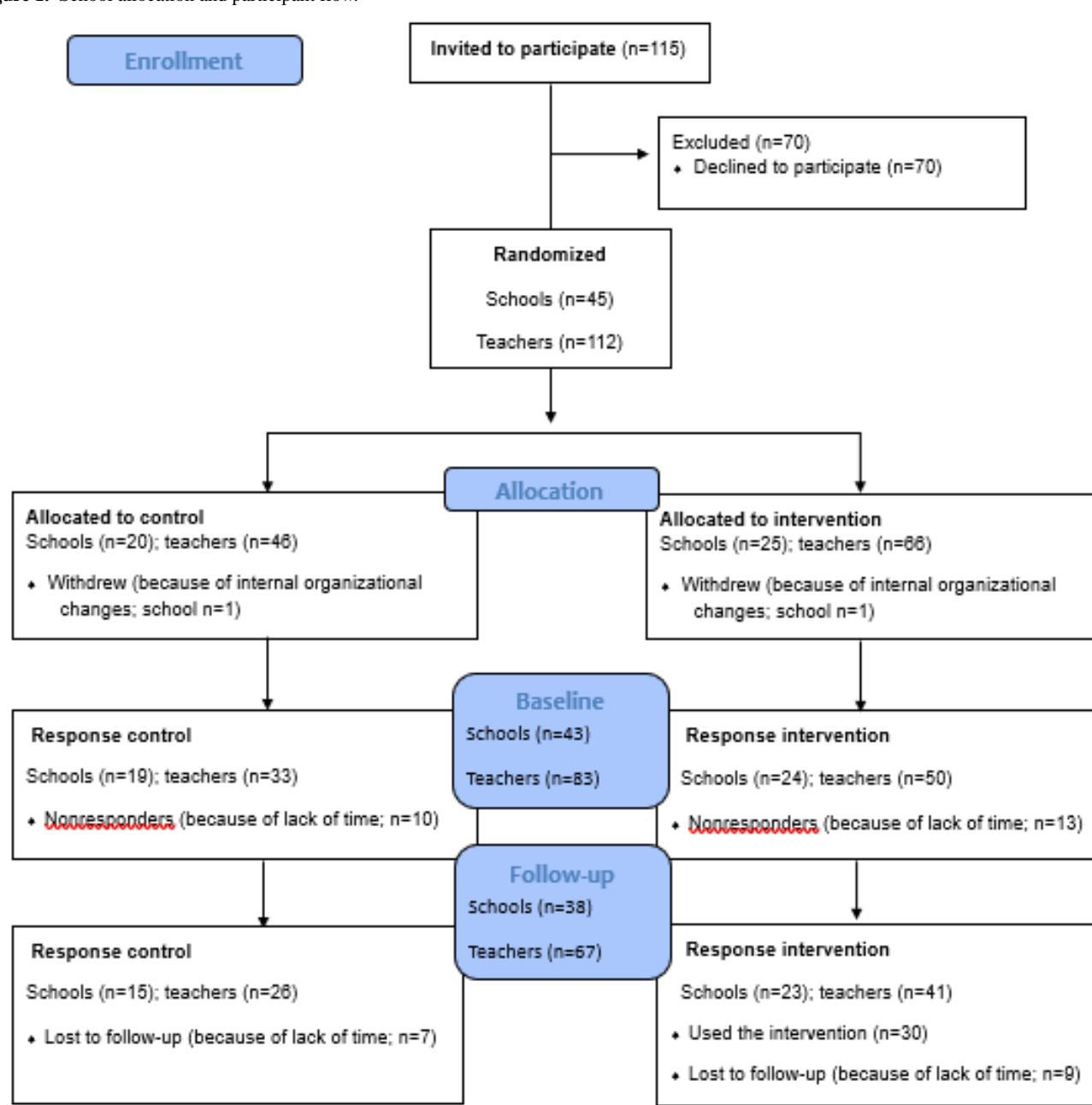


Table 2. Demographic variables of teachers at T0.

Demographic variables (range)	Mean (SD): baseline (T0)		
	Total (n=83)	C-group ^a (n=33)	I-group ^b (n=50)
Age (23-64 years)	43.11 (11.38)	42.97 (11.30)	42.94 (11.59)
Years of teaching experience (1-39)	14.37 (10.43)	13.67 (10.86)	14.60 (10.08)
Years teaching SRH ^c (0-35)	9.20 (7.91)	9.61 (7.73)	9.12 (8.09)
Experience teaching SRH (1-7)	4.63 (1.70)	4.85 (1.64)	4.50 (1.73)
Years teaching LLL ^d (1-10)	4.80 (2.54)	4.44 (2.13)	5.21 (2.97)
Sexual morality (1-5)	2.33 (0.58)	2.28 (0.61)	2.38 (0.56)
Attitude toward teaching SRH (1-7)	5.87 (0.87)	5.86 (0.85)	5.87 (0.88)
Attitude toward reflecting on own SRH teaching methods (1-7)	5.94 (0.98)	6.03 (0.96)	5.86 (1.02)
Perceived need for support in providing SRH (1-5)	3.18 (1.04)	3.36 (.90)	3.08 (1.12)

^aC-group: control group.^bI-group: intervention group.^cSRH: sexual and reproductive health.^dLLL: Long Live Love.**Table 3.** Effect of e-coaching on teachers' implementation behavior.

Implementation behavior	C-group ^a (n=26), mean (SD)	I-group ^b (n=41), mean (SD)	Regression weight B (SE ^c)	P value	95% CI for B
Completeness LLL ^d (%)	73.85 (13.84)	72.35 (14.54)	-2.12 (3.99)	.60	-10.26 to 6.02
Adherence LLL (1-5)	3.38 (0.70)	3.51 (0.78)	0.09 (0.21)	.67	-0.33 to 0.51

^aC-group: control group.^bI-group: intervention group.^cSE: standard error.^dLLL: Long Live Love.**Table 4.** Effect of e-coaching on determinants of implementation behavior.

Determinants	Pretest		Posttest		Regression weight B (SE ^a)	P value	95% CI for B
	C-group ^b (n=26), mean (SD)	I-group ^c (n=41), mean (SD)	C-group (n=26), mean (SD)	I-group (n=41), mean (SD)			
Teacher benefits	3.61 (0.56)	3.57 (0.64)	3.21 (0.71)	3.08 (0.49)	-0.05 (0.16)	.77	-0.36 to 0.26
Subjective norms	4.08 (0.53)	4.03 (0.47)	4.24 (0.55)	4.09 (0.54)	-0.08 (0.15)	.60	-0.38 to 0.22
Social support	4.17 (0.44)	4.14 (0.57)	4.26 (0.55)	4.30 (0.55)	0.09 (0.15)	.54	-0.2 to 0.38
Self-efficacy	4.11 (0.33)	4.07 (0.39)	4.18 (0.47)	4.16 (0.42)	-0.02 (0.11)	.87	-0.23 to 0.19
Student response	5.71 (0.77)	5.51 (1.04)	5.37 (0.99)	5.17 (0.98)	0.03 (0.28)	.92	-0.53 to 0.58

^aSE: standard error.^bC-group: control group.^cI-group: intervention group.

Discussion

Principal Findings

An e-coaching intervention was systematically developed to stimulate adherence and completeness of use of the revised Dutch secondary school-based sex education program LLL. The aim of this study was to improve teachers' implementation behavior through self-reflection and skills development. The

e-coaching was part of a broader implementation strategy that included a teacher training from the MHS before implementation.

Despite e-coaching being systematically developed, and with the input of experienced teachers, e-coaching was not found to be effective in changing teachers' implementation behavior or its determinants. In general, teachers implemented the new LLL program moderately during the pilot study. The lack of effect

could be a reflection of the intervention itself not being effective, either because of its development or implementation [18]. It is however often difficult to prove the effectiveness of interventions or implementation strategy even if they are solidly grounded in theory and evidence [18].

Several factors may explain the lack of effect of e-coaching on implementation of LLL. First, the study design itself may have had some flaws. Our study was a randomized controlled trial design and included a baseline measure and posttest; however, the limited number of teachers and schools at posttest could affect the generalizability of the results. In addition, there were 2 to 3 teachers per school, and this limited number may partly explain the poor implementation outcomes. Clustering practitioners using the same program within an organization, in groups of 3 or more, is usually considered an advantage in successful implementation [38]. Moreover, a “ceiling effect” could be present because of the implementation grade of teachers participating in the study already being high, making it difficult to improve using the e-coaching intervention. This suggests that e-coaching or another form of implementation enhancement may have been redundant in this particular case. Finally, the teachers who agreed to participate in the study may have been a biased sample of motivated, experienced teachers who were already capable of delivering LLL successfully. An additional methodological limitation in this study is the self-reported data [39]. Although this is often used to assess completeness and fidelity, methods of observation could perhaps further validate the results. Although the implementation strategy considered the individual teacher as well as the broader environment such as schools, MHS, and municipality, the focus was predominantly on individuals within these organizations. For example, no comparisons were made in implementation success of LLL and e-coach between schools. Considering the influence of decision-making processes in schools and organizations to influence top-down policy formation at the management level could strengthen sustainability of implementation [40].

Second, for an intervention to have an effect, it is important that the intervention is used and positively perceived. By not being used or insufficiently used by teachers, e-coaching is unlikely to have an effect [40-42]. Despite being designed to support teachers in their implementation of LLL, the website itself also needed to be effectively implemented. Teachers were perhaps not motivated to use the website because of their extensive experience in teaching SRH. Additionally, the broader implementation strategy developed to inform teachers about the e-coaching website was perhaps not optimally utilized, despite involvement of MHS professionals and teachers in the development process, potentially resulting in limited use and lack of effect of the website. For example, in this study, not all MHS provided a training and not all teachers who were offered a training accepted it. The broader implementation strategy could potentially be optimized to increase use of e-coaching by teachers. Motives for teachers’ use or nonuse of e-coaching need to be further explored, as well as means to increase use of the website by teachers. Taking contextual factors and individual factors into consideration remains important when stimulating implementation [6,18,29,43].

Finally, the intervention itself may have been suboptimal. The e-coach was aimed at determinants of completeness and fidelity but may not have addressed the exact needs of the target population, or been able to increase teachers’ awareness of the importance of completeness and adherence, or did not address completeness and adherence sufficiently or adequately. In developing e-coaching, program developers were already aware of the following challenges involved: (1) teachers did not see their suboptimal implementation behavior as problematic and (2) teachers expressed a minimal need for coaching during the interviews in the needs assessment phase of program development [28]. The developers attempted to address these challenges in the development of e-coaching by using an unobtrusive coaching technique and stimulating self-reflection, yet the question remained whether this would be successful or not. The e-coach intervention may have been unable to change teachers’ perceived need for coaching or change their awareness of their suboptimal implementation behavior with regard to completeness and adherence, which may be linked to a lack of effect. This re-emphasizes the importance of having a need for coaching or a desire for change before behavioral change [44]. Means of stimulating teachers to use the website need to be explored.

Implications

Knowledge about implementation of Internet interventions and implementation of eHealth in the school settings particularly needs enhancement [45]. Website use was found to be related to factors associated with the visitor and the intervention website [46-48]. A large study in the Netherlands found that Information Technology use by teachers is limited. They either consult colleagues in their school for information or use the Internet mainly to find information, prepare their lessons, send emails to students, or give homework assignments and thus less for professional development [49]. Digital technologies are being increasingly used in the education system, bringing exciting opportunities for innovative ways of teaching and learning [50].

The strength of e-coaching is that, in addition to being aimed at specific determinants, it provides more than just a one-time training. Instead, it provides assistance during real-life implementation situations and has a longitudinal character in that teachers can visit the website as desired or need be [51]. Although completeness and adherence to program delivery are crucial to the effectiveness of the program, teaching quality of SRH lessons encompasses other teacher classroom-related skills, such as creating a safe and trusted environment, which form the conditions for providing these lessons. Such skills are addressed by e-coaching. In addition, other studies on providing sexual education have highlighted the importance of creating a safe environment when teaching this subject for optimal results [52,53]. “Teaching well” is thus more than completeness and fidelity. Therefore, in stimulating implementation of SRH programs, program developers should focus on enhancing completeness and adherence as well as supporting teachers in creating the classroom conditions that enable quality delivery of SRH lessons, as e-coaching has attempted to do.

Conclusions

The lack of effect of e-coaching does not insinuate that Web-based coaching in itself is an ineffective strategy to promote adherence and completeness of program implementation, but in its current form, e-coaching is not the

optimal instrument to achieve adherence and completeness of LLL specifically. To further understand why e-coaching had no effect and how it could potentially be improved, a process evaluation is required to find out how and to what extent teachers made use of the website, how they appreciated it, and what factors affected teachers' use of the website.

Acknowledgments

This work was funded by ZonMw and supported by Maastricht University; Netherlands Organisation for Applied Scientific Research, Expertise Group Child Health; and STI AIDS Netherlands. The authors thank and appreciate all the teachers who participated in the research.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was only retrospectively registered. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 544KB - jmir_v20i6e96_app1.pdf](#)]

References

1. Little MA, Riggs NR, Shin HS, Tate EB, Pentz MA. The effects of teacher fidelity of implementation of pathways to health on student outcomes. *Eval Health Prof* 2015 Mar;38(1):21-41 [[FREE Full text](#)] [doi: [10.1177/0163278713489879](https://doi.org/10.1177/0163278713489879)] [Medline: [23739725](https://pubmed.ncbi.nlm.nih.gov/23739725/)]
2. Lendrum A, Humphrey N. The importance of studying the implementation of interventions in school settings. *Oxf Rev Educ* 2012 Oct 08;38(5):635-652. [doi: [10.1080/03054985.2012.734800](https://doi.org/10.1080/03054985.2012.734800)]
3. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol* 2008 Jun;41(3-4):327-350. [doi: [10.1007/s10464-008-9165-0](https://doi.org/10.1007/s10464-008-9165-0)] [Medline: [18322790](https://pubmed.ncbi.nlm.nih.gov/18322790/)]
4. Bessems KM, van Assema P, de Vries NK, Paulussen TW. Exploring determinants of completeness of implementation and continuation of a Dutch school-based healthy diet promotion programme. *Int J Health Promot Educ* 2014 May 23;52(6):315-327. [doi: [10.1080/14635240.2014.912445](https://doi.org/10.1080/14635240.2014.912445)]
5. Peters LW, Kok G, Ten Dam GT, Buijs GJ, Paulussen TG. Effective elements of school health promotion across behavioral domains: a systematic review of reviews. *BMC Public Health* 2009 Jun 12;9:182 [[FREE Full text](#)] [doi: [10.1186/1471-2458-9-182](https://doi.org/10.1186/1471-2458-9-182)] [Medline: [19523195](https://pubmed.ncbi.nlm.nih.gov/19523195/)]
6. Schutte L, Meertens RM, Mevissen FE, Schaalma H, Meijer S, Kok G. Long Live Love. The implementation of a school-based sex-education program in The Netherlands. *Health Educ Res* 2014 Aug;29(4):583-597 [[FREE Full text](#)] [doi: [10.1093/her/cyu021](https://doi.org/10.1093/her/cyu021)] [Medline: [24817522](https://pubmed.ncbi.nlm.nih.gov/24817522/)]
7. Gottfredson DC, Gottfredson GD. Quality of school-based prevention programs: results from a national survey. *J Res Crime Delinq* 2002 Feb 01;39(1):3-35. [doi: [10.1177/002242780203900101](https://doi.org/10.1177/002242780203900101)]
8. Hoekstra F, Velderman MK, Wiefferink K, van Overveld C, ter Beek J, Buijs G, et al. De adoptiefase van schoolgerichte preventieprogramma's: Van boodschap naar beslissing. *Tijdschrift voor Gezondheidswetenschappen (TSG)* 2009 Jan;87(1):27-36.
9. Fagan AA, Mihalic S. Strategies for enhancing the adoption of school-based prevention programs: Lessons learned from the blueprints for violence prevention replications of the life skills training program. *J Community Psychol* 2003 May;31(3):235-253. [doi: [10.1002/jcop.10045](https://doi.org/10.1002/jcop.10045)]
10. Paulussen T, Kok G, Schaalma H, Parcel GS. Diffusion of AIDS curricula among Dutch secondary school teachers. *Health Educ Q* 1995 May;22(2):227-243. [Medline: [7622390](https://pubmed.ncbi.nlm.nih.gov/7622390/)]
11. Rohrbach LA, Gunning M, Sun P, Sussman S. The project Towards No Drug abuse (TND) dissemination trial: implementation fidelity and immediate outcomes. *Prev Sci* 2010 Mar;11(1):77-88 [[FREE Full text](#)] [doi: [10.1007/s11121-009-0151-z](https://doi.org/10.1007/s11121-009-0151-z)] [Medline: [19757052](https://pubmed.ncbi.nlm.nih.gov/19757052/)]

12. Bishop M, Reiser S, Hall J, Rein J, Taylor A. The hospice model of care: a required web-based program for medical students. *J Palliat Care* 2006 Sep;22(3):219-220.
13. Domitrovich CE, Greenberg MT. The study of implementation: current findings from effective programs that prevent mental disorders in school-aged children. *J Educ Psychol Consult* 2000;11(2):193-221. [doi: [10.1207/S1532768XJEPC1102_04](https://doi.org/10.1207/S1532768XJEPC1102_04)]
14. Durlak JA. Why program implementation is important. *J Prev Interv Community* 1998;17(2):5-18. [doi: [10.1300/J005v17n02_02](https://doi.org/10.1300/J005v17n02_02)]
15. Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Educ Res* 2003 Apr;18(2):237-256. [Medline: [12729182](https://pubmed.ncbi.nlm.nih.gov/12729182/)]
16. Kramer L, Laumann G, Brunson L. Implementation and diffusion of the rainbows program in rural communities: implications for school-based prevention programming. *J Educ Psychol Consult* 2000;11(1):37-64. [doi: [10.1207/s1532768Xjepec1101_04](https://doi.org/10.1207/s1532768Xjepec1101_04)]
17. Rohrbach LA, Graham JW, Hansen WB. Diffusion of a school-based substance abuse prevention program: predictors of program implementation. *Prev Med* 1993 Mar;22(2):237-260. [doi: [10.1006/pmed.1993.1020](https://doi.org/10.1006/pmed.1993.1020)] [Medline: [8483862](https://pubmed.ncbi.nlm.nih.gov/8483862/)]
18. Bartholomew LK, Markham CM, Ruiter R, Fernandez ME, Kok G, Parcel GS. Planning health promotion programs: an intervention mapping approach. San Francisco: Jossey-Bass; 2016.
19. Hansen WB, Bishop DC, Bryant KS. Using online components to facilitate program implementation: impact of technological enhancements to all stars on ease and quality of program delivery. *Prev Sci* 2009 Mar;10(1):66-75. [doi: [10.1007/s11121-008-0118-5](https://doi.org/10.1007/s11121-008-0118-5)] [Medline: [19067165](https://pubmed.ncbi.nlm.nih.gov/19067165/)]
20. Kam C, Greenberg MT, Walls CT. Examining the role of implementation quality in school-based prevention using the PATHS curriculum. Promoting Alternative Thinking Skills Curriculum. *Prev Sci* 2003 Mar;4(1):55-63. [Medline: [12611419](https://pubmed.ncbi.nlm.nih.gov/12611419/)]
21. Mihalic SF, Fagan AA, Argamaso S. Implementing the LifeSkills Training drug prevention program: factors related to implementation fidelity. *Implement Sci* 2008;3:5 [FREE Full text] [doi: [10.1186/1748-5908-3-5](https://doi.org/10.1186/1748-5908-3-5)] [Medline: [18205919](https://pubmed.ncbi.nlm.nih.gov/18205919/)]
22. Farmer - Dougan V, Viechtbauer W, French T. Peer - prompted social skills: the role of teacher consultation in student success. *Educ Psychol* 1999;19(2):207-219. [doi: [10.1080/0144341990190207](https://doi.org/10.1080/0144341990190207)]
23. Jones KM, Wickstrom KF, Friman PC. The effects of observational feedback on treatment integrity in school-based behavioral consultation. *Sch Psychol Q* 1997;12(4):316-326. [doi: [10.1037/h0088965](https://doi.org/10.1037/h0088965)]
24. Mortenson BP, Witt JC. The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psych Rev* 1998;27(4):613-627.
25. Noell GH, Witt JC, Gilbertson DN, Ranier DD, Freeland JT. Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *Sch Psychol Q* 1997 Mar;12(1):77-88. [doi: [10.1037/h0088949](https://doi.org/10.1037/h0088949)]
26. Witt JC, Noelle GH, Lafleur LH, Mortenson BP. Teacher use of interventions in general education settings: measurement and analysis of the independent variable. *J Appl Behav Anal* 1997 Dec;30(4):693-696. [doi: [10.1901/jaba.1997.30-693](https://doi.org/10.1901/jaba.1997.30-693)]
27. Schaalma HP, Kok G, Bosker RJ, Parcel GS, Peters L, Poelman J, et al. Planned development and evaluation of AIDS/STD education for secondary school students in The Netherlands: short-term effects. *Health Educ Q* 1996 Nov;23(4):469-487. [Medline: [8910025](https://pubmed.ncbi.nlm.nih.gov/8910025/)]
28. Schutte L, van den Borne M, Kok G, Meijer S, Mevissen FE. Innovatively supporting teachers' implementation of school-based sex education: developing a web-based coaching intervention from problem to solution. *J Med Internet Res* 2016 Jul 12;18(7):e136 [FREE Full text] [doi: [10.2196/jmir.5058](https://doi.org/10.2196/jmir.5058)] [Medline: [27405241](https://pubmed.ncbi.nlm.nih.gov/27405241/)]
29. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol* 2008 Jun;41(3-4):327-350. [doi: [10.1007/s10464-008-9165-0](https://doi.org/10.1007/s10464-008-9165-0)] [Medline: [18322790](https://pubmed.ncbi.nlm.nih.gov/18322790/)]
30. Hofstetter H, Peters LWH, Meijer S, van Keulen HM, Schutte L, van Empelen P. Evaluation of the effectiveness and implementation of the sexual health program Long Live Love. *European Health Psychologist* 2014;16(Suppl):489.
31. Schutte L, Mevissen FE, Meijer S, Paulussen T, van Empelen P, Kok G. Process evaluation of a Web-based coaching intervention to support implementation of sex education among secondary school teachers. *J Med Internet Res* 2017;5-32 (forthcoming).
32. Paulussen T, Wiefferink K, Mesters I. Invoering van effectief gebleken interventies. In: Brug J, van Assema P, Lechner L, editors. Gezondheidsvoorlichting en gedragsverandering: een planmatige aanpak. Assen: Van Gorcum; 2007:151-172.
33. Paulussen T, Kok G, Schaalma H. Antecedents to adoption of classroom-based AIDS education in secondary schools. *Health Educ Res* 1994;9(4):485-496. [doi: [10.1093/her/9.4.485](https://doi.org/10.1093/her/9.4.485)]
34. Ajzen I. The theory of planned behaviour: reactions and reflections. *Psychol Health* 2011 Sep;26(9):1113-1127. [doi: [10.1080/08870446.2011.613995](https://doi.org/10.1080/08870446.2011.613995)] [Medline: [21929476](https://pubmed.ncbi.nlm.nih.gov/21929476/)]
35. Bandura A, National Inst of Mental Health. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, N. J: Prentice-Hall; 1986:1-617.
36. Rogers EM. Diffusion of innovations. New York: Free Press; 2003.
37. Hall GE, Dirksen DJ, George AA. Measuring implementation in schools: levels of use. Austin, TX: SEDL; 2006.
38. Klest SK. Clustering practitioners within service organizations may improve implementation outcomes for evidence-based programs. *Zeitschrift Fur Psychologie* 2014;222:30-36 [FREE Full text] [doi: [10.1027/2151-2604/a000163](https://doi.org/10.1027/2151-2604/a000163)]

39. Lillehoj C, Griffin K, Spoth R. Program provider and observer ratings of school-based preventive intervention implementation: agreement and relation to youth outcomes. *Health Educ Behav* 2004 Apr;31(2):242-257. [Medline: [15090124](#)]
40. Forman SG, Olin SS, Hoagwood KE, Crowe M, Saka N. Evidence-based interventions in schools: developers views of implementation barriers and facilitators. *School Ment Health* 2008 Nov 25;1:26-36. [doi: [10.1007/s12310-008-9002-5](#)]
41. Domitrovich CE, Bradshaw CP, Poduska JM, Hoagwood K, Buckley JA, Olin S, et al. Maximizing the implementation quality of evidence-based preventive interventions in schools: a conceptual framework. *Adv Sch Ment Health Promot* 2008 Jul;1(3):6-28 [FREE Full text] [Medline: [27182282](#)]
42. Payne AA. Do predictors of the implementation quality of school-based prevention programs differ by program type? *Prev Sci* 2009 Jun;10(2):151-167. [doi: [10.1007/s11121-008-0117-6](#)] [Medline: [19101802](#)]
43. Powell BJ, Beidas RS, Lewis CC, Aarons GA, McMillen JC, Proctor EK, et al. Methods to improve the selection and tailoring of implementation strategies. *J Behav Health Serv Res* 2017 Apr;44(2):177-194 [FREE Full text] [Medline: [26289563](#)]
44. Parsloe E, Leedham M. Coaching and mentoring: practical conversations to improve learning. London: Kogan Page Publishers; Mar 03, 2009.
45. Drozd F, Vaskinn L, Bergsund HB, Haga SM, Slinning K, Bjørkli CA. The implementation of internet interventions for depression: a scoping review. *J Med Internet Res* 2016 Sep 8;18(9):e236 [FREE Full text] [Medline: [27608548](#)]
46. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. *J Med Internet Res* 2011 Jan 06;13(1):e2 [FREE Full text] [Medline: [21212045](#)]
47. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. *Health Educ Behav* 2011 Feb;38(1):49-62 [FREE Full text] [Medline: [21189422](#)]
48. 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 Sep 30;13(3):e71 [FREE Full text] [Medline: [21959968](#)]
49. Voogt J, Sligte H, van den Beemt A, van Braak J, Aesaert K. E-didactiek. Welke ict-applicaties gebruiken leraren en waarom?. Amsterdam: Kohnstamm Instituut; 2016.
50. Ertmer PA, Ottenreit-Leftwich AT. Teacher technology change: how knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education* 2010;42(3):255-284 [FREE Full text]
51. Han SS, Weiss B. Sustainability of teacher implementation of school-based mental health programs. *J Abnorm Child Psychol* 2005 Dec;33(6):665-679 [FREE Full text] [Medline: [16328743](#)]
52. Vanwesenbeeck I, Westeneng J, de Boer T, Reinders J, van Zorge R. Lessons learned from a decade implementing comprehensive sexuality education in resource poor settings: the world starts with me. *Sex Educ* 2016;16(5):471-486. [doi: [10.1080/14681811.2015.1111203](#)]
53. Schaalma HP, Abraham C, Gillmore MR, Kok G. Sex education as health promotion: what does it take? *Arch Sex Behav* 2004 Jun;33(3):259-269. [doi: [10.1023/B:ASEB.0000026625.65171.1d](#)] [Medline: [15129044](#)]

Abbreviations

C-group: control group
I-group: intervention group
LLL: Long Live Love
MHS: municipal health services
SE: standard error
SRH: sexual reproductive health

Edited by G Eysenbach; submitted 28.11.16; peer-reviewed by E Jerves, F Drozd; comments to author 28.02.17; revised version received 31.08.17; accepted 30.11.17; published 18.06.18.

Please cite as:

Schutte L, Mevissen FE, Meijer S, Paulussen T, van Empelen P, Kok G
 Effect Evaluation of a Web-Based Coaching Intervention to Support Implementation of Sex Education Among Secondary School Teachers: Randomized Controlled Trial
J Med Internet Res 2018;20(6):e96
 URL: <http://www.jmir.org/2018/6/e96/>
 doi: [10.2196/jmir.7053](#)
 PMID: [29734139](#)

©Lisette Schutte, Fraukje E.F. Mevissen, Suzanne Meijer, Theo Paulussen, Pepijn van Empelen, Gerjo Kok. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 18.06.2018. 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.

Original Paper

Severely Burdened Individuals Do Not Need to Be Excluded From Internet-Based and Mobile-Based Stress Management: Effect Modifiers of Treatment Outcomes From Three Randomized Controlled Trials

Kiona Krueger Weisel¹, MSc; Dirk Lehr^{2,3}, PhD; Elena Heber^{2,4}, PhD; Anna-Carlotta Zarski^{1,2}, Dipl-Psych; Matthias Berking^{1,2}, PhD; Heleen Riper^{2,5}, PhD; David Daniel Ebert^{1,2}, PhD

¹Department of Clinical Psychology and Psychotherapy, Institute of Psychology, Friedrich-Alexander-University Erlangen-Nuernberg, Erlangen, Germany

²Division of Online Health Training, Innovation Incubator, Leuphana University Lueneburg, Lueneburg, Germany

³Department of Health Psychology and Applied Biological Psychology, Institute for Psychology, Leuphana University Lueneburg, Lueneburg, Germany

⁴GET.ON Institute for Online Health Trainings, Hamburg, Germany

⁵Department of Clinical, Neuro and Developmental Psychology, Faculty of Behaviour and Movement Sciences, VU University Amsterdam, Amsterdam, Netherlands

Corresponding Author:

Kiona Krueger Weisel, MSc

Department of Clinical Psychology and Psychotherapy

Institute of Psychology

Friedrich-Alexander-University Erlangen-Nuernberg

4th Floor

Nägelsbachstraße 25a

Erlangen, 91052

Germany

Phone: 49 091318567570

Fax: 49 091318567576

Email: kiona.weisel@fau.de

Abstract

Background: Although internet-based and mobile-based stress management interventions (iSMIs) may be a promising strategy to reach employees suffering from high chronic stress, it remains unknown whether participants with high symptom severity of depression or anxiety also benefit from iSMIs or should be excluded.

Objective: This study aimed to evaluate the efficacy of iSMIs in subgroups with high symptom severity and to test whether baseline symptom severity moderates treatment outcome.

Methods: Data from three randomized controlled trials (N=791) were pooled to identify effect modifiers and to evaluate efficacy in subgroups with different levels of initial symptom severity. The outcomes perceived stress (Perceived Stress Scale, PSS), depression severity (Center for Epidemiological Depression Scale, CES-D), and anxiety (Hospital Anxiety and Depression Scale, HADS) symptom severity were assessed at baseline, 7-week postassessment, and 6-month follow-up. Potential moderators were tested in predicting differences in the change of outcome in multiple moderation analyses. Simple slope analyses evaluated efficacy of the iSMI comparing the intervention group with the waitlist control group in subgroups with low, moderate, and severe initial symptomology based on means and SDs of the study population. In addition, subgroups with clinical values of depression (CES-D \geq 16) and anxiety (HADS \geq 8) at baseline were explored, and response rates (RRs; 50% symptom reduction) and symptom-free (SF) status (CES-D $<$ 16, HADS $<$ 8) were reported.

Results: Individuals with high stress (PSS \geq 30), depression (CES-D \geq 33), anxiety (HADS \geq 15), and emotional exhaustion (MBI \geq 5.6) benefited significantly from the intervention with great reductions of stress ($d_{\text{post}}=0.86\text{--}1.16$, $d_{\text{FU}}=0.93\text{--}1.35$), depression ($d_{\text{post}}=0.69\text{--}1.08$, $d_{\text{FU}}=0.91\text{--}1.19$), and anxiety ($d_{\text{post}}=0.79\text{--}1.19$, $d_{\text{FU}}=1.06\text{--}1.21$), and effects were sustained at 6-month follow-up. Symptom severity moderated treatment outcomes, as individuals with higher symptom severity at baseline benefited significantly more from the intervention than individuals with lower symptom severity. Furthermore, 82.9% (656/791) of individuals had

clinical depression values at baseline, of which significantly more individuals in the intervention group reached at least 50% symptom reduction or fell under clinical cut-off (RR: 29.2%, 93/318; SF: 39.6%, 126/318) compared with the waitlist control group (RR: 8.0%, 27/338; SF: 18.6%, 63/338) at postassessment. Significantly more individuals with clinical anxiety values at baseline (HADS \geq 8, 85.3%, 675/791) in the intervention group achieved at least 50% symptom reduction or fell under clinical cut-off (RR: 27.7%, 94/339; SF: 39.8%, 135/339) compared with the WLC (RR: 4.8%, 16/336; SF: 15.5%, 52/336).

Conclusions: Highly burdened individuals benefit greatly from iSMIs and therefore should not be excluded from participation. Stress management may be a valid entry point to reach highly burdened individuals who otherwise may not seek treatment.

Trial Registration: 1) German Clinical Trials Register DRKS00005112; <https://www.drks.de/DRKS00005112> (Archived by WebCite at <http://www.webcitation.org/6zmIZwvdA>); 2) German Clinical Trials Register DRKS00005384; <https://www.drks.de/DRKS00005384> (Archived by WebCite at <http://www.webcitation.org/6zmIerdtr>); and 3) German Clinical Trials Register DRKS00004749; <https://www.drks.de/DRKS00004749> (Archived by WebCite at <http://www.webcitation.org/6zmIjDQPX>).

(*J Med Internet Res* 2018;20(6):e211) doi:[10.2196/jmir.9387](https://doi.org/10.2196/jmir.9387)

KEYWORDS

anxiety; depression; internet; effect modifier; randomized controlled trials; occupational stress

Introduction

Background

High chronic stress is linked to adverse psychological health outcomes. Left untreated, individuals suffering from high occupational stress can develop common mental disorders such as depression or anxiety [1]. Stress, depression, and anxiety are associated with productivity loss and absenteeism, and can negatively affect workplace safety [2,3].

Meta-analytic evidence on occupational interventions of the last two decades, aiming to improve mental health, has found varying evidence of benefit ranging from nonsignificant to moderate effect sizes depending on, for example, type of intervention, intervention content, and outcome categories [4,5]. In a systematic review, Martin et al [6] found overall small positive effects in pooled data of 17 studies, investigating whether different types of health promotion interventions in the workplace reduce depression (standardized mean difference, SMD=0.28, 95% CI 0.12-0.44) and anxiety (SMD=0.29, 95% CI 0.06-0.51). A systematic review and meta-analysis from 2014 identified nine workplace-based randomized controlled trials (RCTs) aimed at reducing the level of depression symptoms. Pooled effect size estimates showed the interventions to be superior to the control groups by a small positive effect (SMD=0.16, 95% CI 0.07-0.24) [7,8]. Richardson et al found a moderate effect size across studies in a meta-analysis of stress management interventions (SMIs) in occupational settings ($d=0.53$, 95% CI 0.36-0.69), significant moderate intergroup effects for anxiety ($d=0.68$), and a small to moderate effect for mental health ($d=0.44$) [9]. Although the range of effect size varies, occupational interventions and SMIs seem to bear the potential to improve psychological health outcomes. However, the previous studies did not examine whether there were differences in efficacy between subgroups with varying initial symptom severity. It therefore remains an open research question whether individuals with high clinical symptomatology, that is, depression and anxiety, also benefit from SMIs.

One specific form of SMIs, which has proven to be effective in certain contexts, is internet-based and mobile-based stress management interventions (iSMIs) [5,10]. Advantages of iSMIs

include the following: (1) individuals can avoid stigmatization by participating anonymously, (2) such trainings are flexible and adaptable to any work and life situation, (3) material can be revised as often as desired, (4) access to treatment and treatment uptake is facilitated by not having waiting times nor a limitation of resource distribution, and (5) fostering self-efficacy of participants.

Occupational iSMIs could reach individuals who would likely not seek psychological treatment [11,12]. This includes severely burdened individuals who, thus, have developed clinical profiles with high symptoms of depression and anxiety. Most individuals with depression and anxiety do not seek treatment, for example, because of fear of stigma [13]. SMIs may bear the potential to attract individuals who would not make use of mental health interventions explicitly labeled for targeting mental health, that is, depression or anxiety. However, it remains unknown whether individuals with high symptom severity also benefit from low threshold iSMIs, as its exploration, to date, is lacking.

It seems plausible that individuals with severe symptomatology, for example, those who experience clinical levels of depression or anxiety, are too burdened to substantially improve their mental health through occupational SMIs not specifically designed to treat depression or anxiety. It may be that methods and techniques delivered in SMIs are not sufficient, as highly burdened participants may need more therapeutic support than generally provided in (i)SMIs. In addition, with regard to iSMIs, some individuals with high symptom severity may be overwhelmed and unable to apply psychological self-help strategies effectively into their daily lives.

Allowing severely affected individuals to participate in (i)SMIs, who are unlikely to benefit, could be problematic. Participation may result in aggravation, hopelessness, and deterioration of symptoms, and may delay, or in the worst case, inhibit affected individuals from seeking appropriate treatment in time, hence contributing to a chronification of symptoms [14]. In addition, this may result in unnecessary intervention delivery costs. Thus, it is crucial to investigate differential effects of SMIs, and to test whether such approaches are also effective in severely affected populations, or whether these individuals should rather

be excluded from SMIs and referred to clinical treatment for psychological disorders.

Through moderation analysis, differential treatment effects can be investigated. However, primary studies are generally solely powered to detect overall treatment effects, and thus underpowered to adequately perform reliable subgroup and moderator analyses [15]. To overcome this issue, data can be pooled from single studies by adding individual participant data in one dataset for common analyses [16].

Objective

This study aims to investigate the effects of an iSMI in subgroups of individuals who experience severe levels of stress, depression, anxiety, emotional exhaustion, or insomnia at baseline, and to test whether these baseline indicators of clinical impairment moderate the intervention efficacy in the reduction of stress, depression, and anxiety. If proven effective for severely burdened individuals, iSMIs could be a crucial component in the amelioration of mental health in occupational health settings, irrespective of the severity level of clinical impairment populations experience.

Methods

GET.ON Stress Intervention

Secondary analyses were conducted based on pooled individual participant data from three RCTs evaluating the same iSMI (GET.ON Stress). The three studies comparing an iSMI to a waitlist control group (WLC) were identical in design, differing only in respect to guidance intensity [17]. The iSMI is based on the Lazarus and Folkman's transactional model of stress [18], and core components focus on problem solving [19,20] and emotion regulation [21,22]. The studies were approved by the ethics committee of the Philipps University Marburg (2013-20K, AZ 2013-35K, AZ 2012-43K) and registered in the German Clinical Trial Register under DRKS00005112, DRKS00005384, and DRKS00004749. Further details on the intervention can be found in a published study protocol [23].

Textbox 1. Inclusion criteria for the study.

- Currently employed
- Aged 18 years or older
- Scores of 22 or above on Perceived Stress Scale [22,24]
- Internet access and a valid email address
- Sufficient reading and writing skills in German
- Willing to give informed consent

Textbox 2. Exclusion criteria for the study.

- If the participants self-reported having been diagnosed with psychosis or dissociative symptoms
- If the participants showed a notable suicidal risk indicated by a score greater than 1 on the Beck suicide item [25]

Sample

Participants of the three studies were recruited via the occupational health program of a large health insurance company in Germany (company website, newspaper articles, and advertisements in the membership magazine), a study website, mass media (newspapers and television), and announcements by the Ministry of Education. Recruitment was open to the general working population and not restricted to members of a certain health insurance company. In total, 791 participants were included. Inclusion and exclusion criteria are provided in Textboxes 1 and 2, respectively.

A standard procedure was followed when individuals showed a notable suicidal risk. They were advised to seek help from their general practitioner, local psychiatric emergency room, or to contact the official emergency number. Telephone numbers and information on relevant institutions were provided via email. The cut-off on the Perceived Stress Scale (PSS) was chosen to include participants with heightened stress based on 1 SD (6.2) above the mean (15.3) found in a large working population [26]. There was no participation cut-off for individuals with severe impairment because of critically high symptom severity of depression or anxiety. Figure 1 shows the flow of participants.

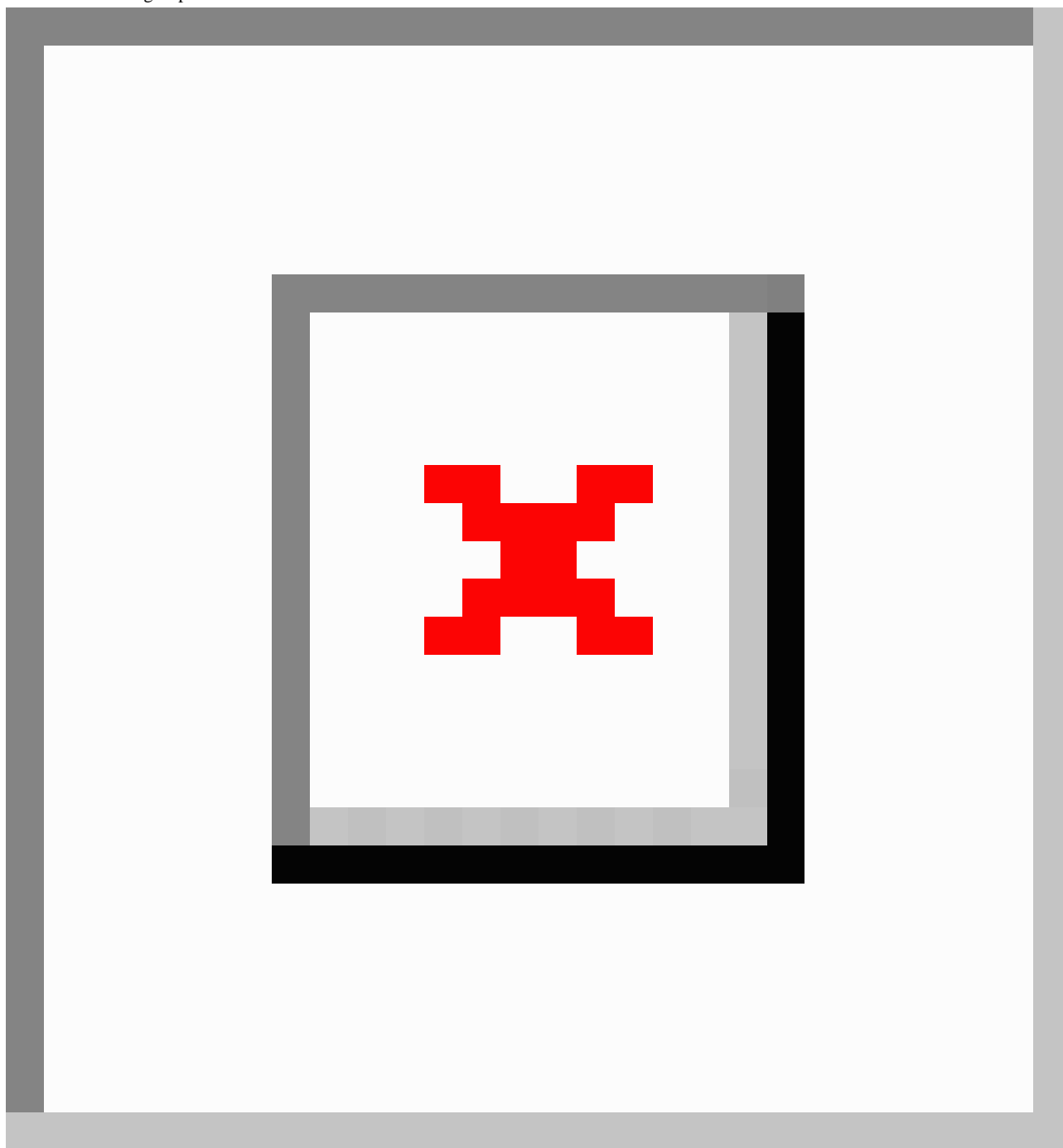
Measures

The outcomes stress, depressive symptom severity, and anxiety symptom severity were collected by self-report at baseline (T1), at postassessment 7 weeks after randomization (T2), and at 6-month follow-up (T3). Furthermore, potential moderators, emotional exhaustion and insomnia severity, were collected at baseline.

Primary Outcome

The primary outcome of the primary studies was perceived stress measured by the German version of the PSS (10 items; score range: 0-4; total score range: 0-40; $\alpha_{T1}=.75$) [22,24]. Higher scores on the PSS indicate more severe perceived stress.

Figure 1. Flow diagram of participants. BDI: Beck Depression Inventory; iSMI: internet-based and mobile-based stress management intervention; WLC: waitlist control group.



Secondary Outcomes

Secondary outcomes included depressive symptom severity measured by the German version of the frequently used Center for Epidemiological Depression Scale (CES-D; 20 items; score range: 0-3; total score range: 0-60; $\alpha_{T1}=.86$) [27]. Scores of CES-D ≥ 16 indicate clinically relevant levels of depression severity [28]. Anxiety symptom severity was measured by the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS; 7 items; score range: 0-3; total score range: 0-21; $\alpha_{T1}=.74$) [29]. A cut-off of ≥ 8 indicates clinically relevant levels of anxiety [30].

Additional mental health assessments included emotional exhaustion, the basic stress dimension of burnout, measured by the German Version Maslach Burnout Inventory (MBI; 5 items; score range: 1-6; total score range 1-6; $\alpha_{T1}=.78$) [31,32], and insomnia severity, assessed by the Insomnia Severity Index (ISI; 7 items; score range: 0-4; total score range 0-28; $\alpha_{T1}=.82$) [33]. Higher scores on all secondary outcome scales indicate more severe or higher symptoms than lower scores.

Statistical Analysis

Missing Data

Treatment outcomes included perceived stress, depressive symptom severity, and anxiety symptoms. The dropout rate ranged from 7.5% (59/791) in the primary outcome stress at postassessment to 13.8% (109/791) at 6-month follow-up in the outcome insomnia severity. Overall, these are low rates of intervention dropout. Comparing the dropout rates between the iSMI and the WLC, it becomes evident that the outcome dropout rates are higher in the intervention condition (12.4%-21%) compared with the WLC (2.5%-6.6%).

Missing data of all outcomes were handled with multiple imputations (MIs) in accordance with the intention-to-treat principle [34,35]. Prerequisites of MI include that the data are missing at random and that the data are normally distributed. Using Little's missing completely at random (MCAR) test, it was determined that the missing data were MCAR. The test was not significant ($\chi^2_{50}=40.2$, $P=.84$). However, the outcome data were not normally distributed, which can potentially lead to implausible low or negative values after using MI. To verify post hoc that this did not affect the imputed data, it was tested whether outliers in the new dataset corresponded with missing data in the original dataset, which was not the case.

Missing data were handled by using the MI procedure in SPSS. Estimations of missing data were based on all available data for all outcome measures at all assessment points (T1, T2, and T3), as well as age, gender, and intervention condition (iSMI and WLC). In addition, 100 imputed datasets were created and aggregated into 1 dataset for the final analysis. Significance level was set to $<.05$.

Multiple Moderation Analyses

To identify subgroups with different levels of initial symptom severity and to assess the efficacy of the intervention for populations with severe impairment on changes in outcomes (stress, depressive, and anxiety symptoms), multiple moderation analyses (MMAs) were performed with the SPSS macro PROCESS [36]. MMAs are separate multiple regression models each containing 3 elements that are tested in 3 steps predicting the outcome: (1) main effect of the baseline variable, (2) main effect of the treatment condition (iSMI and WLC), and (3) interaction effect (baseline \times treatment condition). If the interaction effect is significant in predicting the outcome, it is considered to be an effect modifier moderating the outcome. Effect modifiers of treatment outcome indicate that populations benefit significantly different from each other depending on initial symptom severity. Moderators of the difference in change of the outcomes will be reported for postassessment and 6-month follow-up.

To explore the direction of an interaction effect, improve interpretability, and provide specific estimations for the investigated subgroups of interest, follow-up simple slope analyses were performed. Baseline data on perceived stress, depression, anxiety, emotional exhaustion, and insomnia severity were split into 3 subgroups based on distributional values, mean (M), and SD of the study population (mean-1SD; mean;

mean+1SD), and outcomes in each subgroup were explored. To compare efficacy of the iSMI and WLC in the respected subgroups at postassessment and 6-month follow-up, Cohen d , 95% CI, and the number needed to treat (NNT) were calculated. Baseline differences were controlled for by using the difference in changes of outcome from baseline to postassessment (T1-T2) and from baseline to 6-month follow-up (T1-T3) as dependent variable. The study variable was not included in the final analysis, as it was not associated with the primary outcome stress at 6-month follow-up. Cohen d can be interpreted based on 3 approximations: small effects ($d=0.2$), medium effects ($d=0.5$), and large effects ($d=0.8$) [37]. All continuous variables were standardized for interpretability.

Response Rates and Symptom-Free Status in Clinical Subgroups

In addition, effects in subgroups with clinical values of depression (CES-D ≥ 16) and anxiety (HADS ≥ 8) at baseline were explored. Results were reported as mean between-group differences, Cohen d , and NNT while controlling for baseline severity by using change scores from baseline to postassessment. Improvements at an individual level were investigated by assessing the number of participants who achieved a response rate (RR) through 50% symptom reduction and symptom-free (SF) status (CES-D <16 , HADS <8). Differences between the intervention condition and control group were tested in a chi-square (χ^2) test.

Results

Descriptive Statistics

Participants were aged 42.6 years (SD 9.7) on average at baseline, and the majority were female (76.7%, 607/791). Overall, the sample was highly educated with a minimum of 12 years of schooling and a degree (71.9%, 569/791) or mid-level education with 10 years of schooling and a degree (25.4%, 201/791), with only 2.5% (21/791) of persons having low education levels with no school degree or 9 years of schooling. Participants had an average value of 25.6 (SD 4.3) on the PSS. The mean of depressive symptomology (CES-D) was 23.9 (SD 8.6). Anxiety symptoms (HADS) were at an average of 11.1 (SD 3.4). The average of emotional exhaustion (MBI) was 4.7 (SD 0.7). Insomnia severity (ISI) was at a mean of 13.7 (SD 6.1). Most participants had clinically relevant levels of depression (82.9%, 656/791) and anxiety (85.3%, 675/791) at study uptake. Further details on baseline characteristics can be found in Table 1.

The classification in subgroups with low, moderate, and severe initial symptom severity is based on means and SDs of the study population. Means and SDs for all investigated subgroups (mean-1SD; mean; mean+1SD) with regard to all potential predictors (perceived stress [PSS], depressive symptoms [CES-D], anxiety [HADS], emotional exhaustion [MBI], and insomnia severity [ISI]), and outcomes (PSS, CES-D, and HADS) can be found in Table 2. All subgroups showed a reduction of symptom severity in favor of the iSMI in comparison with the WLC, and effects were sustained at 6-month follow-up.

Primary Outcome

Effect sizes for differences in change of perceived stress at postassessment were medium to large. The smallest effect was observed for individuals with low emotional exhaustion at baseline ($d=0.57$; 95% CI 0.22-0.91; NNT=3.18), and greatest effects were observed for those with high emotional exhaustion ($d=1.16$; 95% CI 0.78-1.54; NNT=1.7). Moderators of change in the outcome stress at postassessment were stress ($\beta=1.04$, $P=.01$) and emotional exhaustion ($\beta=1.12$, $P=.008$). Highly stressed individuals and individuals with high emotional exhaustion profited significantly more from the iSMI than individuals with low symptom severity (low: $PSS_{T1-T2}d=0.65$, $MBI_{T1-T2}d=0.57$; high: $PSS_{T1-T2}d=0.98$, $MBI_{T1-T2}d=1.16$). [Figure](#)

[2](#) shows the estimated course of symptom change in the iSMI compared with the WLC at postassessment and 6-month follow-up.

Looking at the effects at 6-month follow-up, moderate to large effects were found in change of stress, the smallest effect for stress ($d=0.66$; 95% CI 0.3-1.02; NNT=2.78) and the largest effect for emotional exhaustion ($d=1.35$; 95% CI 0.96-1.75; NNT=1.52). Stress ($\beta=0.96$, $P=.049$) was the only moderator for change in stress at 6-month follow-up. Highly stressed individuals showed greater reduction in stress through the intervention than low-stressed individuals (low: $PSS_{T1-T3}d=0.66$, high: $PSS_{T1-T3}d=1.27$).

Table 1. Baseline characteristics of the study population.

Characteristics	All (N=791)	iSMI ^a (n=395)	WLC ^b (n=396)
Age in years, mean (SD)	42.6 (9.7)	42.1 (9.9)	43.1 (9.5)
Gender, n (%)^c			
Female	607 (76.7)	307 (77.5)	300 (75.9)
Male	181 (22.9)	87 (22)	94 (23.8)
Other	2 (0.3)	1 (0.3)	1 (0.3)
Ethnicity, n (%)			
Caucasian or white	657 (83.1)	326 (82.3)	331 (83.8)
Other or no information	134 (16.9)	70 (17.7)	64 (16.2)
Marital status, n (%)			
Unmarried	230 (29.1)	122 (30.8)	108 (27.3)
Married	376 (47.5)	191 (48.2)	185 (46.8)
Cohabited	95 (12)	51 (12.9)	44 (11.1)
Separated or divorced	82 (10.4)	30 (7.6)	52 (13.2)
Widowed	8 (1)	2 (0.5)	6 (1.5)
Education, n (%)			
Low	21 (2.7)	11 (2.8)	10 (2.5)
Middle	201 (25.4)	101 (25.5)	100 (25.3)
High	569 (71.9)	284 (71.7)	285 (72.2)
Employment status, n (%)			
Permanent	652 (82.4)	322 (81.3)	330 (83.5)
Temporary	77 (9.7)	44 (11.1)	33 (8.4)
Self-employed	50 (6.3)	26 (6.6)	24 (6.1)
Other	12 (1.5)	4 (1)	8 (2)
Employment situation, n (%)			
Full-time	601 (76)	302 (76.3)	299 (75.7)
Part-time	180 (22.8)	88 (22.2)	92 (23.2)
Sick leave	10 (1.3)	6 (1.5)	4 (1)
Occupational sectors, n (%)			
Social, education	208 (26.3)	95 (24)	113 (28.6)
Service provision	156 (19.7)	85 (21.5)	71 (18)
Finance, administration	129 (16.3)	63 (15.9)	66 (16.7)
Health	110 (13.9)	63 (15.9)	47 (11.9)
Information technology, computer	48 (6.1)	26 (6.6)	22 (5.6)
Media	23 (2.9)	8 (2)	15 (3.8)
Natural sciences	22 (2.8)	8 (2)	14 (3.5)
Metal, engineering	18 (2.3)	10 (2.5)	8 (2)
Production, manufacture	15 (1.9)	7 (1.8)	8 (2)
Construction, architecture	13 (1.6)	3 (0.8)	10 (2.5)
Social sciences, liberal arts	11 (1.4)	6 (1.5)	5 (1.3)
Art, culture, and design	10 (1.3)	7 (1.8)	3 (0.8)
Infrastructure, logistics	10 (1.3)	4 (1)	6 (1.5)

Characteristics	All (N=791)	iSMI ^a (n=395)	WLC ^b (n=396)
Technology	7 (0.9)	3 (0.8)	4 (1)
Agriculture, environment	7 (0.9)	5 (1.3)	2 (0.5)
Electro	4 (0.5)	3 (0.8)	1 (0.3)
Gross annual income (in Euro), n (%)			
Low (<30,000)	207 (26.2)	116 (29.3)	91 (23)
Middle (30,000-50,000)	200 (25.3)	99 (25)	101 (25.6)
High (>50,000)	309 (39.1)	153 (38.6)	156 (39.5)
Not reported	75 (9.5)	28 (7.1)	47 (11.9)
Experience with health-related programs, n (%)			
Yes	107 (13.5)	49 (12.4)	58 (14.7)
No	684 (86.5)	347 (87.6)	377 (85.3)
Experience with face-to-face psychotherapy, n (%)			
Yes	311 (39.3)	151 (38.1)	160 (40.5)
No	480 (60.7)	245 (61.9)	235 (59.5)
PSS ^d , mean (SD)	25.6 (4.3)	25.6 (4.5)	25.5 (4.1)
CES-D ^e , mean (SD) ^f	23.9 (8.6)	23.9 (9)	24 (8.1)
CES-D ^e ≥16, n (%)	656 (82.9)	318 (40.2)	338 (42.7)
HADS ^g , mean (SD)	11.1 (3.4)	11.1 (3.4)	11.0 (3.4)
HADS ^g ≥8, n (%)	675 (85.3)	339 (42.9)	336 (42.5)
MBI ^h , mean (SD)	4.7 (0.7)	4.7 (0.7)	4.74 (0.7)
ISI ⁱ , mean (SD) ^f	13.7 (6.1)	13.9 (6.1)	13.6 (6.1)

^aiSMI: internet-based and mobile-based stress management interventions.

^bWLC: waitlist control group.

^cDue to missing data, the incidences refer to a sample of n=790.

^dPSS: Perceived Stress Scale.

^eCES-D: Center for Epidemiological Depression Scale.

^fBaseline data were imputed as intention-to-treat population values were used later in the analysis.

^gHADS: Hospital Anxiety and Depression Scale.

^hMBI: Maslach Burnout Inventory.

ⁱISI: Insomnia Severity Index.

Table 3 shows the effects on change in the outcomes stress, depressive symptoms, and anxiety from baseline to postassessment for the 3 subgroups divided by baseline psychopathology (mean-1SD; mean; mean+1SD) of stress, depression, anxiety, emotional exhaustion, and insomnia severity. The table also displays the *P* values of the standardized regression coefficient of the interaction effect between group (iSMI and WLC) and potential moderators. *P* values below .05 indicate that symptom severity of a certain characteristic is a moderator of treatment outcome.

Secondary Outcomes

Effect sizes for differences in change of depression at postassessment were small to large. The smallest effect was

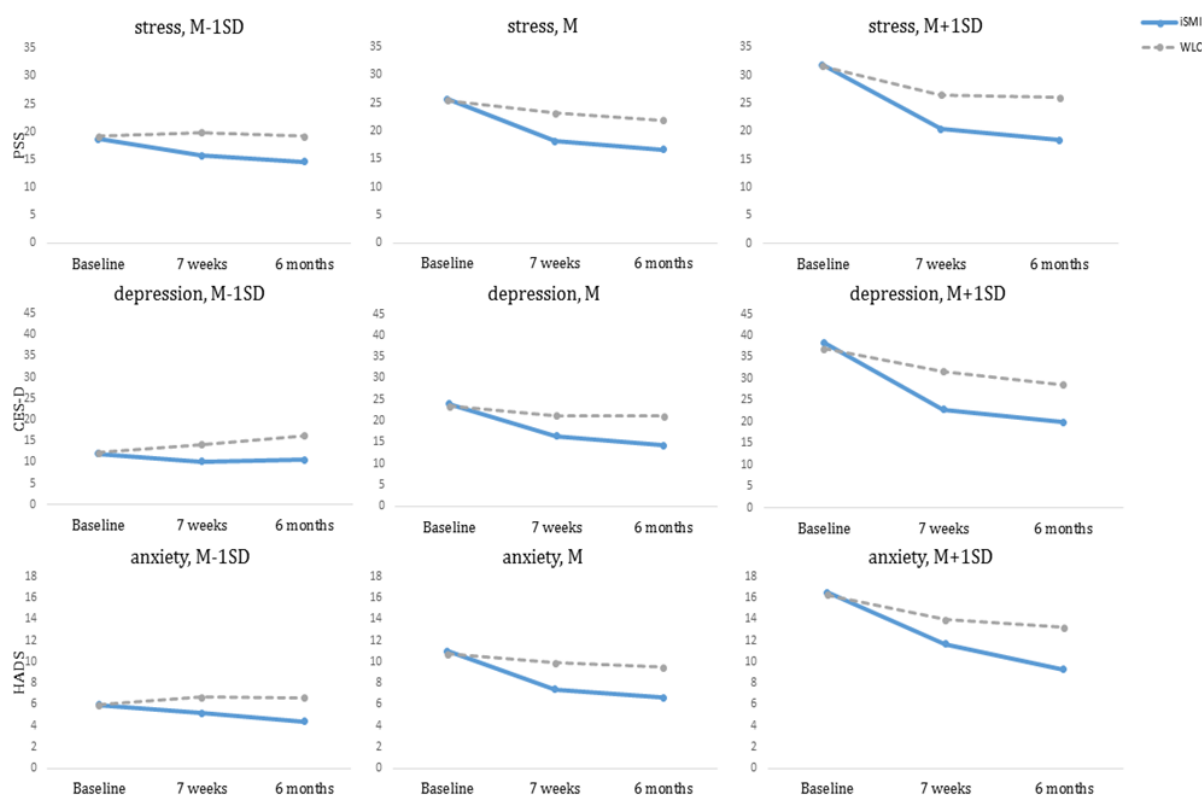
observed in participants with low emotional exhaustion at baseline ($d=0.37$; 95% CI 0.03-0.71; NNT=4.85) and the largest effects for those with high levels of depression severity ($d=1.08$; 95% CI 0.7-1.46; NNT=1.81). Moderators of change in the outcome depression at postassessment were depression ($\beta=1.89$, $P=.001$), stress ($\beta=2.1$, $P<.001$), and emotional exhaustion ($\beta=2.28$, $P<.001$). Participants with initially high depressive symptoms, high perceived stress, and high emotional exhaustion benefitted significantly more from the intervention in reducing depressive symptoms than individuals with lower severity (low: CES-D_{T1-T2} $d=0.67$, PSS_{T1-T2} $d=0.49$, MBI_{T1-T2} $d=0.37$; high: CES-D_{T1-T2} $d=1.08$, PSS_{T1-T2} $d=0.92$, MBI_{T1-T2} $d=1.01$).

Table 2. Overview of internet-based and mobile-based stress management interventions (iSMI) and waitlist control group (WLC) symptom severity of subgroups at baseline (T1), postassessment (T2), and 6-month follow-up (T3).

Symptom severity	Mean-1SD		Mean		Mean+1SD	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
Perceived Stress Scale						
T1						
iSMI ^a	68	18.7 (2.6)	255	25.7 (2.2)	73	31.8 (1.9)
WLC ^b	60	19.2 (2.2)	268	25.4 (2)	67	31.7 (1.9)
T2						
iSMI	68	15.7 (5.4)	255	18.2 (5.4)	73	20.4 (6.9)
WLC	60	19.9 (5.2)	268	23.2 (5.3)	67	26.5 (5.8)
T3						
iSMI	68	14.6 (5.3)	255	16.6 (6)	73	18.4 (6.6)
WLC	60	19.2 (6.4)	268	21.9 (5.8)	67	26 (6.3)
Center for Epidemiological Depression Scale						
T1						
iSMI	78	11.9 (3.1)	255	23.9 (4.7)	63	38.4 (5)
WLC	57	12.2 (2.4)	267	23.5 (4.6)	62	37 (3.9)
T2						
iSMI	78	10.1 (6.2)	255	16.5 (7.4)	63	22.9 (10.9)
WLC	57	14.1 (5.4)	267	21.2 (7.7)	62	31.7 (8.6)
T3						
iSMI	78	10.5 (6.8)	255	14.3 (6.9)	63	19.8 (10)
WLC	57	16.3 (7.3)	267	21.1 (8.7)	62	28.6 (10.7)
Hospital Anxiety and Depression Scale						
T1						
iSMI	57	5.9 (1.2)	279	11 (1.9)	60	16.5 (1.6)
WLC	59	5.9 (1)	264	10.7 (1.8)	72	16.3 (1.4)
T2						
iSMI	57	5.2 (2.5)	279	7.4 (3.1)	60	11.7 (3.9)
WLC	59	6.7 (2.5)	264	9.9 (2.9)	72	14 (3)
T3						
iSMI	57	4.4 (2.3)	279	6.6 (3.1)	60	9.3 (3.7)
WLC	59	6.6 (3.2)	264	9.5 (3.4)	72	13.3 (3.6)
Maslach Burnout Inventory						
T1						
iSMI	74	3.6 (0.4)	258	4.8 (0.4)	64	5.8 (0.2)
WLC	62	3.6 (0.5)	273	4.8 (0.4)	60	5.8 (0.2)
Insomnia Severity Index						
T1						
iSMI	62	4.4 (2.1)	261	13.7 (3.3)	69	22.7 (2.6)
WLC	71	4.4 (2.1)	249	13.8 (3.4)	74	22.3 (2.1)

^aiSMI: internet-based and mobile-based stress management interventions.^bWLC: waitlist control group.

Figure 2. Simple slope: illustrative course of estimated course of symptom change in waitlist control group and internet-based and mobile-based stress management interventions for significant moderator “perceived stress” on stress, “depression severity” on depression, and “anxiety severity” on anxiety between baseline, postassessment, and 6-month follow-up. M: mean; PSS: Perceived Stress Scale; CES-D: Center for Epidemiological Depression Scale; HADS: Hospital Anxiety and Depression Scale; iSMI: internet-based and mobile-based stress management interventions; WLC: waitlist control group.



Effects sizes for differences in change of depression at 6-month follow-up were moderate to large. The smallest effect was found for participants with initially low levels of stress ($d=0.6$; 95% CI 0.24-0.95; NNT=3.05) and the largest effect for participants with high levels of emotional exhaustion ($d=1.19$; 95% CI 0.81-1.58; NNT=1.67). Differences in change of depressive symptoms from baseline to 6-month follow-up were moderated by stress ($\beta=2.28$, $P<.001$), emotional exhaustion ($\beta=1.91$, $P=.002$), and depression ($\beta=1.41$, $P=.025$). Participants with higher levels of stress, emotional exhaustion, and depressive symptoms benefitted more in reducing depressive symptoms (high: $\text{CES-D}_{T1-T3}d=1.02$, $\text{PSS}_{T1-T3}d=1.2$, $\text{MBI}_{T1-T3}d=1.19$) than participants with lower levels (low: $\text{CES-D}_{T1-T3}d=0.79$, $\text{PSS}_{T1-T3}d=0.6$, $\text{MBI}_{T1-T3}d=0.72$). Figure 2 illustrates the simple slope analysis for the moderators stress, depressive, and anxiety symptoms on their respective symptom severity.

Medium to large effect sizes were found for changes in anxiety at postassessment, with the smallest effect for participants with low emotional exhaustion ($d=0.56$; 95% CI 0.22-0.91; NNT=3.25) and the largest effect for participants with high levels of emotional exhaustion at baseline ($d=1.19$; 95% CI 0.81-1.57; NNT=1.67). Moderators of change in the outcome anxiety at postassessment were stress ($\beta=0.47$, $P=.036$) and

emotional exhaustion ($\beta=0.65$, $P=.002$). Individuals with initially higher levels of emotional exhaustion showed significantly higher effectiveness in reducing anxiety (low: $\text{MBI}_{T1-T2}d=0.56$, high: $\text{MBI}_{T1-T2}d=1.19$). Individuals with low stress benefitted slightly more than individuals with moderate or high stress in reducing anxiety; however, all effects are considered large (low: $\text{PSS}_{T1-T2}d=0.88$, moderate: $\text{PSS}_{T1-T2}d=0.8$, high: $\text{PSS}_{T1-T2}d=0.81$).

Effects in change of anxiety were moderate to large at 6-month follow-up; participants with low emotional exhaustion at baseline showed smaller effects ($d=0.7$; 95% CI 0.36-1.05; NNT=2.63) than participants with high emotional exhaustion ($d=1.21$; 95% CI 0.83-1.6; NNT=1.64). Anxiety was moderated by stress ($\beta=0.52$, $P=.034$), emotional exhaustion ($\beta=0.52$, $P=.027$), and anxiety ($\beta=0.59$, $P=.013$); individuals with higher initial symptom severity stress, anxiety, and emotional exhaustion were profiting more from the intervention (low: $\text{PSS}_{T1-T3}d=0.99$; $\text{HADS}_{T1-T3}d=0.76$, $\text{MBI}_{T1-T3}d=0.7$; high: $\text{PSS}_{T1-T3}d=1.17$, $\text{HADS}_{T1-T3}d=1.16$, $\text{MBI}_{T1-T3}d=1.21$). With regard to differences in change from baseline to 6-month follow-up, effects in all investigated subgroups of different initial symptom severity were significant, as shown in Table 4, and in favor of the iSMI compared with the WLC.

Table 3. Effects on change in perceived stress, depressive symptoms, and anxiety from baseline to postassessment for different subgroups of baseline psychopathology severity (Mean–1SD, Mean, and Mean+1SD).

Moderators at postassessment	Mean–1SD			Mean		Mean+1SD	
	<i>P</i> ^a value	<i>d</i> ^b (95% CI)	NNT ^c	<i>d</i> ^b (95% CI)	NNT ^c	<i>d</i> ^b (95% CI)	NNT ^c
Outcome: stress							
MOD_Stress ^d (PSS) ^e	.01	0.65 (0.29-1)	2.82	0.98 (0.8-1.16)	1.95	0.98 (0.63-1.33)	1.95
MOD_Depression ^f (CES-D) ^g	.11	0.82 (0.47-1.18)	2.28	0.85 (0.67-1.03)	2.21	1.05 (0.67-1.42)	1.85
MOD_Anxiety ^h (HADS) ⁱ	.61	0.79 (0.42-1.17)	2.36	0.86 (0.69-1.04)	2.19	0.86 (0.5-1.22)	2.19
MOD_EmotExhaustion ^j (MBI) ^k	.008	0.57 (0.22-0.91)	3.18	0.86 (0.69-1.04)	2.19	1.16 (0.78-1.54)	1.7
MOD_Insomnia ^l (ISI) ^m	.87	1.14 (0.77-1.51)	1.72	0.8 (0.62-0.98)	2.34	0.88 (0.53-1.22)	2.15
Outcome: depression							
MOD_Stress (PSS)	<.001	0.49 (0.14-0.84)	3.68	0.68 (0.51-0.86)	2.7	0.92 (0.57-1.26)	2.07
MOD_Depression (CES-D)	.001	0.67 (0.32-1.02)	2.75	0.7 (0.53-0.88)	2.63	1.08 (0.7-1.46)	1.81
MOD_Anxiety (HADS)	.38	0.76 (0.39-1.14)	2.44	0.67 (0.5-0.85)	2.75	0.69 (0.33-1.04)	2.67
MOD_EmotExhaustion (MBI)	<.001	0.37 (0.03-0.71)	4.85	0.66 (0.49-0.84)	2.78	1.01 (0.63-1.38)	1.91
MOD_Insomnia (ISI)	.11	0.66 (0.31-1.01)	2.78	0.64 (0.47-0.82)	2.86	0.82 (0.48-1.16)	2.28
Outcome: anxiety							
MOD_Stress (PSS)	.04	0.88 (0.52-1.25)	2.15	0.8 (0.62-0.98)	2.34	0.81 (0.47-1.16)	2.3
MOD_Depression (CES-D)	.55	1.04 (0.68-1.4)	1.86	0.75 (0.57-0.93)	2.48	0.89 (0.52-1.26)	2.13
MOD_Anxiety (HADS)	.09	0.58 (0.21-0.95)	3.14	0.93 (0.76-1.11)	2.04	0.79 (0.43-1.14)	2.36
MOD_EmotExhaustion (MBI)	.002	0.56 (0.22-0.91)	3.25	0.77 (0.6-0.95)	2.42	1.19 (0.81-1.57)	1.67
MOD_Insomnia (ISI)	.28	0.98 (0.62-1.34)	1.95	0.69 (0.51-0.87)	2.67	1.12 (0.77-1.47)	1.75

^a*P* values of the standardized regression coefficient of the interaction effect between the potential moderator and group (iSMI and WLC).^bCohen *d*.^cNNT: number needed to treat.^dMOD_Stress: moderator stress.^ePSS: Perceived Stress Scale.^fMOD_Depression: moderator depression.^gCES-D: Center for Epidemiological Depression Scale.^hMOD_Anxiety: moderator anxiety.ⁱHADS: Hospital Anxiety and Depression Scale.^jMOD_EmotExhaustion: moderator emotional exhaustion.^kMBI: Maslach Burnout Inventory.^lMOD_Insomnia: moderator insomnia severity.^mISI: Insomnia Severity Index.

Response Rates and Symptom-Free Status in Clinical Subgroups

Furthermore, subgroups were investigated in which individuals had values indicating clinical depression levels at baseline (CES-D \geq 16, 82.9%, 656/791, mean 26.36, SD 7.19). At postassessment, the iSMI group (n=318) had a mean CES-D value of mean 17.76 (SD 8.69) and a CES-D change score of $M_{T1-T2}=9.05$ (SD_{T1-T2} 8.79) compared with the WLC (n=338, mean 23.1, SD 8.83, $M_{T1-T2}=2.85$, SD_{T1-T2} 7.43). This resulted

in a moderate to large effect of Cohen $d=0.76$ (95% CI 0.6-0.92) and an NNT of 2.44. On the basis of the CES-D, a score reduction of 50% from baseline to postassessment was achieved significantly more often in participants of the iSMI (CES-D: 29.2%, 93/318) as compared with the WLC (CES-D: 8.0%, 27/338; $\chi^2_1=49.5$; $P<.001$; NNT=4.7), and significantly more individuals in the iSMI compared with the WLC met criteria for SF status (CES-D<16) at postassessment (iSMI: CES-D: 39.6%, 126/318; WLC: 18.6%, 63/338; $\chi^2_1=35.2$; $P<.001$; NNT=4.8).

Table 4. Effects on change in perceived stress, depressive symptoms, and anxiety from baseline to 6-month follow-up for different subgroups of baseline psychopathology severity (mean-1SD, mean, and mean+1SD).

Moderators at 6-month follow-up	Mean-1SD			Mean		Mean+1SD	
	<i>P</i> ^a value	<i>d</i> ^b (95% CI)	NNT ^c	<i>d</i> ^b (95% CI)	NNT ^c	<i>d</i> ^b (95% CI)	NNT ^c
Outcome: stress							
MOD_Stress ^d (PSS) ^e	.049	0.66 (0.3-1.02)	2.78	0.91 (0.73-1.09)	2.08	1.27 (0.9-1.63)	1.59
MOD_Depression ^f (CES-D) ^g	.12	0.7 (0.35-1.05)	2.63	0.92 (0.74-1.1)	2.07	1.05 (0.67-1.42)	1.85
MOD_Anxiety ^h (HADS) ⁱ	.55	0.91 (0.53-1.29)	2.08	0.85 (0.67-1.03)	2.21	0.93 (0.57-1.29)	2.04
MOD_EmotExhaustion ^j (MBI) ^k	.05	0.8 (0.45-1.15)	2.34	0.78 (0.61-0.96)	2.39	1.35 (0.96-1.75)	1.52
MOD_Insomnia ^l (ISI) ^m	.53	1.02 (0.65-1.38)	1.89	0.79 (0.61-0.97)	2.36	1.05 (0.7-1.4)	1.85
Outcome: depression							
MOD_Stress (PSS)	.001	0.6 (0.24-0.95)	3.05	0.77 (0.59-0.94)	2.42	1.2 (0.84-1.56)	1.66
MOD_Depression (CES-D)	.025	0.79 (0.44-1.15)	2.36	0.94 (0.76-1.12)	2.02	1.02 (0.65-1.39)	1.89
MOD_Anxiety (HADS)	.33	0.9 (0.52-1.28)	2.1	0.77 (0.6-0.95)	2.42	0.91 (0.55-1.27)	2.08
MOD_EmotExhaustion (MBI)	.002	0.72 (0.37-1.07)	2.56	0.73 (0.56-0.91)	2.54	1.19 (0.81-1.58)	1.67
MOD_Insomnia (ISI)	.11	0.78 (0.43-1.14)	2.39	0.77 (0.59-0.95)	2.42	0.86 (0.52-1.21)	2.19
Outcome: anxiety							
MOD_Stress (PSS)	.034	0.99 (0.63-1.36)	1.94	0.84 (0.66-1.02)	2.23	1.17 (0.81-1.53)	1.69
MOD_Depression (CES-D)	.36	1.11 (0.75-1.48)	1.76	0.87 (0.69-1.05)	2.16	1.06 (0.69-1.44)	1.83
MOD_Anxiety (HADS)	.013	0.76 (0.38-1.13)	2.44	1.01 (0.84-1.19)	1.91	1.16 (0.78-1.53)	1.7
MOD_EmotExhaustion (MBI)	.027	0.7 (0.36-1.05)	2.63	0.89 (0.71-1.07)	2.13	1.21 (0.83-1.6)	1.64
MOD_Insomnia (ISI)	.42	1.04 (0.68-1.4)	1.86	0.82 (0.64-1)	2.28	1.06 (0.71-1.41)	1.83

^a*P* values of the standardized regression coefficient of the interaction effect between the potential moderator and group (iSMI and WLC).^bCohen *d*.^cNNT: number needed to treat.^dMOD_Stress: moderator stress.^ePSS: Perceived Stress Scale.^fMOD_Depression: moderator depression.^gCES-D: Center for Epidemiological Depression Scale.^hMOD_Anxiety: moderator anxiety.ⁱHADS: Hospital Anxiety and Depression Scale.^jMOD_EmotExhaustion: moderator emotional exhaustion.^kMBI: Maslach Burnout Inventory.^lMOD_Insomnia: moderator insomnia severity.^mISI: Insomnia Severity Index.

Individuals with clinically relevant levels of anxiety at baseline (HADS \geq 8, 85.3%, 675/791, mean 11.94, SD 2.82) showed a large effect of Cohen $d=0.87$ (95% CI 0.72-1.03) in favor of the iSMI at postassessment. Individuals in the iSMI ($n=339$) had a mean HADS value of mean 8.15 (SD 3.65) and HADS change score $M_{T1-T2}=3.81$ (SD_{T1-T2} 3.3) compared with the WLC ($n=336$, mean 10.8, SD 3.38, $M_{T1-T2}=1.13$, SD_{T1-T2} 2.81). Treatment response, that is, symptom reduction of 50% in the HADS, was assessed significantly more often in the iSMI (27.7%, 94/339) compared with the WLC (4.8%, 16/336; $\chi^2_1=65.2$; $P<.001$; NNT=4.4). SF status (HADS<8) was assessed significantly more often in the iSMI (39.8%, 135/339) compared with the WLC (15.5%, 52/336; $\chi^2_1=49.9$; $P<.001$; NNT=4).

Discussion

Principal Findings

This study aimed to explore whether iSMIs are effective in severely burdened employees and tested whether baseline indicators of impairment moderated treatment outcome. Highly burdened participants who showed high levels of stress, depression, anxiety, emotional exhaustion, or insomnia severity profited substantially from the intervention on all outcome measures with moderate to large intergroup effect sizes compared with the control condition, both at postassessment and at 6-month follow-up. Moreover, higher impairment (ie, depression, anxiety, perceived stress, and emotional exhaustion)

was associated with greater symptom improvement over time. These findings are in line with (1) studies showing that internet-based self-help interventions can be effective in clinical populations [38-40], (2) studies showing that internet-based occupational health interventions specifically designed to target depression can be effective [41,42], and (3) moderator analyses showing that higher symptom severity is not associated with worse treatment outcome in low-threshold self-help interventions [19,43]. This study extends these findings by showing that highly burdened participants with high levels of stress, depression, anxiety, insomnia severity, or emotional exhaustion can also substantially profit from an intervention that is labelled and specifically designed to reduce negative consequences of occupational stress. Reasons may include that the iSMI contains techniques based on cognitive behavioral therapy (ie, problem-solving and emotion regulation), which have shown to be effective in treating depression and anxiety [21,44].

Many of the participants were first-time help-seekers, indicated by 86.5% having no prior experience with health-related training and less than half (39.3%) having prior experience with face-to-face therapy, although the majority of participants had already reached clinically relevant levels of depression (82.9%) and anxiety (85.3%) before study uptake. This indicates that stress management may be a valid entry point to reach highly burdened individuals who otherwise may not seek treatment. Stress is less stigmatized than depression and anxiety, and online treatment provides the necessary anonymity for uncertain individuals to initially seek help.

Individuals of subgroups with high symptom severity at baseline (CES-D \geq 33; HADS \geq 15; PSS \geq 30; ISI \geq 20; MBI \geq 5.6) experienced significant reductions with large intergroup effect sizes, which were also sustained over time (6-month follow-up: depression, d [0.86-1.19]; anxiety d [1.06-1.21], stress d [0.93-1.27]). Although significantly more individuals in the iSMI intervention group were assessed to have reached a RR of at least 50% symptom reduction (CES-D: 29.2%, 93/318; HADS: 27.7%, 94/339) and even SF status (CES-D: 39.6%, 126/318; HADS: 39.8%, 135/339), many highly burdened individuals, however, did not achieve either. This indicates that iSMIs cannot fully substitute more intensive psychological treatment for all individuals with high levels of anxiety or depression. Such findings highlight the importance of monitoring individual progress throughout treatment to refer individuals to further, more intensive or simply different treatment modalities, based on different theoretical constructs or treatment formats after intervention completion. Another possibility may be to tailor the treatment to the individual symptom profile if the participant does not respond sufficiently to the standardized treatment. However, future studies are needed to explore if such approaches indeed lead to better treatment outcome, as effects were already large in terms of effect sizes, and it is also possible that there is a limit to what can be achieved with psychological interventions.

Limitations

The study also has some limitations. First, as common for randomized trials, there was an elaborate screening process at study entry. This may have caused individuals with lower self-efficacy and less motivation to dropout before study uptake, and those most likely to profit to continue. Second, considering the description of the study population, the variance was rather low, reducing the heterogeneity of variance to explore differences within subgroups of the study sample. Third, subgroups were created based on the mean and 1 SD above or below the mean, which is typical for discerning simple slopes. Ranges are therefore based on this study population and may vary for other populations. Fourth, overall participants had elevated symptoms of impairment because of the study inclusion criterion heightened stress (PSS \geq 22). Therefore, there was no exploration of individuals with very low symptom severity. Fifth, all exploration of subgroups was based on the same iSMI GET.ON Stress. Therefore, the findings may not be generalizable directly to other interventions, especially as the latest meta-analysis on iSMIs indicated high heterogeneity between studies and interventions [5]. Finally, the main finding of this study that individuals with severe impairment profit greatly from the iSMI cannot be generalized to individuals who show indications of suicidality, as the studies excluded individuals with high suicide risk at baseline. Unfortunately, this remains an unresolved issue many internet interventions face, how to adequately deal with at-risk individuals who show active interest in participation but whose monitoring throughout a low-threshold intervention, often, cannot be guaranteed.

Strengths of this study include the strong methodology of RCTs, and the pooling of individual participant data from different studies investigating the same intervention, hereby overcoming the issue of studies generally being underpowered to explore subgroups [16].

Conclusions

This study contains important implications for research and practice. First, populations who experience high stress, are clinically depressed, or have high anxiety can alleviate their disease burden and reduce symptom severity by participating in iSMIs. Second, occupational stress and stress management are effective entry points for initial contact with mental health interventions, and therefore, severely burdened individuals should be targeted for participation. Third, future studies should explore whether tailoring of iSMIs to clinical profiles leads to superior outcomes in severely affected individuals compared with standardized approaches. Finally, as employers benefit from healthy employees, they should consider offering iSMIs in routine occupational settings and encourage participation. Severely burdened employees, especially those with high levels of depression and anxiety, should not be excluded but rather motivated to participate in iSMIs.

Acknowledgments

The authors would like to thank the participants without whom the study would not have been possible. The Barmer and European Union funded the primary studies (EU EFRE: ZW6-80119999, CCI 2007DE161PR001).

Authors' Contributions

DDE and KKW contributed to the design of the study. DDE, DL, EH, HR, and MB developed the intervention content. KKW conducted the outcome analyses. KKW, under the supervision of DDE, drafted the first proof of the manuscript and integrated coauthor comments and edits. All authors contributed to the further writing of the manuscript and approved the final manuscript.

Conflicts of Interest

DDE, MB, and DL are stakeholders of the “GET.ON Institute for Online Health Trainings,” which aims to transfer scientific knowledge related to this field of research into routine health care. EH is working at the “GET.ON Institute for Online Health Trainings.”

References

1. Stansfeld S, Candy B. Psychosocial work environment and mental health-a meta-analytic review. *Scand J Work Environ Health* 2006;32(6):443-462 [FREE Full text] [Medline: 17173201]
2. Lerner D, Henke RM. What does research tell us about depression, job performance, and work productivity? *J Occup Environ Med* 2008;50(4):401-410. [doi: 10.1097/JOM.0b013e31816bae50] [Medline: 18404013]
3. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
4. Bhui KS, Dinos S, Stansfeld SA, White PD. A synthesis of the evidence for managing stress at work: a review of the reviews reporting on anxiety, depression, and absenteeism. *J Environ Public Health* 2012;2012:515874 [FREE Full text] [doi: 10.1155/2012/515874] [Medline: 22496705]
5. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. *J Med Internet Res* 2017;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
6. Martin A, Sanderson K, Cocker F. Meta-analysis of the effects of health promotion intervention in the workplace on depression and anxiety symptoms. *Scand J Work Environ Health* 2008;35(1):7-18. [doi: 10.5271/sjweh.1295]
7. Tan L, Wang MJ, Modini M, Joyce S, Mykletun A, Christensen H, et al. Preventing the development of depression at work: a systematic review and meta-analysis of universal interventions in the workplace. *BMC Med* 2014;12:74 [FREE Full text] [doi: 10.1186/1741-7015-12-74] [Medline: 24886246]
8. Tan L, Wang MJ, Modini M, Joyce S, Mykletun A, Christensen H, et al. Erratum to: preventing the development of depression at work: a systematic review and meta-analysis of universal interventions in the workplace. *BMC Med* 2014 Nov 13;12:212 [FREE Full text] [doi: 10.1186/s12916-014-0212-4]
9. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. *J Occup Health Psychol* 2008;13(1):69-93. [Medline: 18211170]
10. Zetterqvist K, Maanmies J, Ström L, Andersson G. Randomized controlled trial of internet-based stress management. *Cogn Behav Ther* 2003;32(3):151-160. [doi: 10.1080/16506070302316] [Medline: 16291546]
11. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. *J Med Internet Res* 2016;18(1):e21. [Medline: 26818683]
12. Marks IM, Cavanagh K, Gega L. Computer-aided psychotherapy: revolution or bubble? *Br J Psychiatry* 2007;191:471-473 [FREE Full text] [doi: 10.1192/bjp.bp.107.041152] [Medline: 18055948]
13. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychol Med* 2014;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
14. 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;46(13):2679-2693. [doi: 10.1017/S0033291716001562.Does]
15. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004 Mar;57(3):229-236. [doi: 10.1016/j.jclinepi.2003.08.009] [Medline: 15066682]
16. Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychol Methods* 2009;14(2):165-176. [doi: 10.1037/a0015565] [Medline: 19485627]
17. Zarski AC, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. *J Med Internet Res* 2016;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]

18. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer; 1984.
19. Junge MN, Lehr D, Bockting CL, Berking M, Riper H, Cuijpers P, et al. For whom are internet-based occupational mental health interventions effective? Moderators of internet-based problem-solving training outcome. *Internet Interventions* 2015;2(1):39-47. [doi: [10.1016/j.invent.2014.11.007](https://doi.org/10.1016/j.invent.2014.11.007)]
20. Warmerdam L, van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. *J Med Internet Res* 2008;10(4):e44 [FREE Full text] [doi: [10.2196/jmir.1094](https://doi.org/10.2196/jmir.1094)] [Medline: [19033149](https://pubmed.ncbi.nlm.nih.gov/19033149/)]
21. Berking M, Ebert D, Cuijpers P, Hofmann SG. Emotion regulation skills training enhances the efficacy of inpatient cognitive behavioral therapy for major depressive disorder: a randomized controlled trial. *Psychother Psychosom* 2013;82(4):234-245. [doi: [10.1159/000348448](https://doi.org/10.1159/000348448)] [Medline: [23712210](https://pubmed.ncbi.nlm.nih.gov/23712210/)]
22. Michalsen A, Jeitler M, Brunnhuber S, Lüdtke R, Büssing A, Musial F, et al. Iyengar yoga for distressed women: a 3-armed randomized controlled trial. *Evid Based Complement Alternat Med* 2012;2012:408727 [FREE Full text] [doi: [10.1155/2012/408727](https://doi.org/10.1155/2012/408727)] [Medline: [23049608](https://pubmed.ncbi.nlm.nih.gov/23049608/)]
23. Heber E, Ebert DD, Lehr D, Nobis S, Berking M, Riper H. Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial. *BMC Public Health* 2013;13:655 [FREE Full text] [doi: [10.1186/1471-2458-13-655](https://doi.org/10.1186/1471-2458-13-655)] [Medline: [23855376](https://pubmed.ncbi.nlm.nih.gov/23855376/)]
24. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24(4):385-396. [Medline: [6668417](https://pubmed.ncbi.nlm.nih.gov/6668417/)]
25. Beck AT, Steer RA, Brown GK. BDI-II, Beck Depression Inventory: Manual. San Antonio, TX: Psychological Corporation; 1996:490-498.
26. Lesage FX, Berjot S, Deschamps F. Psychometric properties of the French versions of the Perceived Stress Scale. *Int J Occup Med Environ Health* 2012;25(2):178-184 [FREE Full text] [doi: [10.2478/S13382-012-0024-8](https://doi.org/10.2478/S13382-012-0024-8)] [Medline: [22528542](https://pubmed.ncbi.nlm.nih.gov/22528542/)]
27. Meyer TD, Hautzinger M. Allgemeine Depressions-Skala (ADS). *Diagnostica* 2001;47(4):208-215 [FREE Full text] [doi: [10.1026//0012-1924.47.4.208](https://doi.org/10.1026//0012-1924.47.4.208)]
28. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging* 1997;12(2):277-287. [Medline: [9189988](https://pubmed.ncbi.nlm.nih.gov/9189988/)]
29. Hermann-Lingen C, Buss U, Snaith RP. Hospital Anxiety and Depression Scale Deutsche Version (HADS-D). In: Wirtz MA, editor. *Lexikon der Psychologie*. Bern: Verlag Hans Huber; 2014.
30. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002;52(2):69-77. [Medline: [11832252](https://pubmed.ncbi.nlm.nih.gov/11832252/)]
31. Büssing A, Glaser J. Managerial Stress und Burnout: a Collaborative International Study (CISMS); die deutsche Untersuchung. Munich: Technische Universität Lehrstuhl für Psychologie (München); 1998.
32. Leiter MP, Schaufeli W. Consistency of the burnout construct across occupations. *Anxiety Stress Coping* 1996;9(3):229-243. [doi: [10.1080/10615809608249404](https://doi.org/10.1080/10615809608249404)]
33. Morin CM. *Insomnia: Psychological Assessment and Management*. New York, NY: Guilford Press; 1993.
34. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 2002;7(2):147-177. [Medline: [12090408](https://pubmed.ncbi.nlm.nih.gov/12090408/)]
35. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Br Med J* 2009;338:b2393. [Medline: [19564179](https://pubmed.ncbi.nlm.nih.gov/19564179/)]
36. Hayes AF. *Introduction to Mediation, Moderation and Conditional Process Analysis a Regression-Based Approach (Methodology in the Social Sciences)*. New York, NY: The Guilford Press; 2013.
37. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic Press; 1977.
38. Ebert DD, Zarski AC, 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;10(3):e0119895 [FREE Full text] [doi: [10.1371/journal.pone.0119895](https://doi.org/10.1371/journal.pone.0119895)] [Medline: [25786025](https://pubmed.ncbi.nlm.nih.gov/25786025/)]
39. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin Psychol Rev* 2012;32(4):329-342. [doi: [10.1016/j.cpr.2012.02.004](https://doi.org/10.1016/j.cpr.2012.02.004)] [Medline: [22466510](https://pubmed.ncbi.nlm.nih.gov/22466510/)]
40. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. *J Affect Disord* 2017;223:28-40. [doi: [10.1016/j.jad.2017.07.021](https://doi.org/10.1016/j.jad.2017.07.021)] [Medline: [28715726](https://pubmed.ncbi.nlm.nih.gov/28715726/)]
41. Ebert DD, Lehr D, Boß L, Riper H, Cuijpers P, Andersson G, et al. Efficacy of an internet-based problem-solving training for teachers: results of a randomized controlled trial. *Scand J Work Environ Health* 2014 Nov;40(6):582-596. [doi: [10.5271/sjweh.3449](https://doi.org/10.5271/sjweh.3449)] [Medline: [25121986](https://pubmed.ncbi.nlm.nih.gov/25121986/)]
42. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Does Internet-based cognitive behavioral therapy (iCBT) prevent major depressive episode for workers? A 12-month follow-up of a randomized controlled trial. *Psychol Med* 2015;45(9):1907-1917. [doi: [10.1017/S0033291714003006](https://doi.org/10.1017/S0033291714003006)] [Medline: [25562115](https://pubmed.ncbi.nlm.nih.gov/25562115/)]

43. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *Br Med J* 2013;346:f540 [FREE Full text] [Medline: 23444423]
44. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev* 2006;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]

Abbreviations

CES-D: Center for Epidemiological Depression Scale
HADS: Hospital Anxiety and Depression Scale
ISI: Insomnia Severity Index
iSMI: internet-based and mobile-based stress management interventions
MBI: Maslach Burnout Inventory
MCAR: missing completely at random
MI: multiple imputations
MMAs: multiple moderation analyses
NNT: number needed to treat
PSS: Perceived Stress Scale
RCT: randomized controlled trial
RR: response rate
SF: symptom-free
SMD: standardized mean difference
WLC: waitlist control group

Edited by G Eysenbach; submitted 11.11.17; peer-reviewed by JH Baek, P Valente; comments to author 04.01.18; revised version received 11.02.18; accepted 25.03.18; published 19.06.18.

Please cite as:

Weisel KK, Lehr D, Heber E, Zarski AC, Berking M, Riper H, Ebert DD
Severely Burdened Individuals Do Not Need to Be Excluded From Internet-Based and Mobile-Based Stress Management: Effect Modifiers of Treatment Outcomes From Three Randomized Controlled Trials
J Med Internet Res 2018;20(6):e211
URL: <http://www.jmir.org/2018/6/e211/>
doi: [10.2196/jmir.9387](https://doi.org/10.2196/jmir.9387)
PMID: [29921562](https://pubmed.ncbi.nlm.nih.gov/29921562/)

©Kiona Krueger Weisel, Dirk Lehr, Elena Heber, Anna-Carlotta Zarski, Matthias Berking, Heleen Riper, David Daniel Ebert. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 19.06.2018. 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.

Original Paper

The Effectiveness of Web-Based Tailored Smoking Cessation Interventions on the Quitting Process (Project Quit): Secondary Analysis of a Randomized Controlled Trial

Bibhas Chakraborty^{1,2,3}, PhD; Raju Maiti¹, PhD; Victor J Strecher⁴, PhD

¹Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Singapore

²Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore

³Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, United States

⁴Department of Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor, MI, United States

Corresponding Author:

Bibhas Chakraborty, PhD

Centre for Quantitative Medicine

Duke-NUS Medical School

Academia, Level 6

20 College Road

Singapore, 169856

Singapore

Phone: 65 6576 7377

Fax: 65 6225 1244

Email: bibhas.chakraborty@duke-nus.edu.sg

Abstract

Background: Project Quit was a randomized Web-based smoking cessation trial designed and conducted by researchers from the University of Michigan, where its primary outcome was the 7-day point prevalence. One drawback of such an outcome is that it only focuses on smoking behavior over a very short duration, rather than the quitting process over the entire study period.

Objective: The aim of this study was to consider the number of quit attempts during the 6-month study period as an alternative outcome, which would better reflect the quitting process. We aimed to find out whether tailored interventions (high vs low) are better in reducing the number of quit attempts for specific subgroups of smokers.

Methods: To identify interactions between intervention components of smoking cessation and individual smoker characteristics, we employed Poisson regression to analyze the number of quit attempts. This approach allowed us to construct data-driven, personalized interventions.

Results: A negative effect of the number of cigarettes smoked per day ($P=.03$) and a positive effect of education ($P=.03$) on the number of quit attempts were detected from the baseline covariates ($n=792$). Thus, for every 10 extra cigarettes smoked per day, there was a 5.84% decrease in the expected number of quit attempts. Highly educated participants had a 15.49% increase in their expected number of quit attempts compared with their low-educated counterparts. A negative interaction between intervention component *story* and smoker's education was also detected ($P=.03$), suggesting that a high-tailored story given to highly educated people results in 13.50% decrease in the number of quit attempts compared with a low-tailored story.

Conclusions: A highly individually tailored story is significantly more effective for smokers with a low level of education. This is consistent with prior findings from Project Quit based on the 7-day point prevalence.

(*J Med Internet Res* 2018;20(6):e213) doi:[10.2196/jmir.9555](https://doi.org/10.2196/jmir.9555)

KEYWORDS

smoking cessation; number of quit attempts; tailored intervention; treatment regimen; Web-based intervention

Introduction

Smoking is the leading preventable cause of death worldwide [1] and is associated with substantial economic burden [2]. Decades of research efforts have focused on evaluating effective computer-tailored smoking-cessation intervention programs [3,4]. These tailored programs are increasingly being delivered via technology-enabled platforms, for example, the internet [5,6], or more recently through mobile phone apps [7]. However, even with the support of modern technology, developing tailored smoking-cessation interventions is burdensome for both patients and health care providers. Hence, from a precision medicine perspective, it would be interesting to stratify a subgroup of smokers likely to benefit from tailored interventions. In this study, we used data from *Project Quit* [8,9], a randomized trial using Web-based tailored smoking cessation program and conducted secondary data analysis to identify this subgroup.

In modern quantitative precision medicine literature, the idea of personalizing treatments to individual patients is often operationalized as a *treatment regimen* [10-12]. Treatment regimen (TR) is a decision rule that takes available patient information as inputs to recommend some treatment. Constructing evidence-based (ie, data-driven) TR is typically a 2-step process consisting of hypothesis-generating data analysis and conducting a confirmatory trial [13]. An optimal TR estimated from existing data can be used to generate hypotheses on how an individual's case history should guide treatment selection. These hypotheses can then be tested against a suitable control in a randomized controlled trial. Estimated optimal TR from Project Quit data analysis suggests that tailored interventions are most beneficial for smokers with low education and potentially detrimental to those with high education. Project Quit data analysis also suggests that tailored interventions do not have any impact on smokers with very high level of baseline addiction (those smoking >20 cigarettes/day).

Point prevalence is often a popular choice in assessing smoking cessation. In fact, Project Quit study was designed with a 7-day point prevalence as the primary outcome [8]. However, this outcome is based on subjects' smoking status in a very limited time window (last 7 days) rather than the entire study period. An alternative outcome that better reflects the quitting process is the number of quit attempts over the entire study period [14-18]. It reflects participants' involvement, or lack thereof, in the smoking cessation program. Although Project Quit study has collected data on number of quit attempts as a secondary outcome, this information has not yet been analyzed. Thus, we are focusing on this outcome in this study.

This study aims to identify a subgroup of smokers who are most likely to benefit from Web-based tailored behavioral interventions for smoking cessation. We will identify this subgroup based on their willingness and involvement in the quitting process, measured by the number of quit attempts during the study period. Such an approach will potentially allow health care researchers to use the limited public health resources more efficiently in shaping the health care policy.

Methods

Project Quit Trial

Project Quit was a Web-based smoking cessation program developed and conducted by the Center for Health Communications Research at the University of Michigan, Ann Arbor, and was funded by the National Cancer Institute (NCI), USA. The study protocol was reviewed and approved by the Institutional Review Board of each collaborating institution and of the University of Michigan in January 2004. The primary aim of the study was to identify and test the effects of 5 psychosocial and communication intervention components influencing smoking cessation [8]. The content of the Web-based intervention was based on cognitive-behavioral methods of smoking cessation, including an appeal to motives for quitting, stimulus control, self-efficacy enhancement, and suggestion for coping with tempting situations and emotions. Hypothetical success stories were employed within this overall paradigm.

Five intervention components (outcome expectations, efficacy expectations, success stories, message source, and message exposure) were studied in Project Quit. To screen multiple components, the study employed a *multiphase optimization strategy* [19] framework, implemented using a 16-cell (2^{5-1}) fractional factorial design [20,21] in which each of the 5 intervention components were varied at 2 levels, high vs low. However, only 2 components, namely, success stories (hereafter referred to as *story*) and message source (hereafter referred to as *source*), were found to have significant effects on smoking cessation in previous analysis [8]. These 2 components were a priori hypothesized to have the strongest effect on smoking cessation. On the basis of these findings, we considered these 2 intervention components in our analysis. The intervention component *story* refers to success story received by study subjects from a hypothetical character who succeeded in quitting smoking. The *story* was varied at 2 levels—high vs low tailoring depth (ie, the degree to which the character in the story was tailored to subject's baseline characteristics). Similarly, the component *source* refers to the source of Web-based behavioral counseling message received by subjects and was varied at 2 levels—high versus low level—of personalization. High-personalized source included photograph and supportive text from the health maintenance organization's (HMO) smoking cessation team. It was written in a friendly language and included a signature from the team. In contrast, the low-personalized version included a photograph of a building, representing the HMO, and was impersonally written without a closing signature. Strecher et al [8] provided detailed description of these components, including examples of actual Web-based messages.

Adult participants were recruited from 2 HMOs—Group Health Cooperative (GHC), Seattle; and Henry Ford Health System (HFHS), Detroit; both these HMOs were affiliated with the NCI's cancer research network. The study participants had a broad representation of ethnicity, gender, age, health status, and geography. Participants' eligibility criteria included those who (1) had smoked at least 100 cigarettes in his or her lifetime, currently smoked at least 10 cigarettes per day, and had smoked

in the past 7 days; (2) were seriously considering quitting in the next 30 days; (3) were 21 to 70 years; (4) were members of either GHC or HFHS; (5) had home or work internet access and an email account that they used at least twice weekly; (6) were not currently enrolled in other smoking cessation program(s) and not currently using pharmacotherapy for smoking cessation; and (7) had no medical contraindications for nicotine replacement therapy. A total of 1866 subjects participated in Project Quit. All subjects received free Web-based experimental smoking cessation program. Participants were randomized to receive either high or low personalized intervention, as described above. To pharmacologically assist them with smoking cessation, all participants, irrespective of their intervention group, also received a free 10-week supply of nicotine replacement therapy patches. Thus, this study allowed participants to focus on the cognitive-behavioral aspects of smoking cessation through combination of various intervention components.

The primary outcome of the study was the binary 7-day point prevalence in smoking cessation at 6 months following baseline assessment. During the 6-month evaluation survey, each subject was asked if she or he had smoked any cigarettes, even a puff, in the last 7 days. Subject who answered “yes” was marked as smoker and nonsmoker otherwise. In addition, data on the number of quit attempts in the past 6 months were collected as a secondary outcome, which is the focus of this study.

In addition to baseline covariates (age, gender, and race), Project Quit also collected variables deemed relevant for smoking cessation. These included (1) number of cigarettes smoked per day as a measure of baseline addiction; (2) the participant’s level of motivation to quit smoking as a predictor of smoking cessation [8]; (3) the participant’s level of education, which was hypothesized to interact with the intervention component *story* [8]; and (4) participant’s self-efficacy, a consistent predictor of subsequent health-related behavior change based on the social cognitive theory [22].

Data Analysis

Of the 1866 subjects who enrolled in the Project Quit study, 1192 subjects responded to the question on the number of quit attempts. Of these responders, 792 subjects followed the study protocol by not using other smoking cessation aids or programs during the study. As the primary examination in Project Quit [8] utilized *per-protocol analysis* that only included subjects who did not violate study protocol, we used the same strategy to analyze the number of quit attempts in the 792 subjects.

To assess the potential presence of differential missingness across the intervention arms, we conducted chi-square test with 2 categorical variables—intervention arm (4 levels resulting from 2 intervention components, each varied at 2 levels) and nonresponse (2 levels, yes or no).

Baseline covariates considered in this analysis were age (continuous), gender (binary), race (3 levels, but handled by 2 dummy variables—race white and race black), cigarettes smoked per day (continuous), motivation (binary, high vs low, coded 1 or 0), self-efficacy (binary, high vs low, coded 1 or 0), and education (binary, \leq high school vs $>$ high school, coded 0 or 1).

The source and story levels were coded as 1 (high) and 0 (low), respectively.

We used the Poisson regression model to analyze the number of quit attempts. The Poisson regression model can be applied to settings where the outcome is a count-type variable with its expectation (mean) varying as a log-linear function of the covariates and intervention components. The model used in this analysis can be specified as $\log(E(Y|X_1, X_2, \dots, X_8, A_1, A_2)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_8 X_8 + (\delta_0 + \delta_1 X_7)A_1 + (\eta_0 + \eta_1 X_8)A_2$, where Y denotes the number of quit attempts during the 6-month study period; $X_i, i=1, 2, \dots, 8$, denote the baseline characteristics, viz, age, gender, race white, race black, cigarettes smoked per day, motivation, education, and self-efficacy, respectively; and A_1 and A_2 denote the intervention components story and source, respectively. The notation $E(Y|X_1, X_2, \dots, X_8, A_1, A_2)$ denotes the conditional expectation (conditional mean) of Y , given all the baseline covariates and intervention components. The unknown parameters ($\beta_i, i=0, 1, \dots, 8; \delta_0, \delta_1, \eta_0, \eta_1$) in Poisson regression are estimated by the maximum likelihood method. We used open-source software R, version 3.2.3 [23] for the analysis.

Regression coefficients $\beta_i, i=1, \dots, 8$, denote the main effects of the covariates $X_i, i=1, 2, \dots, 8$; β_0 denotes the model intercept; δ_0 and η_0 denote the main effects of intervention components A_1 and A_2 , respectively; and finally, δ_1 and η_1 denote the preconceived interaction effect between X_7 and A_1 and that between X_8 and A_2 , respectively. Instead of reporting the estimates of β_i , we reported the corresponding adjusted incidence rate ratios, or simply the rate ratios (RRs). These quantities offer a more interpretable way to report results from a Poisson regression model (analogous to reporting odds ratios from a logistic regression model for binary data). Under the above setup, we defined RR for a covariate X_i as the ratio of the expectation of Y given that $X_i=1$ and the expectation of Y given that $X_i=0$ (for binary X_i) or as the ratio of the expectation of Y given that $X_i=x+1$ and the expectation of Y given that $X_i=x$ for some arbitrary value x (for continuous X_i), given that other variables in the model (both covariates and interventions) are fixed. This RR can then be computed as the exponential transform of the regression coefficient ($\exp(\beta_i)$). RR measures change in the expected outcome when X_i increases by 1 unit (for continuous X_i), or when X_i moves from 1 category to the other (for categorical X_i) on a multiplicative scale.

Re-expression of intervention effects may further facilitate interpretation. The effect of a particular intervention component, say A_1 (story), can be expressed as $E(Y|X_1, X_2, \dots, X_8, A_1=1, A_2) - E(Y|X_1, X_2, \dots, X_8, A_1=0, A_2) = (\exp(\delta_0 + \delta_1 X_7) - 1) E(Y|X_1, X_2, \dots, X_8, A_1=0, A_2)$, which in turn can be interpreted as—given all other covariates are fixed, a highly tailored story ($A_1=1$) increases the expected number of quit attempts by $(\exp(\delta_0 + \delta_1 X_7) - 1)100\%$ compared with the low-tailored story ($A_1=0$). Similarly, for A_2 (source), it can be interpreted that a highly personalized source increases the expected number of quit attempts by $(\exp(\eta_0 + \eta_1 X_8) - 1)100\%$, compared with the

low-personalized source. Furthermore, for any of the baseline covariates, the effect of the i -th covariate can be expressed as $(\exp(\beta_i) - 1)100\%$, $i=1, \dots, 8$.

We used the standard 5% alpha level to assess statistical significance in our analyses. Whenever appropriate, we also reported the 95% CIs of various effects. On the basis of the Poisson regression results, we then derived the corresponding TRs for recommending personalized smoking cessation interventions and drew decision trees to visually represent TRs.

We expected smokers' baseline level of addiction, as measured by the number of cigarettes smoked per day (and found in our analysis results presented below), to influence the number of quit attempts. Therefore, once the Poisson regression analysis on the full data was completed, we divided the participants into 2 subgroups: (1) those who used to smoke less than or equal to the observed median of the number of cigarettes smoked per day and (2) those who used to smoke more than the observed median of the number of cigarettes smoked per day. We then repeated the Poisson regression analysis for each of the subgroups.

Results

Before the primary data analyses, we examined potential differential missingness across the intervention arms and found no significant difference ($P=.64$).

Descriptive Data Summary

The per-protocol participants' baseline characteristics ($n=792$) are summarized in Table 1. These subjects had a mean age of 46.32 (SD 10.64) years. Of these, 60.6% (480/792) were female, 77.7% (615/792) were white, 12.3% (97/792) were African Americans, 63.4% (502/792) were highly educated, 53.2% (421/792) had high self-efficacy, and 45.7% (362/792) were highly motivated. On average, the participants used to smoke 21.51 (SD 8.94) cigarettes per day at baseline. With respect to randomized interventions, 51.4% (407/792) subjects received a highly tailored story, and 49.9% (395/792) subjects received a highly personalized source. During the 6-month study period, 93.3% (739/792) participants attempted to quit at least once. The number of quit attempts varied from 0 to 10, and the mean quit attempt was 2.74 (SD 2.50) times.

Table 1. Participant characteristics. Descriptive summary refers to mean (SD) for continuous characteristics and frequency (percentage) for categorical variables.

Participant characteristics	Descriptive summary (n=792)
Age in years, mean (SD)	46.32 (10.64)
Gender, n (%)	
Female	480 (60.6)
Race, n (%)	
African American	97 (12.3)
White	615 (77.7)
Other	80 (10.1)
Education, n (%)	
>High school	502 (63.4)
≤High school	290 (36.6)
Number of cigarettes smoked per day, mean (SD)	21.51 (8.94)
Motivation, n (%)	
High	362 (45.7)
Low	430 (54.3)
Self-efficacy, n (%)	
High	421 (53.2)
Low	371 (46.8)
Story, n (%)	
Deeply tailored	407 (51.4)
Low-tailored	385 (48.6)
Source, n (%)	
Highly personalized	395 (49.9)
Low-personalized	397 (50.1)

Poisson Regression Results

The estimated Poisson regression coefficients, z-scores, RR values along with their 95% CIs, and corresponding *P* values are reported in Table 2. After adjusting for relevant covariates and treatment components, only the number of cigarettes smoked per day at baseline (RR=0.994; 95% CI 0.989-0.999; *P*=.03), education (RR=1.155; 95% CI 1.018-1.311; *P*=.03), and education-by-story interaction (RR=0.825; 95% CI 0.692-0.985; *P*=.03) were significant. This means that when all other covariates are fixed in the model, for every extra cigarette smoked per day at baseline, the expected number of quit attempts in the 6-month study period changes by a multiplicative factor of 0.994. In other words, for every 10 extra cigarettes smoked per day at baseline, there is a 5.84% decrease in the expected number of quit attempts over 6 months. On the other hand, there is a 15.49% increment in the expected number of quit attempts for highly educated participants (*P*=.03), as compared with those with low education.

Interaction between education and story is interpreted differently from the main effects of individual covariates. For this scenario, the main effect and interaction effect should be interpreted jointly. Table 3 shows the effects of both high- and low-tailored stories on the 2 levels of education. Giving a high-tailored story to a highly educated smoker results in a 13.5% decrease in the number of quit attempts compared with a low-tailored story (*P*=.03). However, the result is completely reverse for the low-educated group, where a high-tailored story given to a

low-educated person increases the expected number of quit attempts by 4.8% compared with a low-tailored story. On the basis of this result, we derived a TR that recommends personalized smoking cessation interventions (high-tailored story for low-educated subjects and low-tailored story for those who are highly educated). A decision tree to visualize this TR is shown in Figure 1.

We have shown that smokers' baseline level of addiction, as measured by the number of cigarettes smoked per day, has a negative impact on the number of quit attempts. Using the observed median of 20 cigarettes smoked/day as a threshold, we further divided the participants into 2 subgroups: (1) those who used to smoke ≤20 cigarettes/day and (2) those who used to smoke >20 cigarettes/day. We found that severe smokers (>20 cigarettes/day at baseline) were not influenced by any intervention components. However, less severe smokers with lower education were more influenced by the highly tailored story, which is similar to the whole group of smokers in the study. Results from the Poisson regression analyses for the less severe subgroup of smokers (≤20 cigarettes/day at baseline) are shown in Tables 4 and 5, and can be interpreted in a similar fashion as above.

In addition to the effects that were significant in the full dataset, the main effects of self-efficacy and the intervention component story also came out significant in this subgroup analysis. Figure 2 shows the decision tree for the associated TR.

Table 2. Summary results of the Poisson regression model for the number of quit attempts outcome (n=792).

Variable	Regression parameter estimate	Z-score	Adjusted rate ratio (95% CI)	<i>P</i> value
Age (years)	0.001	0.662	1.001 (0.997-1.005)	.51
Gender (male)	0.070	1.572	1.073 (0.983-1.171)	.12
Race (dummy for white)	−0.041	−0.577	0.959 (0.833-1.105)	.56
Race (dummy for Black)	0.134	1.504	1.143 (0.960-1.361)	.13
Number of cigarettes smoked per day (NCigs per day)	−0.006	−2.185	0.994 (0.989-0.999)	.03 ^a
Motivation	0.055	1.148	1.056 (0.962-1.160)	.25
Education	0.144	2.230	1.155 (1.018-1.311)	.03 ^a
Self-efficacy	−0.054	−0.840	0.948 (0.836-1.074)	.40
Story	0.047	0.643	1.048 (0.908-1.209)	.52
Source	−0.082	−1.285	0.921 (0.813-1.044)	.20
Story × education	−0.192	−2.125	0.825 (0.692-0.985)	.03 ^a
Source × self-efficacy	0.080	0.926	1.084 (0.914-1.284)	.36

^aDenotes *P*<.05.

Table 3. Estimated intervention effect of story, expressed as a percentage change in the expected number of quit attempts, stratified by education level, mathematically expressed as $(\exp(\delta_0 + \delta_1 \text{ Edu}) - 1)100\%$ (n=792).

Education	Estimate (95% CI)
>High school (Edu=1)	−13.50 (−22.58 to −4.42)
≤High school (Edu=0)	4.798 (−10.17 to 19.76)

Figure 1. An estimated treatment regimen to recommend personalized smoking cessation intervention for the whole population. Edu: education; HS: high school.

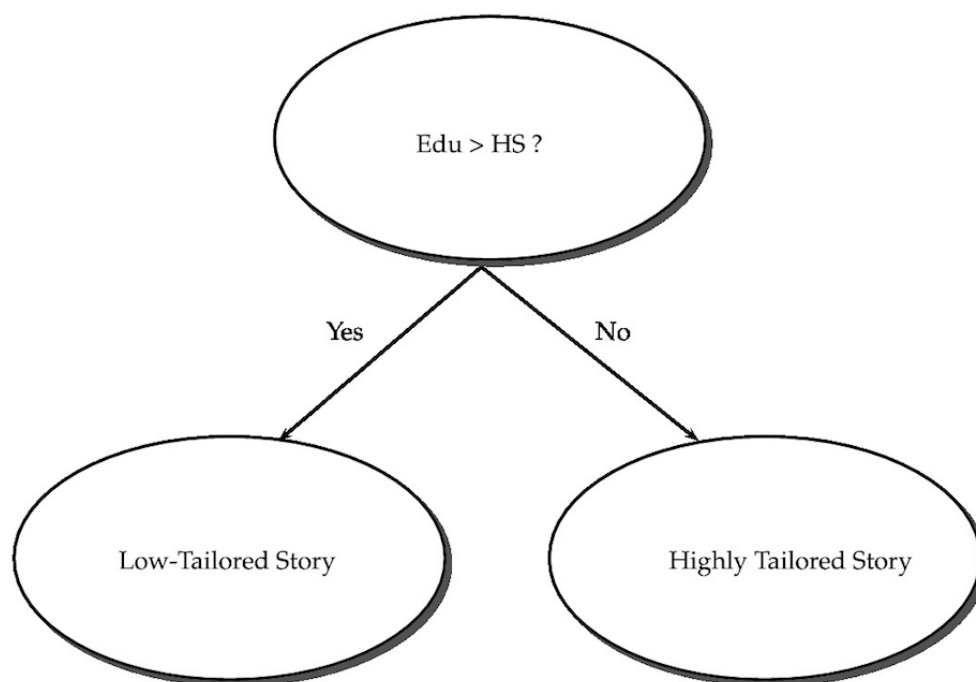


Table 4. Summary results of the Poisson regression model for participants who smoked less than or equal to 20 cigarettes per day (n=546).

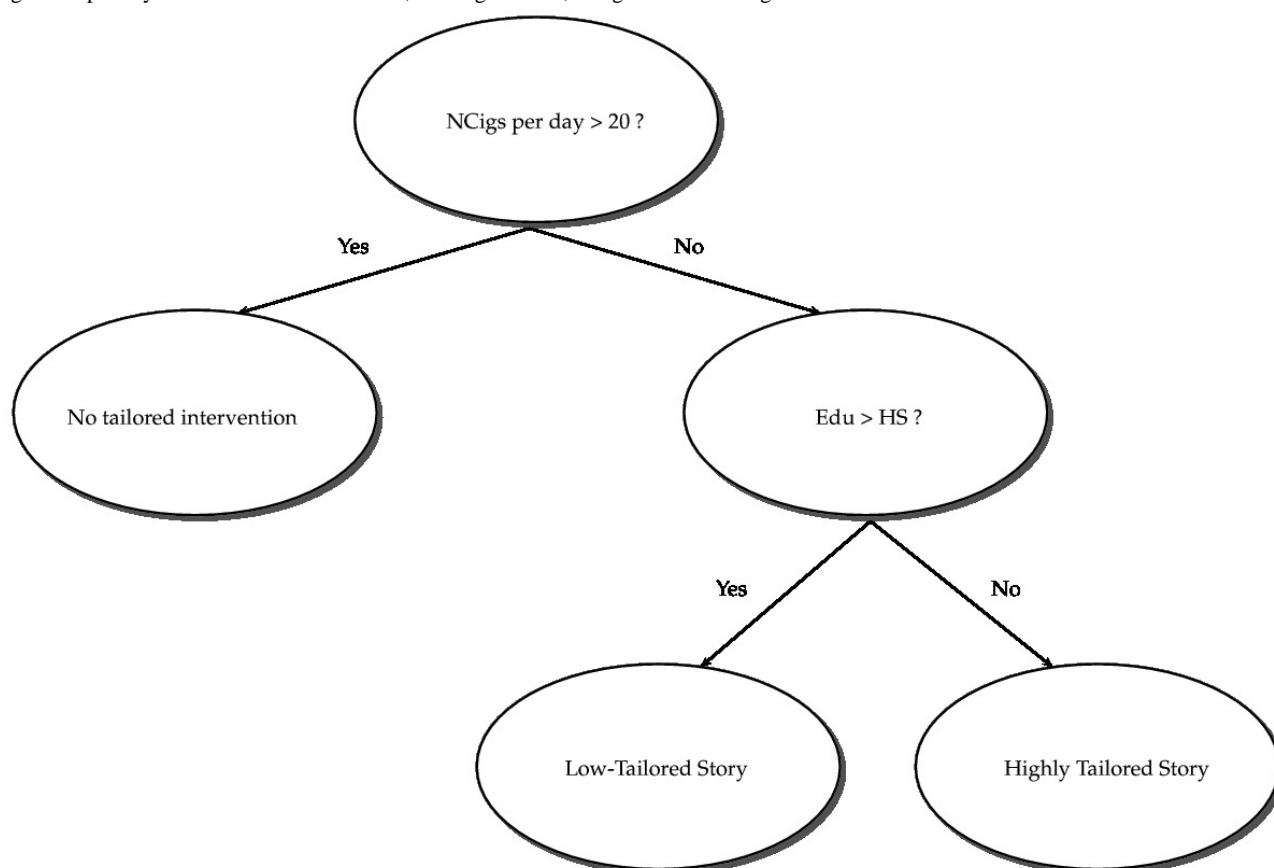
Variable	Estimate	Z value	Adjusted rate ratio (95% CI)	P value
Age (years)	0.002	0.676	1.002 (0.997-1.006)	.50
Gender (male)	0.083	1.539	1.087 (0.972-1.202)	.12
Race white	−0.148	−1.813	0.862 (0.724-1)	.07
Race black	−0.051	−0.504	0.950 (0.762-1.139)	.61
Number of cigarettes smoked per day (NCigs per day)	−0.014	−2.108	0.986 (0.973-0.999)	.03 ^a
Motivation	0.095	1.687	1.099 (0.978-1.221)	.09
Education	0.256	3.142	1.291 (1.085-1.497)	.002 ^a
Self-efficacy	−0.159	−2.131	0.853 (0.729-0.978)	.03 ^a
Story	0.189	2.022	1.208 (0.987-1.428)	.04 ^a
Source	−0.077	−1.028	0.926 (0.790-1.062)	.30
Story × education	−0.414	−3.700	0.661 (0.516-0.806)	<.001 ^a
Source × self-efficacy	0.052	0.499	1.053 (0.839-1.267)	.62

^aDenotes $P < .05$.

Table 5. Estimated intervention effect of story, expressed as a percentage change in the expected number of quit attempts, stratified by the education level for the persons who smoked less than or equal to 20 cigarettes per day (n=546).

Education	Estimate (95% CI)
>High school (Edu=1)	−20.20 (−29.93 to −10.47)
≤High school (Edu=0)	20.76 (−1.32 to 42.83)

Figure 2. An estimated treatment regimen to recommend personalized smoking cessation intervention for smokers who smoked less than or equal to 20 cigarettes per day at baseline. Edu: education; HS: high school; NCigs: number of cigarettes smoked.



Discussion

Principal Findings

This study aimed to stratify smokers who are likely to benefit from tailored smoking cessation intervention programs and those who are not. This will allow us to develop personalized smoking cessation interventions. Outcomes of this study can potentially help policy makers to allocate limited public health resources to target subgroups of smokers who are more likely to be successful from tailored smoking cessation interventions. This study analyzed existing data from a large randomized, Web-based smoking cessation trial (Project Quit) to answer the above research question.

In Project Quit, Strecher et al [8] previously studied the impact of 5 Web-based intervention components on the 7-day point prevalence. Using a multivariable logistic regression model, they found that 2 intervention components, namely, story and source, have significant effect on smoking cessation at 6 months. Furthermore, they showed that participants with lower education were more influenced by highly tailored stories, and a highly personalized source had marginally greater impact on smoking cessation in participants with higher self-efficacy. However, 1 drawback with the 7-day point prevalence is that it does not take into account the quitting process over the entire study period. Instead, it only focuses on smoking behavior in a limited time window at the end of the 6-month follow up. Our investigation was designed to specifically overcome this limitation by considering the number of quit attempts during

the whole study period as the outcome of interest and examine whether similar effects still hold. The number of quit attempts quantified participants' willingness and involvement in the smoking cessation process throughout the study period. Our study incorporated 2 findings from prior analysis of Project Quit [8] into this analysis. First, of the 5 intervention components from the original trial, we considered only 2 (story and source) into our model because of their significant effects (remaining components were insignificant). Second, based on the a priori hypothesis and data analysis from Strecher et al [8], we only included 2 interaction effects—one between story and education, and another between source and self-efficacy.

We found that participants with lower education (high school graduates or less) were positively influenced by a high-tailored story to quit smoking, whereas those with higher education were better off with a low-tailored story. Our findings are consistent with those from Strecher et al [8], who found similar effect modification, albeit in the context of the 7-day point prevalence. Findings on the low education group are not surprising as participants in this category can easily associate themselves with fictitious characters in the story that are tailored to them (socioeconomically or otherwise), as opposed to untailored characters. Such association allows the low-educated subjects to “transport” themselves into the story, thus resulting in successful smoking cessation. Strecher et al [8] suggested that the extent of being “transported” has a strong influence on persuasion, which in particular applies to smokers' attempt and behavior to quit [24,25]. In contrast to Strecher et al [8], we did not find any significant interaction between source and

self-efficacy in influencing the number of quit attempts. We speculate that such discrepancy may be due to different outcome under consideration or the smaller sample size in this investigation.

Here, we summarize the strengths of this study that was designed to address gaps in the extant literature on smoking cessation. First, ours is the first analysis of quit attempts data from Project Quit. This will potentially shed new light on smokers' quitting process experience while participating in a Web-based smoking cessation program. Second, although analysis of quit attempts data are available in the literature [14-18], they are based on observational cohort studies. Our study is the first to analyze quit attempts data from a randomized trial. Third, we utilize TR as a perspective from the precision medicine literature to help better understand the type of smokers who will benefit from tailored intervention in optimizing their quitting effort. The results suggest that smokers with low education are more likely to benefit from tailored interventions. This is consistent with prior findings based on the 7-day point prevalence data [8,20,26]. Thus, this analysis validates the significance of number of quit attempts as an alternative to the commonly used point prevalence outcome. Finally, from a methodological perspective, number of quit attempts is a count-type variable rather than a continuous measurement or binary indicator. Hence, we employed Poisson regression to analyze the dataset.

There are a few limitations in our study. First, because the number of quit attempts was a secondary outcome in Project Quit, this variable had higher rate of missingness compared with the primary outcome of the 7-day point prevalence. For simplicity and easy interpretation, we only conducted a complete-case analysis. One could potentially employ missing data analysis techniques (eg, multiple imputation) to impute the missing values before conducting the analysis. However, as we did not find any evidence of differential missingness across the intervention arms, we argue that the missingness in the current data is mostly noninformative. Thus, data imputation techniques would not offer much benefit over a complete-case analysis.

Second, the intervention components in the original study were designed to influence the 7-day point prevalence. One could conceive other potential intervention components not studied in Project Quit, which may potentially better influence smokers' involvement in their difficult journey toward quitting and their number of quit attempts in particular. We believe that new studies specifically designed to understand the impact of tailored interventions on quit attempts are necessary to answer such questions. Third, the number of quit attempts is a self-reported outcome over a reasonably long period. As it is unrealistic for the participants to remember their exact number of quit attempts in the past 6 months, this variable may have recall bias. However, this concern can be addressed in the current era of mobile health and sensor technologies. New-generation studies should employ mobile apps and wearable devices to capture quit attempts data more accurately and thus minimize measurement errors.

Conclusions

In this study, we aimed to shed new lights on the impacts of Web-based tailored psychosocial and communication intervention components on smoking cessation. Using data from a randomized Web-based trial, we examined the number of quit attempts during a 6-month study period. We also investigated how these impacts are modified by individual characteristics. Collectively, we aimed to identify subgroups of smokers who would successfully benefit from Web-based tailored interventions. We found that highly individually tailored story is significantly more effective for smokers with low education (high school graduate or less) compared with those with higher education (at least some college exposure). Our findings can provide evidence and potentially help policy makers to utilize limited public health resources to cease smoking in low-educated smokers. Nevertheless, we must cautiously note that the number of quit attempts in this study is self-reported, and thus subjected to recall bias. Future studies that incorporate sensor and/or mobile technologies to collect precise data on quit attempts are clearly warranted.

Acknowledgments

The original Project Quit trial was funded by the US NCI grants P50 CA101451 and R01 CA101843. This work (secondary analysis of Project Quit data) was supported by BC's start-up grant from Duke-NUS as well as MOE2015-T2-2-056 grant from Singapore's Ministry of Education.

Conflicts of Interest

None declared.

Editorial notice: The original randomized study analyzed in this paper was not registered, explained by authors by the fact that the original RCT (Project Quit) started around 2005 when trial registration was not yet mandatory. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

References

1. World Health Organization. 2011. WHO Report on the Global Health Epidemic, 2011: Warning About the Dangers of Tobacco URL: http://apps.who.int/iris/bitstream/10665/44616/1/9789240687813_eng.pdf [WebCite Cache ID 6zG44JcuU]

2. Ekpu VU, Brown AK. The economic impact of smoking and of reducing smoking prevalence: review of evidence. *Tob Use Insights* 2015 Jul 14;8:1-35 [[FREE Full text](#)] [doi: [10.4137/TUI.S15628](#)] [Medline: [26242225](#)]
3. Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. *Patient Educ Couns* 1999 Feb;36(2):107-117. [Medline: [10223016](#)]
4. Piper ME, Fiore MC, Smith SS, Fraser D, Bolt DM, Collins LM, et al. Identifying effective intervention components for smoking cessation: a factorial screening experiment. *Addiction* 2016 Jan;111(1):129-141 [[FREE Full text](#)] [doi: [10.1111/add.13162](#)] [Medline: [26582269](#)]
5. Swartz LH, Noell JW, Schroeder SW, Ary DV. A randomised control study of a fully automated Internet based smoking cessation programme. *Tob Control* 2006 Feb;15(1):7-12 [[FREE Full text](#)] [doi: [10.1136/tc.2003.006189](#)] [Medline: [16436397](#)]
6. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction* 2005 May;100(5):682-688. [doi: [10.1111/j.1360-0443.2005.01093.x](#)] [Medline: [15847626](#)]
7. Ghorai K, Akter S, Khatun F, Ray P. mHealth for smoking cessation programs: a systematic review. *J Pers Med* 2014 Jul 18;4(3):412-423 [[FREE Full text](#)] [doi: [10.3390/jpm4030412](#)] [Medline: [25563359](#)]
8. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. *Am J Prev Med* 2008 May;34(5):373-381 [[FREE Full text](#)] [doi: [10.1016/j.amepre.2007.12.024](#)] [Medline: [18407003](#)]
9. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. *J Med Internet Res* 2008 Nov 04;10(5):e36 [[FREE Full text](#)] [doi: [10.2196/jmir.1002](#)] [Medline: [18984557](#)]
10. Chakraborty B, Moodie EE. *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. New York: Springer-Verlag; 2013.
11. Kosorok MR, Moodie EE, editors. *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*. Philadelphia, PA and Alexandria, VA: ASA-SIAM; 2015.
12. Nahum-Shani I, Ertefaie A, Lu XL, Lynch KG, McKay JR, Oslin DW, et al. A SMART data analysis method for constructing adaptive treatment strategies for substance use disorders. *Addiction* 2017 May;112(5):901-909. [doi: [10.1111/add.13743](#)] [Medline: [28029718](#)]
13. Wu F, Laber EB, Lipkovich IA, Severus E. Who will benefit from antidepressants in the acute treatment of bipolar depression? A reanalysis of the STEP-BD study by Sachs et al. 2007, using Q-learning. *Int J Bipolar Disord* 2015 Apr 03;3:7 [[FREE Full text](#)] [doi: [10.1186/s40345-014-0018-5](#)] [Medline: [25844303](#)]
14. Rose JS, Chassin L, Presson CC, Sherman SJ. Prospective predictors of quit attempts and smoking cessation in young adults. *Health Psychol* 1996 Jul;15(4):261-268. [Medline: [8818672](#)]
15. Sendzik T, McDonald PW, Brown KS, Hammond D, Ferrence R. Planned quit attempts among Ontario smokers: impact on abstinence. *Addiction* 2011 Nov;106(11):2005-2013. [doi: [10.1111/j.1360-0443.2011.03498.x](#)] [Medline: [21592249](#)]
16. Borland R, Partos TR, Yong HH, Cummings KM, Hyland A. How much unsuccessful quitting activity is going on among adult smokers? Data from the International Tobacco Control Four Country cohort survey. *Addiction* 2012 Mar;107(3):673-682 [[FREE Full text](#)] [doi: [10.1111/j.1360-0443.2011.03685.x](#)] [Medline: [21992709](#)]
17. Diemert LM, Bondy SJ, Brown KS, Manske S. Young adult smoking cessation: predictors of quit attempts and abstinence. *Am J Public Health* 2013 Mar;103(3):449-453. [doi: [10.2105/AJPH.2012.300878](#)] [Medline: [23327264](#)]
18. Chaiton M, Diemert L, Cohen JE, Bondy SJ, Selby P, Philipneri A, et al. Estimating the number of quit attempts it takes to quit smoking successfully in a longitudinal cohort of smokers. *Br Med J Open* 2016 Jun 09;6(6):e011045 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2016-011045](#)] [Medline: [27288378](#)]
19. 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](#)]
20. Nair V, Strecher V, Fagerlin A, Ubel P, Resnicow K, Murphy S, et al. Screening experiments and the use of fractional factorial designs in behavioral intervention research. *Am J Public Health* 2008 Aug;98(8):1354-1359 [[FREE Full text](#)] [doi: [10.2105/AJPH.2007.127563](#)] [Medline: [18556602](#)]
21. Chakraborty B, Collins LM, Strecher VJ, Murphy SA. Developing multicomponent interventions using fractional factorial designs. *Stat Med* 2009 Sep 20;28(21):2687-2708 [[FREE Full text](#)] [doi: [10.1002/sim.3643](#)] [Medline: [19575485](#)]
22. Bandura A. Health promotion by social cognitive means. *Health Educ Behav* 2004 Apr;31(2):143-164. [doi: [10.1177/1090198104263660](#)] [Medline: [15090118](#)]
23. Comprehensive R Archive Network. The R Project for Statistical Computing URL: <http://www.cran.r-project.org/> [accessed 2018-05-08] [[WebCite Cache ID 6zG3dFkak](#)]
24. Taylor SE, Thompson SC. Stalking the elusive "vividness" effect. *Psychol Rev* 1982;89(2):155-181.
25. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. *J Pers Soc Psychol* 2000 Nov;79(5):701-721. [Medline: [11079236](#)]
26. Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat Methods Med Res* 2010 Jun;19(3):317-343 [[FREE Full text](#)] [doi: [10.1177/0962280209105013](#)] [Medline: [19608604](#)]

Abbreviations

GHC: Group Health Cooperative
HFHS: Henry Ford Health System
HMO: health maintenance organization
NCI: National Cancer Institute
NCig: number of cigarettes smoked per day
RR: rate ratio
TR: treatment regimen

Edited by G Eysenbach; submitted 02.12.17; peer-reviewed by F Khatun, IN Gomez; comments to author 08.02.18; revised version received 28.03.18; accepted 11.04.18; published 20.06.18.

Please cite as:

Chakraborty B, Maiti R, Strecher VJ

The Effectiveness of Web-Based Tailored Smoking Cessation Interventions on the Quitting Process (Project Quit): Secondary Analysis of a Randomized Controlled Trial

J Med Internet Res 2018;20(6):e213

URL: <http://www.jmir.org/2018/6/e213/>

doi: [10.2196/jmir.9555](https://doi.org/10.2196/jmir.9555)

PMID: [29925494](https://pubmed.ncbi.nlm.nih.gov/29925494/)

©Bibhas Chakraborty, Raju Maiti, Victor J Strecher. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.06.2018. 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.

Original Paper

An Online Hand Exercise Intervention for Adults With Rheumatoid Arthritis (mySARAH): Design, Development, and Usability Testing

Cynthia Srikesavan^{1*}, PhD; Esther Williamson^{1*}, PhD; Tim Cranston², BSc (Hon); John Hunter¹, BSc (PT); Jo Adams³, PhD; Sarah E Lamb¹, DPhil

¹Rehabilitation Research in Oxford, Nuffield Department of Orthopedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

²Oxford Clinical Trials Research Unit, Nuffield Department of Orthopedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Headington, Oxford, United Kingdom

³Centre for Innovation and Leadership in Health Sciences, Faculty of Health Sciences, University of Southampton, Southampton, United Kingdom

*these authors contributed equally

Corresponding Author:

Cynthia Srikesavan, PhD

Rehabilitation Research in Oxford

Nuffield Department of Orthopedics, Rheumatology, and Musculoskeletal Sciences

University of Oxford

Windmill Road

Oxford, OX3 7LD

United Kingdom

Phone: 44 01865737907

Email: cynthia.srikesavan@ndorms.ox.ac.uk

Abstract

Background: The Strengthening and Stretching for Rheumatoid Arthritis of the Hand (SARAH) program is a tailored, progressive 12-week exercise program for people with hand problems due to rheumatoid arthritis. The program was shown to be clinically and cost-effective in a large clinical trial and is recommended by the UK National Institute for Health and Care Excellence (NICE) guidelines for rheumatoid arthritis in adults.

Objective: We have developed an online version of the SARAH program (mySARAH) to make the SARAH program widely accessible to people with rheumatoid arthritis. The purposes of this study were to develop mySARAH and to evaluate and address its usability issues.

Methods: We developed mySARAH using a three-step process and gaining feedback from patient contributors. After initial development, mySARAH was tested in two iterative usability cycles in nine participants using a simplified think-aloud protocol and self-reported questionnaires. We also evaluated if participants executed the SARAH exercises correctly after watching the exercise videos included on the website.

Results: A preliminary version of mySARAH consisting of six sessions over a 12-week period and delivered via text, exercise videos, images, exercise plan form, exercise calendar, and links to additional information on rheumatoid arthritis was developed. Five participants (1 male; 4 females; median age 64 years) and four participants (four females; median age 64.5 years) took part in the first and second usability testing cycles respectively. Usability issues identified from Cycle 1 such as having a navigation tutorial video and individualised feedback on pain levels were addressed prior to Cycle 2. The need for more instructions to complete the mySARAH patient forms was identified in Cycle 2 and was rectified. Self-reports from both cycles indicated that participants found the program useful and easy to use and were confident in performing the SARAH exercises themselves. Eight of the nine participants correctly demonstrated all the exercises.

Conclusions: mySARAH is the first online hand exercise intervention for people with rheumatoid arthritis. We actively involved target users in the development and usability evaluation and ensured mySARAH met their needs and preferences.

(*J Med Internet Res* 2018;20(6):e10457) doi:[10.2196/10457](https://doi.org/10.2196/10457)

KEYWORDS

rheumatoid arthritis; hand joints; exercise training; web-based

Introduction

Background

The Strengthening And stretching for Rheumatoid Arthritis of the Hand (SARAH) program is a tailored, progressive 12-week exercise program for people with hand problems due to Rheumatoid Arthritis (RA) [1]. The SARAH program was designed as an addition to best practice usual care (joint protection education and functional splinting and assistive devices) for adults with RA who had pain and hand function problems and had been on a stable drug regimen for at least 3 months. A pragmatic clinical trial [2,3] was conducted at 17 National Health Service (NHS) sites across the United Kingdom in 490 people who were randomized to receive best practice usual care or best practice usual care plus the SARAH program [3].

Patients who received the SARAH program in the trial attended 6 face-to-face appointments with a registered physiotherapist or occupational therapist who was a hand therapist or was experienced in rheumatology. The program included 7 upper limb mobility exercises—metacarpophalangeal flexion, tendon gliding (hook, straight, and full fist), radial walking, finger abduction, wrist circumduction, hand-behind-head, and hand-behind-back and 4 strength exercises for the hand—gross grip, pinch grip, finger adduction, and eccentric wrist extension.

Integral to the SARAH program are behavioral support strategies such as self-monitoring, goal setting, and action planning to improve patients' self-efficacy, that is, the patients' confidence to carry out the SARAH exercises independently. At the start of the program, patients were assisted by therapists to complete a personal exercise guide to set functional goals relating to their hand problems in accordance with SMART (Specific, Measurable, Attainable, Relevant, and Timely) principles and to make an exercise plan of "when" and "where" to do the SARAH exercises. Patients were also asked to complete an exercise diary to monitor the completion of the exercises [1-3].

During the subsequent appointments, the therapists and patients jointly reviewed the exercise diary and the personal exercise guide to set new goals and an exercise plan or modify them, if required. If a patient had difficulties adhering to the SARAH program, the patient and the therapist worked together to identify barriers to complete the exercises and to discuss realistic solutions and ways to maximize facilitating factors. A barriers and facilitators form was completed to guide this discussion. Exercises were progressed or regressed using a standardized protocol to ensure that the exercises were tailored to each patient. Patients were provided with a discharge advice sheet, exercise booklet and copies of exercise diary, personal exercise guide, and barriers and facilitators form during their final clinical appointment. They were encouraged to continue the exercises independently at home.

The key findings of the SARAH trial [3] were as follows. At 4 months, the group that received the SARAH program showed improvements in hand function double that of the usual care group (8.7 points improvement in the hand function subscale of the Michigan Hand Outcome Questionnaire (0-100) compared

with an improvement of 4 points; mean difference 4.7 points). At 12 months, the group that received the SARAH program had improvements in hand function double that of the usual care group (7.9 points improvement in the hand function subscale of the Michigan Hand Outcome Questionnaire compared with an improvement of 3.6 points; mean difference 4.3 points). The SARAH program did not result in any adverse effects, for example, increased joint pain, stiffness, or "flare-ups." The SARAH program was also cost-effective.

These findings led to an update of the 2015 NICE guidelines recommending the SARAH program for adults who have hand problems due to RA [4]. Following this, it was important to develop a plan to disseminate the evidence-based SARAH program and ensure uptake by the target users, that is, people with RA and, thereby, facilitate improved patient care.

The overarching purpose of this project was to evaluate the adaptation of the SARAH program, originally designed to be delivered face-to-face by a therapist, to a self-guided online version (mySARAH), in which people with RA undertake the SARAH program without therapist supervision. We propose a knowledge translation initiative of an online version of the SARAH program (mySARAH) to disseminate the SARAH program to the target users by taking advantage of the increasing accessibility and use of the Internet [5-7].

The dissemination of mySARAH is guided by the 5 steps of Analysis, Design, Development, Implementation, and Evaluation (ADDIE) instructional system design model [8,9]. We will report on the first 3 steps of this process in this paper.

Step 1: Analysis

A needs assessment was undertaken to understand what target users required in a knowledge translation tool (mySARAH) to bridge the gap between knowledge (SARAH program) and action (making the SARAH program available in an easily accessible format for people with RA).

Step 2: Design

The online prototype of mySARAH was designed and users provided feedback.

Step 3: Development

The preliminary version of mySARAH was developed and usability issues were evaluated for a final version. Future work will focus on the final 2 steps.

Step 4: Implementation

mySARAH will be launched, first with a small group of target users and then into the public domain.

Step 5: Evaluation

Reach, Effectiveness, Adoption, Implementation, and Maintenance of mySARAH will be measured in a small group of target users and then a large target population.

Objectives

Step 1 consists of the needs assessment. Here our objectives were to collect users' opinions and preferences for mySARAH and adapt the SARAH program to fit mySARAH. Step 2

involves the design of mySARAH. Here we designed the online prototype of mySARAH and collected user feedback on the prototype. Step 3 involves the development of mySARAH. Our objectives at this stage were to revise the mySARAH prototype toward the preliminary version, evaluate and address usability issues, evaluate if participants could replicate the SARAH exercises correctly, and produce the final version of mySARAH.

Methods

A flow diagram of the study is presented in [Figure 1](#).

Step 1: Needs Assessment of mySARAH

We invited patient contributors from the public and the local branch of the National Rheumatoid Arthritis Society (patient support group) to assist with the development of mySARAH. We conducted face-to-face or phone meetings with 5 patient contributors (age range 50 to 66 years, duration range since RA diagnosis 1-12 years). We explained the components, including patient advice, the type of exercises, the number of exercise sessions, and the use of exercise diaries, goal setting and exercise planning, and how these components might be transferred to mySARAH, of the SARAH program delivered in the SARAH clinical trial. We explored their needs, preferences, and expectations for an online hand exercise program. We then collectively summarized their input. In response to this, the research team adapted the SARAH program for the mySARAH online prototype. The user feedback and SARAH program adaptations agreed upon by the team are presented in the Results section.

Step 2: Designing mySARAH

We incorporated the common heuristic principles recommended by Baumel and Muench [10] into the mySARAH prototype. For example, simple functionality and navigation features, a tunneled approach to every exercise, and review session to respond to users' needs with respect to the standard hand clinical appointments, "In-house" tools such as a "Contact us" button to facilitate user engagement, automatic email reminders to remind users about sessions that were missed or incomplete, and an exercise checklist and exercise calendar features for self-monitoring.

We also reviewed the published literature [11] to identify the types of features of successful internet-delivered self-guided health interventions and included them in mySARAH, for example, 1) having progressive modules over weeks or months requiring active user engagement, 2) having external links for additional health information, 3) having cognitive behavioral strategies, 4) having self-monitoring tools, and 5) providing patient education.

We further used the Behavioral Intervention Technology model [12,13] to schematically map the components of mySARAH ([Table 1](#)). The Behavioural Intervention Technology (BIT) Model summarizes the following 5 components to describe an eHealth intervention: aims, behavior change strategies, elements, characteristics, and workflow. The aims and behavior change strategies cover the conceptual aspects of "Why" and "How" and the other 3 components cover the technical aspects of "What," "How," and "When" of an eHealth intervention.

Figure 1. Study flow diagram.

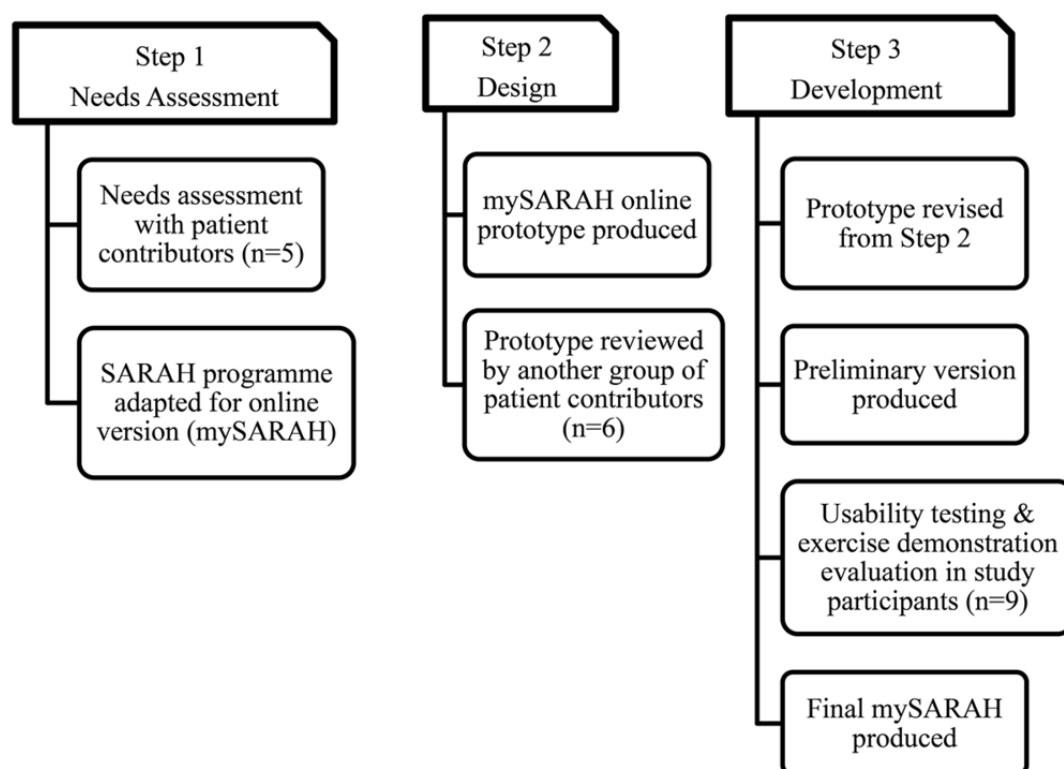


Table 1. The Behavioural Intervention Technology (BIT) Model mapped to the online version of the Strengthening And stretching for Rheumatoid Arthritis of the Hand (mySARAH) exercise intervention.

Conceptual and technical components	BIT Model components	mySARAH Examples
Why (Conceptual)	Aims	<ul style="list-style-type: none"> To provide adults with RA^a access to the SARAH^b program as part of a strategy for self-management To teach them to carry out the SARAH exercises correctly To promote long-term adherence to the SARAH exercises To improve and maintain hand function
How (Conceptual)	Behavioral change strategies	<ul style="list-style-type: none"> Knowledge: Information about RA, why hand exercises are important Goal setting: Set SMART^c goals related to hand function Action planning: Plan “When” and “Where” to do the SARAH exercises Problem solving: Identify and overcome barriers to exercise adherence or to maximize the use of facilitators Self-monitoring: Monitor one’s own exercise adherence behavior Review goals: Modify or set new SMART goals Instruction and Demonstration: Advice on joint protection and demonstration of SARAH exercises
What (Technical)	Elements	<ul style="list-style-type: none"> Reminder emails, exercise calendars, messaging, session notes, summary reports, and user fillable forms
How (Technical)	Characteristics	<ul style="list-style-type: none"> Medium: Text, images and videos Complexity: Easy to complete tasks, forms, and exercise calendars Aesthetics: Simple and less distractive Personalization features: None
When (Technical)	Workflow	<ul style="list-style-type: none"> Tunneled, Task-Based (example: user must complete mySARAH Session 1 to access Session 2), and time-based delivery (example: Successive mySARAH Sessions scheduled based on previous session completion date) Continued user access to mySARAH elements after 12 weeks

^aRA: rheumatoid arthritis.

^bSARAH: Strengthening And stretching for Rheumatoid Arthritis of the Hand.

^cSMART: Specific, Measurable, Attainable, Relevant, and Timely.

The conceptual “Why” defines the clinical aims of the intervention and “How” defines the behavioral strategies used to achieve the aims. The technical “What” defines the components of the intervention, “How” describes the characteristics (the medium used, the complexity of the content, aesthetics, and personalization features) of the components, and “When” defines the workflow of the intervention delivery.

We designed the prototype on Drupal (Version 7.0), a widely used open source content management system [14]. A simple layout and a relatively plain background were used in mySARAH. A total of 6 SARAH exercise and review sessions were designed over a 12-week period where users learned to perform SARAH exercises, set their own SMART goals, and planned their exercise schedule. mySARAH included exercise calendars, a pain scale, summary reports of patient completed goals and exercise plan form, a frequently asked questions section, exercise illustrations, exercise and instructional videos, and a facility to download and save a digital copy of the content of each completed session. It also had additional patient information on RA and links to online suppliers to purchase the SARAH exercise equipment, such as gel balls, thera putty, and resistance bands. A multimedia approach with a combination of text, videos, and images was utilized for intervention delivery.

All sessions were tunnel-based, that is, users must complete a session before advancing to the next one. The consecutive session was automatically scheduled on the user’s calendar based on the completion date of the previous session. For example, upon completing session one, the second session would be released to the user after a weeks’ time.

We used the professionally produced videos on joint protection advice and SARAH exercises that were produced to train hand therapists to deliver the SARAH program using an online training program [15]. We also produced preliminary videos with a member of the team on goal setting, action planning, and simple ways to adhere to SARAH exercises.

We aimed to write the mySARAH content in plain English language without any technical or medical jargon. We checked this using the Gunning Fog index [16], a formula which estimates the readability of a piece of text by considering the number of words per sentence and the proportion of words which contain 3 or more syllables. The Fog readability index indicates the number of years of formal education required for a reader to understand the text. We used an online tool [16] to check the readability of mySARAH. In order to be readable and comprehensible to patients from a broad range of educational

backgrounds, we aimed for school grades between 5 and 9 that correspond to 10 and 14 years of reading age [17].

The prototype and mySARAH logo were reviewed by another group of 6 patient contributors (age range 59-75 years, duration since RA diagnosis 1-5 years) in a half-day meeting. Following this meeting, we produced additional preliminary videos to describe patients' experiences on how RA affected their hand function and why exercising the hands was important. The modifications made in the mySARAH prototype are presented in the Results section.

Step 3: Developing mySARAH

Guided by Step 2, we modified the prototype and developed the preliminary version of mySARAH. We then tested mySARAH to identify and resolve usability issues for producing the final version. The usability testing protocol was reviewed and approved by the Medical Sciences Inter-Divisional Research Ethics Committee, University of Oxford (R52172/RE001).

Based on the existing evidence that over 80% of usability issues can be identified with 5-9 participants and 95% with 9 participants [18,19], we proposed a convenience sample of 10 participants for usability testing. Adults having problems with hand function due to RA and living within 2 hours of travel to the study site were considered eligible to participate in the usability testing. Participants were invited via online advertisements, e-newsletters, local patient groups, and social media of Arthritis Research UK, National Rheumatoid Arthritis Society and Patients Active in Research organizations, and by word of mouth. We asked interested volunteers to contact the SARAH implementation team directly by email or phone. Appointments were arranged for the individual volunteers to attend a one-off 90-minute usability testing session at the study site.

Two researchers from the SARAH implementation team conducted the sessions. One researcher observed and took notes, whereas the other was a session facilitator. The facilitator explained the testing procedures to each participant, emphasizing that it was the evaluation of the website and not the user. Participants were asked to provide information on their age, gender, educational level, employment status, and ethnicity, years since RA diagnosis, and hours spent on internet each day. We conducted two iterative cycles, the first cycle with 5 volunteers and the second with 4 volunteers. We used the following procedures in our usability testing.

Simplified Concurrent Think-Aloud Protocol

The facilitator asked participants to navigate through mySARAH, complete assigned tasks [20,21], for example, creating an account, watching videos, and completing a session, and simultaneously talk about what they feel, see, or think while browsing. When participants had difficulties verbalizing, the facilitator encouraged them by a "Keep talking" signboard and assisted with prompts only if required. The think-aloud sessions were audio-recorded.

Exercise Demonstration

Next, the facilitator asked participants to watch one SARAH exercise video at a time and repeat the exercise while watching.

The video was then closed, and the facilitator asked them to demonstrate 3 repetitions of the exercise they had just watched. The facilitator also monitored participants for pain or discomfort in their fingers and/or wrists while demonstrating the exercises, for example, if they were stretching out a stiff joint. If a participant reported pain beyond slight discomfort, we reduced the number of repetitions from 3 to 1 or discontinued the demonstration session.

The note-taker observed the participants' ability to correctly demonstrate each exercise, including choosing the right baseline resistance level for strength exercises, and documented any difficulties, doubts, and comments reported. A simple 1-3-point scale (1=correctly demonstrated, 2=assistance required from evaluator or by replaying the video, and 3=difficulty demonstrating the exercise correctly after being assisted), developed by the SARAH implementation team, was used to rate the correct execution of exercises and baseline load setting.

Subjective Reports

We used the Computer System Usability Questionnaire [22] to evaluate the user satisfaction, ease of use, information, and interface of online program with a 7-point Likert scale, 1 representing "Strongly disagree" and 7 representing "Strongly agree." We measured the perceived usefulness with a 1-5 Likert scale that is scored from 1=Not at all useful to 5=Extremely useful; ease of use with a 1-5 Likert scale that is scored from, 1=Very difficult to 5=Very easy; and confidence in doing the SARAH exercises with a 1-5 Likert scale that is scored from 1=Not at all confident to 5=Very confident. The findings from this step are presented in the Results section.

Data Analysis

We listened to the audio files of the think-aloud sessions along with notes from each session and created a list of key usability issues reported by participants. The demographic characteristics of usability testing participants were summarized, and the questionnaire scores were reported as medians and interquartile ranges.

Results

Step 1: Needs Assessment of mySARAH

The patient contributors preferred a hand exercise website having (1) a simple layout, (2) short exercise and instructional videos, (3) brief paragraphs with content written in a clear and straightforward language, (4) links to additional information on rheumatoid arthritis, (5) email reminders with option to select the frequency of reminders received, (6) a simple screening process with questions confirming age and RA diagnosis, and (7) a separate section for "Frequently asked questions."

In addition to the above, the following modifications were made from the SARAH program in the trial for the mySARAH prototype. In the clinical trial, participants attended face-to-face appointments with a hand therapist at weeks 1, 2, 4, 6, 9, and 12. For mySARAH, we made some modifications to the timing of sessions to better fit the online delivery of the program. mySARAH sessions were scheduled to occur at weeks 1, 2, 3, 6, 9, and 12. We used a simple lay term "exercise plan form"

for “personal exercise guide.” We simplified the personal exercise guide by removing the confidence scale asking patients how confident they were to achieve their goals and the exercise contract section between the patient and therapist. Instructional videos provided tips on adhering with the program instead of using the barriers and facilitators form and two mySARAH logos were designed.

Step 2: Designing mySARAH

The analysis of the text (excluding tables, figures, hyperlinks) contained in mySARAH produced a Gunning Fog readability index ranging from approximately 7.9-11 years of education, which corresponds to an approximate reading age of 13-16 years. Some text was higher in the readability levels than ideal, but this was due to the high proportion of polysyllabic words and medical terms (examples: exercise and RA) that could not be modified further. We provided simple explanations of any medical terms.

The approximate session-wise Gunning Fog indexes were as follows: Session 1: 9.6 (reading age 15 years), Session 2: 7.9 (reading age 13 years), Session 3: 10.6 (reading age 16 years), Session 4: 10.8 (reading age 16 years), Session 5: 11 (reading age 16 years), and Session 6: 11 (reading age 16 years).

The patient contributors who participated in the half-day meeting liked the layout and look of the website. They felt that the information and the language were clear and easy to follow. They also liked the features of exercise calendar, email reminders, exercise videos, frequently asked questions section, and facility to have a summary record of their goals and exercise plan. They also agreed on one of the two mySARAH logos produced by the team. Several revisions were suggested, including breaking long paragraphs into shorter paragraphs, using bullet points to break up lengthier sections of text, ensuring the pages were not cluttered, page proofreading, the addition of contact details for exercise equipment suppliers, and additional details, at the end of the program, pertaining to a continued access to mySARAH.

Step 3: Developing mySARAH

The preliminary version of mySARAH was produced and the revisions suggested by users in Step 2 were incorporated. The resultant version still closely resembled the prototype.

We enrolled 10 participants in the usability testing, which took place in two cycles. A total of 9 participants completed the testing and one volunteer withdrew after consenting due to a family member's sickness.

The demographic characteristics of participants who took part in the usability testing Cycle 1 and Cycle 2 sessions are presented in [Table 2](#). Participants felt that the mySARAH website was self-explanatory, easy to use, and contained all the information needed about SARAH. They reported that the registration process was straightforward, the goal setting and exercise planning form and exercise calendar was helpful, the

website content was relevant, the exercise and other videos were helpful and engaging, and the forms were easy to fill out. The main usability issues identified by participants from each cycle and the subsequent revisions are listed in [Table 3](#).

The Computer System Usability Questionnaire and Likert Scale scores from both cycles of usability testing are listed in [Table 4](#). All participants had a good agreement (with scores above 6) on almost all the items of the Computer System Questionnaire, especially in terms of satisfaction, ease of use, and the content. Overall, 3 participants found Item 8 stating “I believe I became productive quickly using this system” irrelevant to the tasks they completed and, therefore, did not score it, and 7 participants felt that Item 9 stating “The system gives error messages that clearly tell me how to fix problems” was not relevant because they did not encounter any issues while filling in the fields in mySARAH forms. Therefore, we have not reported these two items. However, participants raised the concern of being notified to rectify any errors while filling the forms. We duly addressed this usability issue by adding pop-up error notifications in the final version.

The Likert scale scores indicate that all users found the program useful and easy to use and were confident in their ability to do the SARAH exercises themselves.

Overall, majority of the participants (8/9, 89%) correctly demonstrated all 11 SARAH exercises, scoring 1 on the 1-3 exercise demonstration scale. One participant required guidance for the “Spread fingers” and “Hand squeeze” exercises and found holding the resistance band between fingers for the “Wrist backward bends” strength exercise difficult to demonstrate.

After addressing Cycle 2 usability issues, a few additional revisions were made by the research team in the final version of mySARAH. Specifically, a “Go to homepage” tab was created to signpost to respective sessions on logging in. We produced and filmed a professional mySARAH promotional video, a navigation tutorial video, and informational videos on clinical aspects of RA and behavioral strategies for exercise adherence. A patient video demonstrating how to set baseline load for each strengthening exercise was additionally produced ([Multimedia Appendix 1](#)). Both therapist-patient and patient-demonstrated videos for wrist backward bends exercise were combined as a single video and added to mySARAH. We had the mySARAH pages proofread by a patient volunteer and a member of the research team.

Final version of mySARAH

A brief description of mySARAH ([Multimedia Appendix 1](#)) and screenshots of mySARAH ([Multimedia Appendix 2](#)) are presented herein. [Table 5](#) provides an overview of the final mySARAH sessions' content. Sessions were accessed by users on a preset timetable, so they had adequate time to perform the exercises between sessions. [Figure 2](#) shows the navigation pathway of mySARAH.

Table 2. Characteristics of participants taking part in mySARAH usability Cycles 1 and 2.

Characteristics	Cycle 1	Cycle 2
Number of participants (N=9)	5	4
Age (years), median (IQR ^a)	64 (60-66)	64.5 (55-70.5)
Male/Female	1/4	0/4
Educational qualification		
High school	1	1
Bachelor	1	2
Doctoral	3	0
Certificate course	0	1
Ethnicity		
White British, n (%)	4 (80)	4 (100)
Indian n (%)	1 (20)	0 (0)
Employment status n (%)		
Retired	3 (60)	2 (50)
Part-time	1 (20)	1 (25)
Full-time	1 (20)	0 (0)
Not working	0 (0)	1 (25)
Duration since diagnosis (years), median (IQR)	17 (12-25)	12 (4-26)
Internet use/day (hours), median (IQR)	3 (2-6)	1 (1-1.5)

^aIQR: interquartile range.

Table 3. Main usability issues and rectifications made in the online version of the Strengthening And stretching for Rheumatoid Arthritis of the Hand (mySARAH) exercise intervention.

Usability issues	Rectifications
Cycle 1 (n=5)	
0-10 numerical pain scale was not clear	Verbal descriptors were added to the 0-10 pain scale
The background color was very plain	A pleasant blue background was added
A feedback report on the pain levels would be helpful	A graph feature to provide a visual feedback on the pain levels recorded during every session was added
There was no separate section for email reminders	A separate section with an option for selecting the frequency of reminders was added
A separate patient video demonstrating wrist backward bends exercise would be helpful	An additional patient video was added
The size of some images was too small	Small images were increased in size
The progression bar across sessions was not noticeable	Progression bar was increased in width
Some pages were too wordy	Bullet points were used
A navigation tutorial video would be helpful	A preliminary navigation tutorial video was added
Cycle 2 (n=4)	
The instructions for completing forms/exercise calendar was not adequate	Clear instructions for completing forms/exercise calendar were added
There was too much scrolling in some pages	Page screen spaces were managed to reduce scrolling
Try more colors with pages: add 1 or 2 images on the landing page	A welcome image was added on the landing page
How do you know if a mistake was made on a form?	A pop-up message was set up to notify any omission or error prior to submission

Table 4. Questionnaire scores of the usability testing (N=9) for the online version of the Strengthening And stretching for Rheumatoid Arthritis of the Hand (mySARAH) exercise intervention.

Questionnaire	Cycle 1	Cycle 2
Computer System Usability Questionnaire items on 1-7 scale^a, median (IQR^b)		
1) Overall, I am satisfied with how easy it is to use this system	6.5 (6-7)	6.5 (6-7)
2) It was simple to use this system	6 (6-7)	6.5 (6-7)
3) I can effectively complete my work quickly using this system	5.8 (5.1-6.3)	6 (6-6.3)
4) I am able to complete my work quickly using this system	5.8 (5.1-6.3)	6 (6-6.3)
4) I am to efficiently complete my work using this system	6.5 (5.5-7)	6 (6-6)
6) I feel comfortable using this system	6 (6-7)	7 (6.8-7)
7) It was easy to learn to use this system	6 (6-7)	6.5 (6-7)
8) I believe I became productive quickly using this system	N/A ^c	N/A
9) The system gives error messages that clearly tell me how to fix problems	N/A	N/A
10) Whenever I make a mistake using this system, I recover easily and quickly	6 (5-6.5)	6 (5.3-6)
11) The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	6.5 (5.5-7)	7 (7-7)
12) It is easy to find the information I needed	6 (6-7)	6.5 (6-7)
13) The information provided for the system is easy to understand	6 (6-7)	7 (6.8-7)
14) The information is effective in helping me complete the tasks and scenarios	6 (6-7)	7 (6.8-7)
15) The organization of information on the system screens is clear	6 (6-7)	6.5 (6-7)
16) The interface of the system is pleasant	6 (6-7)	6 (6-7)
17) I like using the interface of this system	6.5 (5.8-7)	6 (5.8-6.3)
18) This system has all the functions and capabilities I expect it to have	6.5 (5.8-7)	6 (5.5-6.3)
19) Overall, I am satisfied with this system.	6.5 (6-7)	6 (6-6.3)
Likert scale of perceived usefulness, ease of use, and confidence, median (IQR)		
On a scale of 1-5, with 1 representing "Not at all useful" and 5 representing "Extremely useful," how would you rate the overall usefulness of mySARAH?	5 (4-5)	5 (4.8-5)
On a scale of 1-5, with 1 representing "Very difficult" and 5 representing "Very easy," how would you rate the overall ease of use of mySARAH?	4 (3-4)	4.5 (4-5)
On a scale of 1-5, with 1 representing "Not at all confident" and 5 representing "Very confident," how would you rate your confidence in doing the SARAH ^d exercises by yourself?	4.5 (4-5)	5 (5-5)

^a1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=neither, 5=somewhat agree, 6=agree, 7=strongly agree.

^bIQR: interquartile range.

^cN/A: not applicable.

^dSARAH: Strengthening And stretching for Rheumatoid Arthritis of the Hand.

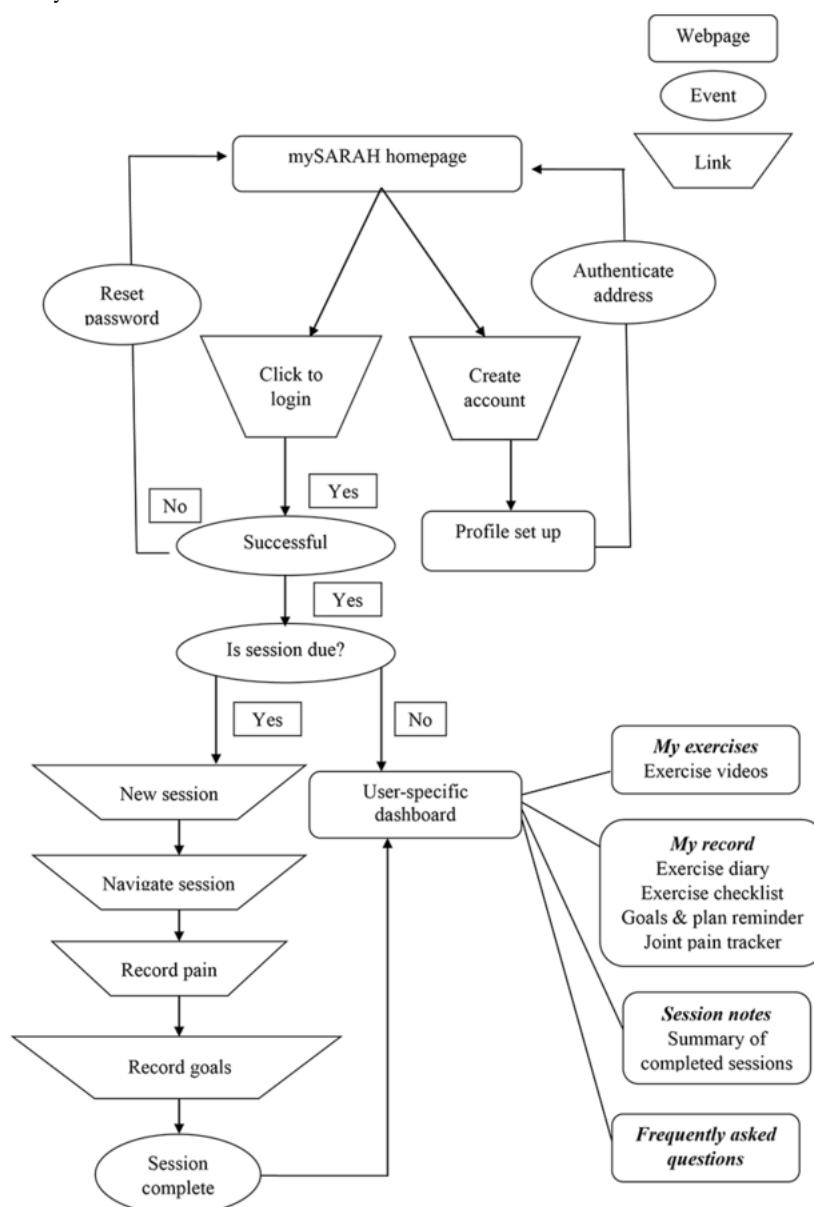
Table 5. Outline of mySARAH sessions' content.

Session	Suggested week of completion	Outline of content
1	Week 1	<ul style="list-style-type: none"> • Users fill out demographic information and hand function questionnaire • Users rate the pain in their hands on a 0-10 numerical scale • Information is provided about the clinical aspects of RA^a and its management • The SARAH^b mobility exercises are introduced • Users are taught how to set SMART^c goals and plan when and where to complete exercises. • Users are encouraged to complete the mobility exercises daily from this point onwards
2	Week 2	<ul style="list-style-type: none"> • The SARAH strength exercises are introduced • Users are taught baseline setting for strength exercises • Users review and update their goal and plan at the end of each session from this point onwards
3	Week 3	<ul style="list-style-type: none"> • The session covers how and when users should adjust their exercises if they: <ul style="list-style-type: none"> • Are finding them too challenging • Need to make them harder
4	Week 6	<ul style="list-style-type: none"> • The session encourages users to consider any barriers to completing their exercises, which have become apparent since beginning the program • It also asks users to think about how they have overcome barriers and what else they could do in the future
5	Week 9	<ul style="list-style-type: none"> • The session discusses the challenges to adhere to the program in the long-term • Users are taught how to restart the program if they need to stop for any reason
6	Week 12	<ul style="list-style-type: none"> • The session focuses on the continuation of the exercises after completion of the program. • Users are encouraged to continue to access the resources on the website if they need to • Users complete the Michigan hand function subscale, Global Rating of Change scale to measure their progress

^aRA: Rheumatoid arthritis.

^bSARAH: Strengthening And stretching for Rheumatoid Arthritis of the Hand.

^cSMART: Specific, Measurable, Attainable, Relevant, and Timely.

Figure 2. Navigation workflow of mySARAH.

Discussion

Principal Findings

The objectives of this study were to develop an online version of the evidence-based SARAH hand exercise program in collaboration with patients with RA (target users of mySARAH) and to identify and address its usability issues for a refined final version.

In general, patient contributors wanted a simple, less-cluttered, and less-wordy website. They were pleased with the purpose and content of the mySARAH prototype and found it a useful resource for people with hand function difficulties due to RA. Usability testing session participants also found mySARAH a useful and easy to use online exercise program and were confident to execute the SARAH exercises and set their baseline load for strength exercises on their own.

End User Involvement

We wanted to develop a user-centered mySARAH website and, hence, involved patient contributors in the initial phases of developing mySARAH. This is one of the very few web-based systems that were formally tested in adults with RA [21,23]. Users are expected to learn the SARAH exercises correctly, do them daily, and progress or regress the dosage according to their capabilities because mySARAH is self-directed in nature. Therefore, in addition to the usability testing of the website, we captured how far people managed to learn and perform the exercises correctly and set their baseline load for each strength exercise. We believed that evaluating this at these earlier stages would inform whether participants found the exercise demonstration videos easy to follow and replicate them, as intended.

Strengths of mySARAH Program

RA is a chronic condition; therefore, long-term adherence is required to maintain joint mobility and muscle strength. From

the findings of the SARAH trial, we know that participants found it challenging to continue the exercises long-term, and by 2 years follow-up, many ceased their exercises [24]. One of the major strengths of mySARAH is that it will allow users continued access to the SARAH exercises from home without the need for hospital appointments and, thus, overcoming practical problems such as transportation difficulties or lack of availability of hand therapy appointments. It would also give them access to videos and information if they need reminding of the exercises following a break. They can continue to use the exercise calendar to potentially promote long-term adherence to the program. The other strength is that the program was built upon a theoretical model incorporating heuristics principles toward a user needs-based intervention.

Limitations

This study has some limitations. Due to resource and time restraints, we did not transcribe the think-aloud sessions or code their content. The exercise demonstration evaluation scale used in this study was not a validated scale, but this evaluation was done by experienced physiotherapists and the scale was designed as a simple and pragmatic tool to record the evaluation. We

intended to include people from a wide range of educational and computer literacy backgrounds. However, the majority of the participants (8/9, 89%) were white British women, and most of them (6/9, 67%) had Bachelor to Doctoral levels of formal education. The median number of years since participants were diagnosed with RA was greater than 10 years in both the usability cycles. Therefore, we are uncertain whether men, people from other cultural or educational backgrounds, or those diagnosed more recently would offer a different perspective on mySARAH. Our sample size still meets the recommended number for testing usability issues.

Next Steps

We plan to carry out further testing in a proof-of-concept study to establish if people with RA are willing and able to complete the mySARAH program, do the exercises correctly, and undertake daily exercises

Conclusions

Involving target users in the development process ensured that mySARAH resulted in a user-centered and user-friendly online exercise resource. The results from the usability testing show mySARAH to be an efficient and user-friendly program.

Acknowledgments

This research was funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust. This work was supported by the NIHR Biomedical Research Centre, Oxford, UK. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. JH was funded by the ARUK Nurse and AHP research career development internship scheme, reference number 20867.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Brief description of mySARAH.

[PDF File (Adobe PDF File), 228KB - [jmir_v20i6e10457_app1.pdf](#)]

Multimedia Appendix 2

mySARAH Screenshots.

[PDF File (Adobe PDF File), 777KB - [jmir_v20i6e10457_app2.pdf](#)]

References

1. Heine PJ, Williams MA, Williamson E, Bridle C, Adams J, O'Brien A, SARAH Team. Development and delivery of an exercise intervention for rheumatoid arthritis: strengthening and stretching for rheumatoid arthritis of the hand (SARAH) trial. *Physiotherapy* 2012 Jun;98(2):121-130. [doi: [10.1016/j.physio.2011.03.001](#)] [Medline: [22507362](#)]
2. Adams J, Bridle C, Dosanjh S, Heine P, Lamb S, Lord J, et al. Strengthening and stretching for rheumatoid arthritis of the hand (SARAH): design of a randomised controlled trial of a hand and upper limb exercise intervention-ISRCTN89936343. *BMC musculoskeletal disorders* Dec 2012;13(1):230. [Medline: [23176133](#)]
3. Lamb S, Williamson E, Heine P, Adams J, Dosanjh S, Dritsaki M, StrengtheningStretching for Rheumatoid Arthritis of the Hand Trial (SARAH) Trial Team. Exercises to improve function of the rheumatoid hand (SARAH): a randomised controlled trial. *Lancet* 2015 Jan 31;385(9966):421-429. [doi: [10.1016/S0140-6736\(14\)60998-3](#)] [Medline: [25308290](#)]
4. NICE Guidance - Rheumatoid arthritis in adults: management. Rheumatoid arthritis in adults: management URL: <https://www.nice.org.uk/guidance/cg79/chapter/Recommendations> [accessed 2018-06-06] [WebCite Cache ID 6zyNEgbV1]

5. 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 Jun 23;8(2):e10 [FREE Full text] [doi: [10.2196/jmir.8.2.e10](https://doi.org/10.2196/jmir.8.2.e10)] [Medline: [16867965](https://pubmed.ncbi.nlm.nih.gov/16867965/)]
6. Jahangiry L, Farhangi M, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. *Public Health* 2017 Nov;152:36-46. [doi: [10.1016/j.puhe.2017.06.005](https://doi.org/10.1016/j.puhe.2017.06.005)] [Medline: [28734170](https://pubmed.ncbi.nlm.nih.gov/28734170/)]
7. Ludden G, van RT, Kelders S, van GJ. How to increase reach and adherence of web-based interventions: a design research viewpoint. *Journal of medical Internet research* Jul 2015;17(7):a.
8. Levac D, Glegg S, Camden C, Rivard L, Missiuna C. Best practice recommendations for the development, implementation, and evaluation of online knowledge translation resources in rehabilitation. *Phys Ther* 2015 Apr;95(4):648-662. [doi: [10.2522/ptj.20130500](https://doi.org/10.2522/ptj.20130500)] [Medline: [25301966](https://pubmed.ncbi.nlm.nih.gov/25301966/)]
9. Battles J. Improving patient safety by instructional systems design. *BMJ Quality & Safety* 2006;15(suppl 1):25-29.
10. Baumel A, Muench F. Heuristic evaluation of ehealth interventions: establishing standards that relate to the therapeutic process perspective. *JMIR mental health* Jan 2016;3(1):E [FREE Full text] [doi: [10.2196/mental.4563](https://doi.org/10.2196/mental.4563)] [Medline: [26764209](https://pubmed.ncbi.nlm.nih.gov/26764209/)]
11. Rogers M, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. *Journal of medical Internet research* Mar 2017;19(3):1-28.
12. Mohr D, Schueller S, Montague E, Burns M, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. *Journal of medical Internet research* Jun 2014;16(6):a.
13. Crutzen R. The behavioral intervention technology model and intervention mapping: the best of both worlds. *Journal of medical Internet research* Aug; 16 (8) 2014:1-3.
14. Drupal-Open Source CMS. 2018-03-19 URL: <https://www.drupal.org/> [accessed 2018-06-06] [WebCite Cache ID 6zyc7eAFI]
15. Srikesavan C, Williamson E, Eldridge L, Heine P, Adams J, Cranston T, et al. A Web-Based Training Resource for Therapists to Deliver an Evidence-Based Exercise Program for Rheumatoid Arthritis of the Hand (iSARAH): Design, Development, and Usability Testing. *Journal of medical Internet research* Dec 2017;19(12):1-14 [FREE Full text]
16. Gunning- Fog Index. 2018-03-19 URL: <http://gunning-fog-index.com/> [accessed 2018-06-06] [WebCite Cache ID 6zycINiHs]
17. Adams J, Chapman J, Bradley S, Ryan S. Literacy levels required to complete routinely used patient-reported outcome measures in rheumatology. *Rheumatology (Oxford)* 2013;52(3):460-464.
18. Virzi R. Human Factors. 1992. Refining the test phase of usability evaluation: how many subjects is enough URL: <https://doi.org/10.1177/001872089203400407> [accessed 2018-06-06] [WebCite Cache ID 6zycJitNU]
19. User testing. 2018-03-19 URL: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [accessed 2018-06-06] [WebCite Cache ID 6zycnfbRB]
20. Payne AY, Surikova J, Liu S, Ross H, Mechetiuc T, Nolan RP. Usability Testing of an Internet-Based e-Counseling Platform for Adults With Chronic Heart Failure. *JMIR Hum Factors* 2015 May 08;2(1):e7 [FREE Full text] [doi: [10.2196/humanfactors.4125](https://doi.org/10.2196/humanfactors.4125)] [Medline: [27026267](https://pubmed.ncbi.nlm.nih.gov/27026267/)]
21. Li L, Adam P, Townsend A, Lacaille D, Yousefi C, Stacey D, et al. Usability testing of ANSWER: a web-based methotrexate decision aid for patients with rheumatoid arthritis. *BMC Med Inform Decis Mak* 2013 Dec 01;13:131 [FREE Full text] [doi: [10.1186/1472-6947-13-131](https://doi.org/10.1186/1472-6947-13-131)] [Medline: [24289731](https://pubmed.ncbi.nlm.nih.gov/24289731/)]
22. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 1995 Jan;7(1):57-78. [doi: [10.1080/10447319509526110](https://doi.org/10.1080/10447319509526110)]
23. Nikolaus S, Bode C, Taal E, Vonkeman H, Glas C, van DLM. Acceptance of new technology: a usability test of a computerized adaptive test for fatigue in rheumatoid arthritis. *JMIR Human Factors* Jul 2014;1(1):a.
24. Williamson E, McConkey C, Heine P, Dosanjh S, Williams M, Lamb S. Hand exercises for patients with rheumatoid arthritis: an extended follow-up of the SARAH randomised controlled trial. *BMJ Open* 2017 Dec 12;7(4):e013121 [FREE Full text] [doi: [10.1136/bmjopen-2016-013121](https://doi.org/10.1136/bmjopen-2016-013121)] [Medline: [28404610](https://pubmed.ncbi.nlm.nih.gov/28404610/)]

Abbreviations

ADDIE: Analysis, Design, Development, Implementation, and Evaluation

BIT model: Behavioural Intervention Technology Model

mySARAH: Online version of the Strengthening And stretching for Rheumatoid Arthritis of the Hand exercise intervention

NHS: National Health Service

NICE: National Institute for Health and Care Excellence

RA: Rheumatoid Arthritis

SARAH: Strengthening And stretching for Rheumatoid Arthritis of the Hand

SMART: Specific, Measurable, Achievable, Relevant, and Timely

Edited by G Eysenbach; submitted 21.03.18; peer-reviewed by S Kardeş, T Jiang; comments to author 11.04.18; revised version received 16.04.18; accepted 24.04.18; published 27.06.18.

Please cite as:

Srikesavan C, Williamson E, Cranston T, Hunter J, Adams J, Lamb SE

An Online Hand Exercise Intervention for Adults With Rheumatoid Arthritis (mySARAH): Design, Development, and Usability Testing
J Med Internet Res 2018;20(6):e10457

URL: <http://www.jmir.org/2018/6/e10457/>

doi: [10.2196/10457](https://doi.org/10.2196/10457)

PMID: [29950288](https://pubmed.ncbi.nlm.nih.gov/29950288/)

©Cynthia Srikesavan, Esther Williamson, Tim Cranston, John Hunter, Jo Adams, Sarah E Lamb. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 27.06.2018. 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.

Original Paper

Web-Based Intervention for Family Carers of Persons with Dementia and Multiple Chronic Conditions (My Tools 4 Care): Pragmatic Randomized Controlled Trial

Wendy Duggleby¹, RN, PhD; Jenny Ploeg², RN, PhD; Carrie McAiney³, PhD; Shelley Peacock⁴, RN, PhD; Kathryn Fisher², PhD; Sunita Ghosh^{5,6}, PStats; Maureen Markle-Reid², RN, PhD; Jennifer Swindle¹, PhD; Allison Williams⁷, PhD; Jean AC Triscott⁸, CCFP (COE), FFAFP, FCFP; Dorothy Forbes⁹, RN, PhD; Kathya Jovel Ruiz¹, RN

¹Faculty of Nursing, University of Alberta, Edmonton, AB, Canada

²School of Nursing, McMaster University, Hamilton, ON, Canada

³Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada

⁴College of Nursing, University of Saskatchewan, Saskatoon, SK, Canada

⁵Department of Medical Oncology, University of Alberta, Edmonton, AB, Canada

⁶Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada

⁷School of Geography & Earth Sciences, McMaster University, Hamilton, ON, Canada

⁸Department of Family Medicine, Glenrose Rehabilitation Hospital, Edmonton, AB, Canada

⁹School of Nursing, University of Western Ontario, London, ON, Canada

Corresponding Author:

Wendy Duggleby, RN, PhD

Faculty of Nursing

University of Alberta

Level 3, ECHA, 11405 87 Avenue

Edmonton, AB, T6G 1C9

Canada

Phone: 1 780 492 6764

Email: wendy.duggleby@ualberta.ca

Abstract

Background: My Tools 4 Care (MT4C) is a Web-based intervention that was developed based on the transitions theory. It is an interactive, self-administered, and portable toolkit containing six main sections intended to support carers of community-living persons with Alzheimer's disease and related dementia and multiple chronic conditions through their transition experiences.

Objective: The objective of our study was to evaluate the effectiveness of MT4C with respect to increasing hope, self-efficacy, and health-related quality of life in carers of community-living older persons with Alzheimer's disease and related dementia and multiple chronic conditions.

Methods: A multisite, pragmatic, mixed methods, longitudinal, repeated-measures, randomized controlled trial was conducted between June 2015 and April 2017. Eligible participants were randomized into either treatment (MT4C) or educational control groups. Following baseline measures, carers in the treatment group received 3 months of password-protected access to MT4C. Trained research assistants collected data from participants via phone on hope (Herth Hope Index [HHI]), self-efficacy (General Self-Efficacy Scale), and health-related quality of life (Short Form-12 item [version 2] health survey; SF-12v2) at baseline, 1, 3, and 6 months. The use and cost of health and social services (Health and Social Services Utilization Inventory) among participants were measured at baseline, 3, and 6 months. Analysis of covariance was used to identify group differences at 3 months, and generalized estimating equations were used to identify group differences over time.

Results: A total of 199 carers participated in this study, with 101 participants in the treatment group and 98 in the educational control group. Of all, 23% (45/199) participants withdrew during the study for various reasons, including institutionalization or death of the person with dementia and lack of time from the carer. In the treatment group, 73% (74/101) carers used MT4C at least once over the 3-month period. No significant differences in the primary outcome measure (mental component summary score from the SF-12v2) by group or time were noted at 3 months; however, significant differences were evident for HHI-factor 2 ($P=.01$), with higher hope scores in the treatment group than in the control group. General estimating equations showed no

statistically significant group differences in terms of mental component summary score at all time points. Attrition and the fact that not all carers in the treatment group used MT4C may explain the absence of statistically significant results for the main outcome variable.

Conclusions: Despite no significant differences between groups in terms of the primary outcome variable (mental component score), the significant differences in terms of one of the hope factors suggest that MT4C had a positive influence on the lives of participants.

Trial Registration: ClinicalTrials.gov NCT02428387; <https://clinicaltrials.gov/ct2/show/NCT02428387> (Archived by Webcite at <http://www.webcitation.org/708oFCR8h>).

(*J Med Internet Res* 2018;20(6):e10484) doi:[10.2196/10484](https://doi.org/10.2196/10484)

KEYWORDS

carers; dementia; treatment; internet

Introduction

Web-based interventions have become increasingly popular as a means to support family and friend caregivers (hereafter carers) because of their flexibility and ease of access [1,2]. The importance of support for carers of persons with Alzheimer's disease and related dementias (ADRD) has been acknowledged as a worldwide issue given the growing numbers of persons with ADRD and the recognition that the majority of their care is provided by carers [3]. The complexity of the care provided by carers is further complicated when the person with ADRD lives with the carer, in the community, and has multiple chronic conditions (MCC) [4]. As a result of caregiving, carers undergo significant life-altering transitions, such as changes in roles and relationships that can have a negative impact on their quality of life (QOL) [5]. Transitions are significant changes that individuals need to incorporate into their lives to obtain positive health outcomes [6]. The lack of resources to support carers through these transitions is compounded by the 24/7 nature of caregiving for persons with ADRD, which makes accessing resources difficult. Web-based interventions show promise because they can be accessed at times and places that are convenient for carers [2,7].

Three systematic reviews of Web-based interventions for carers of persons with ADRD [2,8,9] reported that interventions resulting in improved carer health outcomes, for example, reduced anxiety and stress, had the following features: (1) they could be individually tailored by incorporating choices in different parts of the intervention [10-12], (2) they offered multiple components [10,13,14], and (3) they were psychoeducational interventions [10,12,15-20]. Most of the reviewed studies were pilot studies with the authors recommending future research using pragmatic randomized controlled trial (RCT) designs to evaluate Web-based interventions. The most common theoretical foundations for the interventions were stress and burden theories, with a focus on strain and depression [21]. However, these theories do not address the multiple, complex, reoccurring, and significant life-altering changes and the processes of the changes (transitions) that carers experience throughout their caregiving experiences [21,22].

Based on an adaptation of Meleis' theory of transition [6] and on our previous research that focused on transitions, hope, and

QOL [5], a self-administered, Web-based intervention titled My Tools 4 Care (MT4C) [23] was developed in partnership with the Alzheimer's Society of Alberta and Northwest Territories to support carers during their transition experiences. Transition experiences are the processes triggered by significant changes and involve carers acknowledging the changes, connecting with others, and redefining their perception of normal [22]. Redefining the perceptions of what is normal results in decreased stress and increased hope and makes one feel confident [24]. As part of the development of MT4C, Duggleby et al's [22] adaptation of Meleis' transition theory, involving core concepts of acknowledging the situation, connecting, and redefining normal, was mapped to the specific components of MT4C (for more information see Duggleby et al) [22].

The Web developers for this intervention were ATMIST [25]. A hard copy version of MT4C was pilot-tested by 20 carers of persons with ADRD, who found it feasible, acceptable, and potentially able to support them through transitions [22]. The principles that guided the development of the Web-based MT4C intervention were as follows: (1) inclusion of choice, the carers choose which sections they would like to use and when; (2) encouragement of user-generated content, carers can write in sections, add stories, pictures, music, etc; (3) portability, available on the Web, tablet, or mobile phone; and (4) privacy of information, only the carers can view their entries and share the contents.

The purpose of this pragmatic RCT was to evaluate the effectiveness of MT4C with respect to increasing health-related quality of life (HRQOL), self-efficacy, and hope among carers of older adults with ADRD and MCC in the community. Participants in the treatment group used MT4C for 3 months. We hypothesized that participants using MT4C would have increased hope, improved general self-efficacy, and increased HRQOL scores at 3 months compared with those at baseline and compared with those exhibited by an educational control group, at no additional cost. The following research questions were addressed in this study:

1. Does use of MT4C result in a 3-month (immediately post intervention) and 6-month (3 months post intervention) increase in HRQOL, self-efficacy, and hope in carers of persons with ADRD and MCC compared with that in an educational control group?

2. Are the effects or benefits of MT4C achieved at no additional cost in the treatment group compared with that in an educational control group?

Methods

Design

A detailed protocol for this study has been published elsewhere [26]; thus, only a summary has been provided here. This trial has been reported in accordance with the CONSORT-EHEALTH checklist [27]. The study design was a multisite, pragmatic, mixed-methods, longitudinal, repeated-measures, RCT. Baseline data were collected, followed by random assignment to a treatment (MT4C) or educational control group. Measures were repeated at 1, 3, and 6 months. As this was a mixed-methods concurrent study with a predominately quantitative design [28], qualitative data (semistructured interviews) were collected concurrently with, and were used to inform, the quantitative data. Quantitative and qualitative data were integrated at the results stage. Data quality was checked throughout the study by research assistants and at monthly meetings by the research team.

There were no changes to the content of MT4C, bug fixes, or unexpected events in association with MT4C use during this study. The study received ethical approval from the University of Alberta Health Research Ethics Board (# Pro0004872) and the Hamilton Integrated Research Ethics Board (#15-309).

Recruitment and Participants

Participant recruitment occurred offline over a 2-year period in 2 Canadian provinces (Ontario and Alberta) using multiple strategies. Trained research assistants targeted local branches of the Alzheimer's Society, in both provinces, and attended education groups for carers and shared study-related information, including its purpose and inclusion criteria. In addition, staff at the Alzheimer's Society and coordinators at community-based carer support groups, geriatric outpatient or memory clinics, adult day programs, and senior support services were provided with recruitment materials (eg, brochures or postcards). Staff members from these groups approached potential participants and obtained their consent to be contacted by the research team. In Alberta, advertisements in local community newspapers requested that interested carers contact the research coordinator using a toll-free number or via email.

Research assistants contacted interested carers via phone to screen for eligibility and to schedule the first interview. Participants were considered eligible to participate in the study if they were above the age of 18 years and were providing physical, emotional, or financial care for a community-living care recipient aged 65 years or older who had AD/DR and two or more chronic conditions. The participants were all English speaking and either a family or friend of the care recipient, with access to a computer and a valid email address. Exclusion criteria included non-English speaking carers and those caring for a family member or friend who was under the age of 65 years, who did not have AD/DR and MCC, or who was not a community-living care recipient.

Randomization

Allocation of participants into treatment and educational control groups was achieved using a 1:1 ratio. A biostatistician, not involved in recruitment, generated group allocations using stratified permuted block randomization. Random number sequences were fed into RedCap, a secure, password-protected, Web-based randomization service offered at the University of Alberta, which allocated clients to the two groups according to a random sequence.

Blinding

Given the nature of the study, the research team was unblinded to group allocation. Recruitment materials referred to evaluating different strategies to help carers and did not mention MT4C. Furthermore, to prevent contamination, participants were asked to keep the information about what group they were in confidential. To prevent participants from identifying the group to which they were allocated, two different consent forms were created, one for each group. These were used to obtain verbal telephone informed consent from each study participant. The consent form also acknowledged the potential risks that participants might encounter during the study (see [Multimedia Appendix 1](#)). Immediately following the first interview, participants received an email from the research assistants containing information on the study along with copies of the consent form and data collection tools. Participants allocated to the treatment group were provided with password-protected, no-cost access to MT4C.

Data collection occurred from June 2015 to April 2017. [Figure 1](#) outlines the data collection procedures, which are available in more detail in the study protocol article [24]. Trained research assistants, in each province, completed audiotaped telephone interviews (quantitative and qualitative) that lasted anywhere from 15 to 60 min.

Intervention

Participants in the treatment group were instructed to access MT4C at their convenience on a computer, tablet, or mobile phone for 3 months. Once the carer logged on to the site, the first page ("How to use MT4C") provided instructions on how to use MT4C and contained a menu outlining the sections constituting the toolkit. MT4C consists of six main sections: (1) about me, (2) common changes to expect, (3) frequently asked questions, (4) resources, (5) important health information, and (6) calendar. In the *About Me* section, participants have the option to add formatted text, pictures, and PDF files (see [Multimedia Appendix 2](#) for screenshots of all MT4C pages). All data entered by participants into the site remained confidential, even from the study team. Participants also received an electronic copy of the Alzheimer's Society's *The Progression of Alzheimer's Disease* booklet [29], a copy of the study questionnaires, and the MT4C toolkit checklist intended for participants to record their use of the MT4C site.

Educational Control Group

Participants in the educational control group (usual care) received a copy of the Alzheimer's Society's *The Progression of Alzheimer's Disease* booklet, via email, after the first interview. This booklet is freely available through the

Alzheimer's Society of Canada and provides information for the person with dementia, his or her family, and carers on the stages of Alzheimer's disease. At the end of data collection, participants in the control group received an email containing a one-page summary of the study, including preliminary findings, and instructions on how to contact the research team if they wished to have access to the MT4C site.

Measures

Demographics

All participants completed a demographic questionnaire during the baseline interview. Information collected included age, gender, marital status, ethnicity, citizenship, level of education, employment status, occupation, income, relationship to the person with ADRD, carer-specific chronic health conditions, and length of time spent caregiving. This form also collected information on the care recipients' age, gender, and number of chronic conditions.

Toolkit Checklist

Participants accessing the MT4C site were asked to document, offline, the frequency and amount of time (in minutes) they spent on each section of the site. Research assistants reviewed this information with participants at the 1- and 3-month interviews. At the 1-month interview, participants were reminded that at 3 months, their access to MT4C would end, and they were encouraged to use the site, if not done so already, and continue completing the checklist.

Primary Outcome Measure: Short Form-12 Item (version 2) Health Survey

All participants completed the Short Form-12 item (version 2) health survey (SF-12v2) at all time points. It is a widely used measure of HRQOL consisting of 12 questions, measuring 8 domains of functioning and well-being (physical functioning, role functioning, bodily pain, general health, vitality, social functioning, emotional health, and mental health) [30,31]. Overall scores are summarized in 2 domains: a physical component summary score (PCS) and a mental component summary score (MCS). Scores range from 0 to 100, with higher scores indicating a better HRQOL. SF-12v2 is a reliable tool with estimated PCS and MCS test-retest reliabilities of $r=0.89$ and $r=0.86$, respectively [32]. MCS was selected as the primary outcome for this study, given the psychoeducational nature of the intervention.

Secondary Outcome Measures: SF-12v2 PCS, General Self-Efficacy Scale, and Herth Hope Index

Secondary outcome measures included the SF-12v2 PCS score (described above) and the General Self-Efficacy Scale (GSES) and Herth Hope Index (HHI) scores. Participants completed the GSES and HHI at all data collection time points. GSES was used to assess participants' perceived self-efficacy or belief that they can complete novel or difficult tasks or cope with diversity [33]. It is a 10-item, 4-point scale with a Cronbach's alpha coefficient of reliability r ranging from 0.76 to 0.90 ($P<.05$). Total scores range from 10 to 40, with higher scores indicating a greater level of self-efficacy. GSES has been used in countries around the world and has been adapted to 26 languages [33].

HHI is a 12-item Likert-type scale [34]. Items are scored from 1 "strongly disagree" to 4 "strongly agree." Total scores range from 12 to 48, with higher scores indicating a greater hope. The items in the scale can be grouped into three factors of hope: (1) temporality and future, (2) positive readiness and expectancy, and (3) interconnectedness. HHI has been used in a variety of populations and has a test-retest reliability of 0.91 ($P<.05$) and criterion-related validity r of 0.81 to 0.92 ($P<.05$) [34].

Health and Social Services Utilization Inventory

A modified version of the Health and Social Services Utilization Inventory (HSSUI) was used to collect service use information to calculate costs from a societal perspective [35] from all study participants at baseline, 3 months, and 6 months. The initial version of HSSUI was developed by Browne and colleagues and has been shown to have good reliability and validity ($r=0.72$ to 0.99) [36,37]. This survey measures the use of services by asking respondents to think back over a specific period of time (here, 3 months) about the type of health and social services accessed, number of times used, and any out-of-pocket expenses related to these services. HSSUI was developed for persons with different types of illnesses. For our study with carers, it was modified to reflect service use such as accessing the Alzheimer's Society.

Qualitative Interviews

All participants were interviewed at 1, 3, and 6 months using semistructured, audiotaped telephone interviews. Questions were asked about the significant changes they had experienced as carers as well as what had helped them deal with these changes. At 3 and 6 months, participants using MT4C were asked about the following: (1) their perceptions of MT4C, (2) how MT4C helped them deal with transitions, (3) what they liked most and least about MT4C, and (4) changes they would make to MT4C.

Data Analysis

Primary and Outcome Measures

All quantitative data were entered in SPSSv24 and cleaned and checked by a research assistant for accuracy. SAS version 9.4 was used for all statistical analyses. Statistical tests assumed a 0.05 two-tailed level of significance and 95% CIs.

Baseline characteristics data are presented as means and SDs for continuous variables and as numbers and percentages for categorical variables. Analysis of covariance (ANCOVA) was used to test the differences in outcome variables between the intervention and control groups at 3 months. The 3-month analysis represents the primary analysis for this study because this period corresponds to the duration of the intervention. Separate ANCOVA models were run for each outcome, with the 3-month outcome as the dependent variable, group (intervention, control) as the independent variable, and baseline value of the outcome as the covariate, rather than determining differences between groups at baseline (CONSORT guidelines) [38]. Intention-to-treat principles were used in all analyses; thus, all participant data were analyzed in the groups in which they were originally allocated (including those in the treatment group who did not utilize MT4C); imputation was applied to address

missing data. Multiple imputation is considered the best method for addressing the most common and realistic missing data patterns seen in RCTs [39]. We performed multiple imputations using the general procedure and employing the fully conditional specification procedure with predictor mean matching [40]. A range of auxiliary and outcome variables were used in the imputation model to improve accuracy; thus, the imputation model included baseline variables (age and gender of carer, number of carer chronic conditions, and number of care recipient chronic conditions) as well as the secondary outcome variables. After missing data were imputed (five imputations), each dataset was analyzed using ANCOVA, and the results from these multiple analyses were pooled to obtain an overall inference. Sensitivity analyses were performed using the complete case dataset.

Subgroup analyses were performed to determine whether the differences in intervention effectiveness for the primary outcome were observed for specific baseline groups. These were restricted to the following six baseline factors: age, gender, carer employment status, number of carer chronic conditions, number of care recipient chronic conditions, and income. Subgroup differences in the intervention effect were determined based on the significance of the group or subgroup interaction term in the ANCOVA model.

Generalized estimating equations (GEE) were used to determine any group differences over the 6-month period. GEE was selected because it is a robust method that does not rely on normality assumptions to address the dependency in repeated-measures data [41]. This 6-month analysis is regarded as a secondary or supplemental analysis in this study because the 3- to 6-month period corresponds to a time when the intervention was no longer available to treatment group participants. Separate GEE models were run for each outcome (primary, secondary). GEE models included group (intervention, control), time, and group \times time interaction; the group \times time interaction was of primary interest because statistical significance for this variable indicates the presence of a treatment effect [41].

Cost analyses were conducted to compare the cost of health service use in the intervention and control groups. Neither the intervention nor control group had program-specific costs. The service use that clients reported, using HSSUI at baseline and at 3 months (end of intervention period), was multiplied by the unit costs for the service to obtain total service costs. Unit costs were obtained from Ontario and Alberta databases, which provide the costs of all services paid for by the publicly-funded health care systems in each province. Cost data are substantially positively skewed and have traditionally been handled using nonparametric methods [42]. Mann-Whitney U-test was used to evaluate the between-group differences in median costs and to compare the total service costs at baseline and 3 months between the two groups.

Qualitative Interview Data

Qualitative data were analyzed using a qualitative descriptive approach in which coding categories were derived directly from the data [43]. Each transcript was read as a whole looking for

similarities, differences, and patterns. Categories were grouped into themes. Trustworthiness of the data was maintained using the words of participants as much as possible and keeping an audit trail of analytic decisions. The findings from the qualitative analysis were integrated with the quantitative findings; an in-depth analysis of the qualitative findings for this study is reported elsewhere [44].

Results

Participants

Recruitment began on May 2015; a total of 382 persons were contacted, resulting in 199 carers of community-living persons with ADRD and MCC who participated in the study (Ontario: 106/199, 53.3% and Alberta: 93/199, 46.7%; [Figure 1](#)). Participants were randomly assigned to either treatment (n=101) or control (n=98) group. Of all participants, 22.6% (45/199) withdrew during the study ([Figure 2](#)), and a total of 154 (154/199, 77.4%) carers completed the study (treatment group: 73/154, 47.4%; control group, 81/154, 52.6%).

Participants (Baseline Characteristics)

[Table 1](#) shows participants' baseline characteristics. Randomization resulted in no significant differences between the two groups. Most carers in the study (161/199, 80.9%) were female, married or in a common law relationship (168/199, 84.4%), Caucasian (185/199, 93.0%), unemployed (111/199, 55.8%), and either the spouse (98/199, 49.2%) or adult child (91/199, 45.7%) of the person with dementia. Of all carers, 69.3% (138/199) resided with the person with dementia and 68.8% (137/199) received assistance with caregiving. On average, carers in the study were 63.6 (SD 11.6) years of age and had been carers for approximately 4.3 (SD 4.2) years. Care recipients in the study were on average 80.3 (SD 7.7) years of age and had an average of 10.0 (SD 4.1) chronic conditions ([Table 1](#)).

Intervention "Dose" (Use of My Tools 4 Care)

In the intervention group, 73% (74/101) carers used MT4C at least once over the 3-month period. At 1 month, participants who used the site spent the most time on Section 1 *My story* (median 9.5 min, interquartile range [IQR] 28.4). By 3 months, participants spent most of their time on Section 2 *Common changes to expect* (median 15 min, IQR 45.0) and Section 4 *Resources* (median 10.00 min, IQR 20.0).

Intervention Effectiveness

The results of the multiple imputation ANCOVA testing for group differences in mean changes in the outcome scores from baseline to 3 months are provided in [Table 2](#). No significant group differences were observed in the primary or secondary outcome measures. A significant group difference was observed for factor 2 of HHI, although no difference was observed for the overall HHI. The complete case ANCOVA results were consistent with the multiple imputation ANCOVA results (data not shown). Given the absence of effects at 3 months and termination of the intervention use at this time, GEE was expected to show no overall effect for the 6-month period.

Figure 1. Number of Participants at Each Data Collection Period and Data Collection Procedure. GSES: General Self-Efficacy Scale; HHI: Herth Hope Index; HSSUI: Health and Social Services Utilization Inventory; MT4C: My Tools 4 Care; SF-12v2: Short Form-12 item health survey.

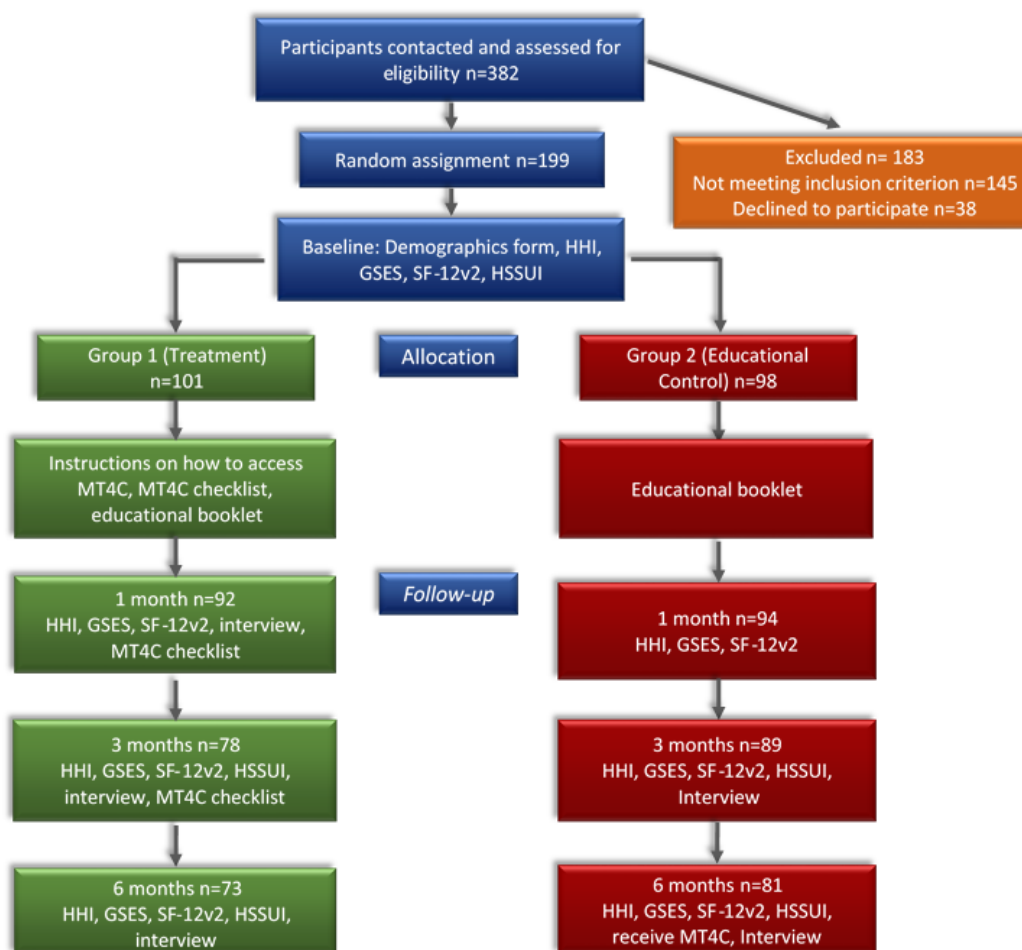


Figure 2. Number of Participant Withdrawals. ADRD: Alzheimer's disease and related dementias.

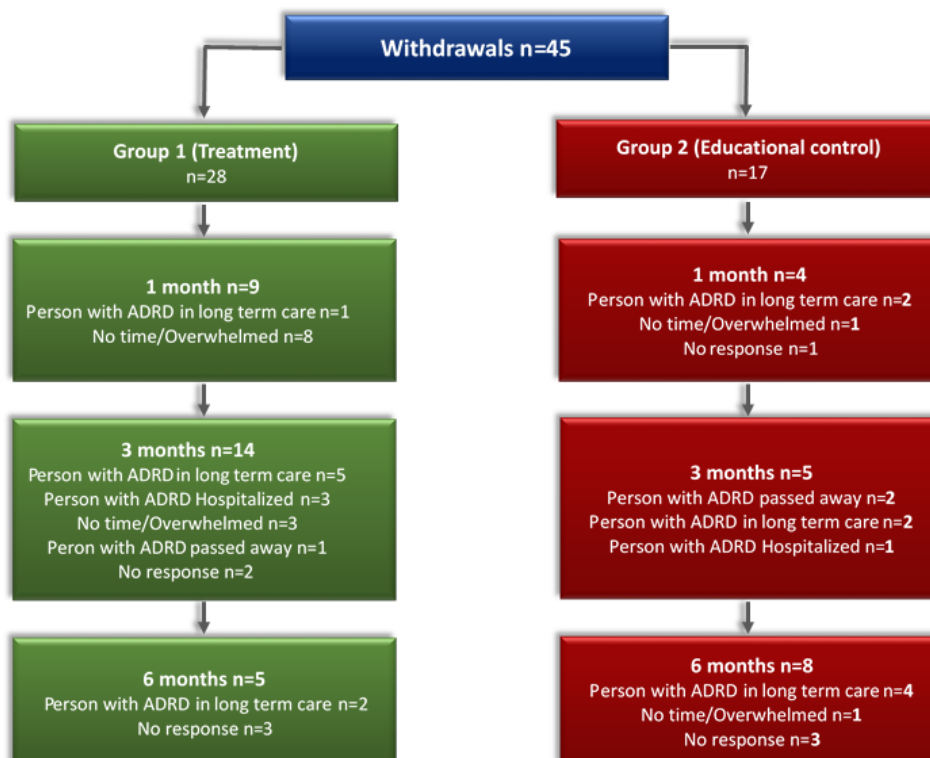


Table 1. Demographic characteristics of caregivers and care recipients.

Characteristic	Intervention (n=101)	Control (n=98)	Total Sample (N=199)
Carers			
Age, mean (SD)	63.4 (12.2)	63.9 (11.1)	63.6 (11.6)
Number of years caregiving, mean (SD)	4.1 (3.8)	4.6 (4.5)	4.3 (4.2)
Years of education, mean (SD)	14.1 (2.9)	14.27 (3.0)	14.18 (2.9)
Chronic conditions, mean (SD)	2.2 (1.5)	2.5 (1.6)	2.4 (1.6)
Gender, n (%)			
Male	22 (22)	16 (16)	38 (19)
Female	79 (78)	82 (84)	161 (81)
Marital status, n (%)			
Married or living with someone	84 (83)	85 (87)	168 (85)
Single, widowed, divorced or separated	17 (17)	13 (13)	30 (15)
Ethnicity, n (%)			
Caucasian	93 (92)	92 (94)	185 (93)
Other	8 (8)	6 (6)	13 (7)
Employed, n (%)			
Yes	39 (39)	47 (48)	86 (44)
No	61 (61)	50 (52)	111 (56)
Relationship to care recipient, n (%)			
Spouse or life partner	51 (50)	47 (48)	98 (49)
Son or daughter	47 (47)	44 (45)	91 (46)
Other	3 (3)	7 (7)	10 (5)
Living with care recipient, n (%)			
Yes	70 (69)	68 (69)	138 (69)
No	31 (30)	30 (31)	61 (31)
Finances meet needs, n (%)			
Completely or Very well or Adequately	81 (80)	76 (79.2)	157 (80)
With some difficulty or Not very well or Totally inadequate	20 (20)	20 (20.8)	40 (20)
Household income, n (%)			
Less than Can \$40,000	25 (29)	24 (30.4)	49 (30)
Can \$40,000 to \$70,000	23 (27)	16 (20.3)	39 (23)
Greater than Can \$70,000	38 (44)	39 (49.3)	77 (47)
Assistance with caring, n (%)			
Yes	70 (69)	67 (68)	137 (69)
No	31 (31)	31 (32)	62 (31)
Care Recipient			
Age, mean (SD)	80.5 (7.4)	80.2 (8.0)	80.3 (7.7)
Chronic conditions, mean (SD)	10.4 (4.1)	9.6 (4.0)	10.0 (4.1)
Gender, n (%)			
Male	55 (54)	49 (50)	104 (52)
Female	46 (46)	49 (50)	95 (48)

Table 2. ANCOVA results using multiple imputation (baseline to 3 months).

Outcome Measures	Pooled LSM ^a Group Difference (95% CI)	<i>P</i> value for Null Model (No Group Effect)
SF-12 v2 PCS ^b	−0.02 (−2.07 to 2.01)	.98
SF-12 v2 MCS ^c	−0.23 (−3.25 to 2.80)	.88
HHI^d	0.56 (−0.25 to 1.36)	.17
HHI-Factor 1	0.05 (−0.28 to 0.38)	.77
HHI-Factor 2	0.56 (0.11 to 1.01)	.01 ^e
HHI-Factor 3	0.11 (−0.25 to 0.47)	.55
GSES ^f	0.22 (−0.78 to 1.22)	.67

^aLSM:Least Square Means.^bSF-12 PCS: Short Form-12 item health survey physical component summary score.^cSF-12 MCS: Short Form-12 item health survey mental component summary score.^dHHI: Herth Hope Index.^eSignificant at $P \leq .05$.^fGSES: General Self-Efficacy Scale.**Table 3.** Generalized estimating equations with Short Form-12 item (version 2) mental component summary score as the dependent variable (all time points).

Parameter	Beta	SE	95% Wald CI	Hypothesis Test		
				Wald chi-square	<i>df</i> ^a	<i>P</i> value
Time point (Baseline)						
1 month	1.00	0.55	−0.07 to 2.08	3.32	1	.07
3 months	0.64	0.65	−0.64 to 1.9	0.92	1	.33
6 months	0.68	0.74	−0.76 to 2.12	0.86	1	.35
Caregiver gender (versus female)						
Male	3.99	1.20	1.63 to 6.36	11.02	1	.001 ^b
Financial needs met (versus no)						
Yes	1.06	1.02	−0.94 to 3.05	1.08	1	.30
Caregiver age	0.14	0.07	0.00 to 0.29	3.88	1	.049 ^b
Study group (versus control)						
Treatment	−0.32	0.97	−2.22 to 1.58	0.11	1	.74
Relationship to care recipient (versus other)						
Spouse	−1.58	1.53	−4.59 to 1.42	1.07	1	.30
Years in caregiver role	−0.07	0.13	−0.33 to 0.19	0.28	1	.60
GSES ^c total score	0.54	0.10	0.34 to 0.741	28.26	1	<.001 ^a
SF-12v2 PCS ^d	−0.26	0.04	−0.35 to −0.174	35.37	1	<.001 ^a
HHI total score	0.69	0.09	0.52 to 0.870	60.09	1	<.001 ^a

^a*df*: degrees of freedom.^bSignificant at $P \leq .05$.^cGSES: General Self-Efficacy Scale.^dSF-12v2 PCS: Short Form-12 item (version 2) physical component summary score.

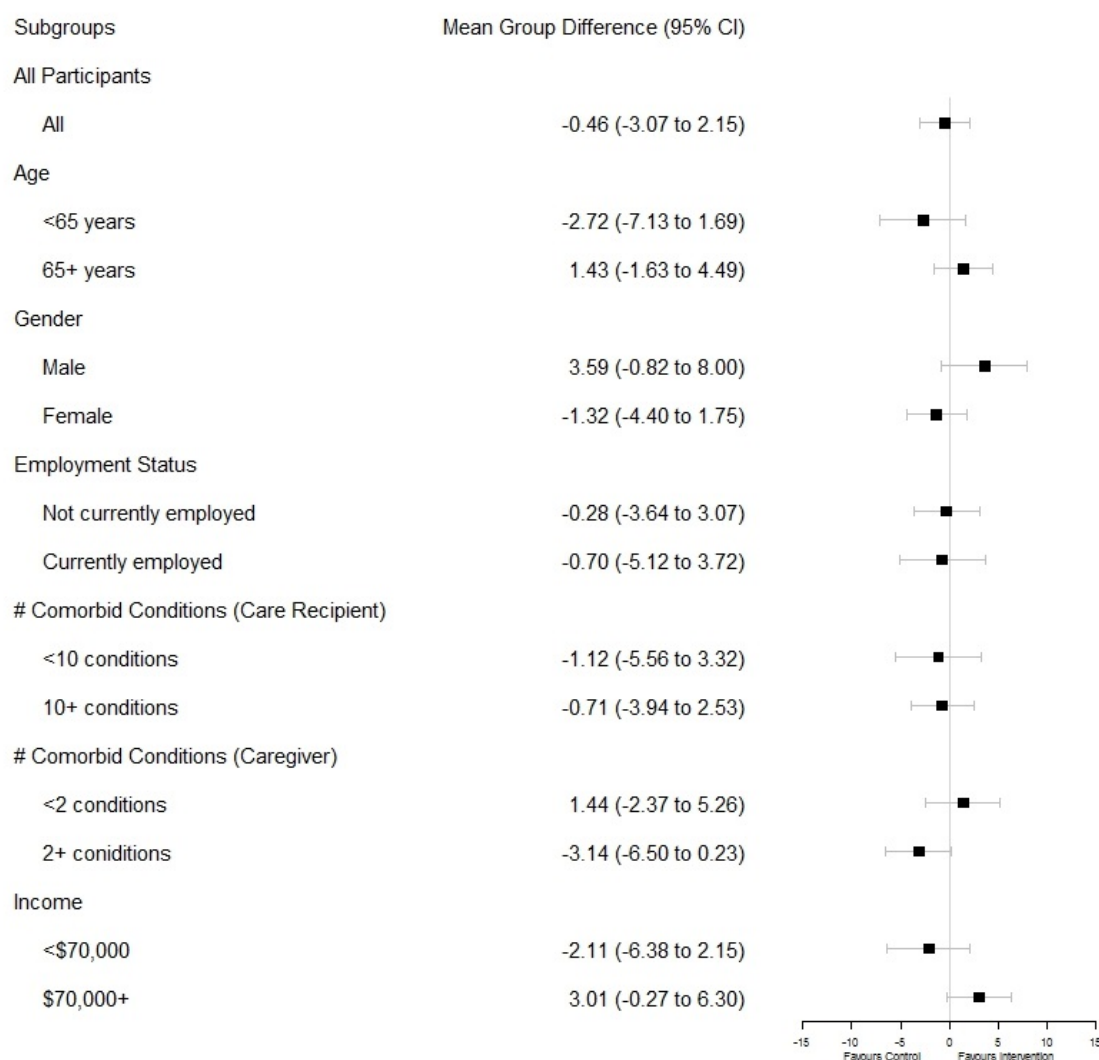
GEE indicated no statistically significant differences between groups over time in MCS scores (Table 3). Statistically significant variables in the MCS model were male sex ($P=.001$); older age (≥ 65 years; $P=.049$); and HHI ($P<.001$), GSES ($P<.001$) and SF-12v2 PCS ($P<.001$) scores.

Subgroup analysis results also indicated that there were no statistically significant findings for differences in MCS scores (Figure 3).

Although no statistically significant differences between the groups were observed in the primary and secondary outcomes or the subgroup analyses, participants in the treatment group indicated, when asked in interviews, that MT4C helped them with their transitions. For example, one participant said:

“Yeah, it does, because—but it’s how I—you know, I really wish I would have written those things down on the day that they had happened, you know, because it would give me something concrete to go back and see how the—how he digressed...”

Figure 3. Subanalysis.



Participants who felt MT4C did not help them with transitions suggested it was because they were already privy to sufficient resources. For example, one participant noted that MT4C “didn’t help me significantly...I had gone to some carers’ group and got some information there.”

Health and Social Service Costs

Table 4 provides the results of comparing the two groups in terms of the change in costs from baseline to 3 months. No statistically significant between-group differences were observed in the cost of individual health and social services or in overall service costs.

Table 4. Cost comparison: intervention versus control group (baseline vs 3 months).

Health and social services utilization inventory	Intervention (Can \$)			Control (Can \$)			Independent samples difference, Wilcoxon Mann–Whitney z -statistic (P value) ^a
	Baseline	3-month	Median cost	Baseline	3-month	Median cost	
	median (Q1, Q3)	median (Q1, Q3)	difference (Q1, Q3)	median (Q1, Q3)	median (Q1, Q3)	difference (Q1, Q3)	
Physicians (primary care & specialists)	185.00 (129.94, 340.89)	185.00 (132.95, 370.00)	0.00 (–77.20, 73.48)	185.00 (148.85, 281.71)	185.00 (77.20, 354.88)	0.00 (–123.33, 157.25)	–0.41 (.68)
Hospital & emergency department	523.02 (283.02, 703.54)	386.04 (283.02, 772.08)	–1726.06 (–3452.12, 0.00)	386.04 (283.02, 1158.12)	386.04 (386.04, 386.04)	772.08 (772.08, 772.08)	0.61 (.54)
Other health & social service providers	180.00 (87.86, 416.64)	242.87 (102.93, 455.90)	0.00 (–131.20, 131.79)	293.00 (94.87, 690.19)	220.69 (470.52, 111.05)	0.00 (–75.69, 150.96)	–0.03 (.98)
Laboratory services	24.51 (12.40, 109.44)	41.94 (6.74, 240.50)	–5.58 (–70.00, 25.05)	38.32 (16.45, 145.64)	19.15 (3.29, 114.63)	0.00 (–35.24, 18.18)	–0.93 (.35)
Prescription medications	80.77 (49.38, 243.82)	240.36 (130.54, 558.34)	–24.60 (–59.72, –10.38)	106.38 (47.43, 227.50)	78.08 (31.55, 134.64)	0.00 (0.00, 0.00)	–1.20 (.23)
Community support services	165.00 (91.43, 344.93)	206.15 (90.45, 346.25)	0.00 (–115.00, 93.66)	183.24 (121.38, 335.94)	178.50 (67.02, 483.35)	18.00 (–125.84, 167.50)	–0.37 (.71)
Other services	227.70 (109.98, 609.15)	203.05 (51.26, 450.49)	0.00 (–290.00, 189.98)	201.53 (60.00, 378.00)	226.53 (51.26, 406.10)	3.86 (–95.19, 116.42)	–0.23 (.82)
Total costs	751.20 (316.55, 1285.13)	587.35 (347.51, 1064.31)	32.21 (–319.38, 412.69)	640.92 (354.96, 1091.31)	659.51 (267.62, 1077.79)	129.16 (–380.16, 394.69)	–0.24 (.81)

^aThe hypothesis tested was there would be no group differences in median scores baseline–3 month.

Discussion

Principal Findings

The purpose of this study was to examine the influence of using a self-administered, multicomponent, Web-based intervention (MT4C) in increasing hope, general self-efficacy, and mental health among carers of community-living persons with ADRD and MCC. Despite there being no significant differences between the treatment (MT4C) and educational control groups in the primary or secondary outcome measures, the treatment group had significantly higher factor 2 hope scores than the control group at 3 months. Factor 2 on HHI is a subscale entitled “positive readiness and expectancy” and reflects the confidence of people in their ability to have a positive future [34]. Statements in this subscale include feelings that “there is a light at the end of the tunnel” and “I have a direction.” This increase in hope is consistent with our adapted transition theory, which suggests that when carers are able to redefine what they perceive as normal, they report increases in hope.

The findings also suggest that hope and general self-efficacy continue to be important variables influencing mental health among carers of community-living older persons with ADRD and MCC. MCC adds additional stress and complexity to the caregiving experience [4], and hope and general self-efficacy have been found to influence the carer’s HRQOL in studies of

carers with dementia [5]. As hope and general self-efficacy are significant variables influencing mental health, this finding suggests that the model for the intervention has promise and that activities within MT4C targeted at increasing hope and general self-efficacy should be strengthened. For example, the current activity focused on hope is entitled “Everyday hope” in which participants are asked to consider what would give them hope that day. To strengthen this activity, participants could be encouraged to also view a 15-min film entitled “Connecting with Hope” in which carers of persons with ADRD describe how they maintain hope. An activity focused on self-efficacy currently includes participants identifying their own inner strength. This activity could be strengthened by having participants identify what went well each day, to focus on the positive aspects of caregiving.

As a tailored intervention, instructions for use of MT4C suggest that participants use whatever sections they want, for as long as they want. If the treatment effect for MT4C is reliant upon increasing hope and general self-efficacy, a treatment effect might have been realized if the instructions required participants to focus on the activities specifically designed to increase hope and general self-efficacy in this population. Tailored interventions that consist of multiple components are complex. Moreover, pragmatic trials typically do not focus on mechanisms of action, but instead simply ask whether the intervention worked. For these reasons, we did not focus on determining

which component and how much of a component is needed to achieve a treatment effect [45]. Although we tried to capture how much and what sections of MT4C were used through self-report, a significant amount of data was missing. Future research should incorporate the measurement of time spent on individual components into the Web design.

Other Web-based interventions for family carers that reported statistically significant findings focused on outcomes such as anxiety, distress, and depression [1]. Other aspects of mental health and HRQOL might have been more sensitive to a treatment effect as a result of using MT4C. A review of multicomponent interventions for family carers of people with dementia suggests that changing HRQOL through interventions is difficult because HRQOL can quickly deteriorate in carers of persons with dementia [46]. Future evaluation of MT4C should target more specific outcomes, such as anxiety, distress, and depression.

The qualitative data suggested that MT4C helped some participants with the significant changes they experienced as carers. Those who did not feel that MT4C helped them indicated that it was because they were already receiving support from an Alzheimer's Society. As the majority of the participants were recruited from Alzheimer's Societies, this could have influenced the outcome of the study.

Limitations

This study has several limitations. Attrition over time resulted in a study cohort consisting of 166 participants at 3 months (post intervention), which is below the sample size required to determine significance. Although multiple strategies were used,

similar to another research, recruiting carers for research was difficult [47]. In addition, since this was a convenience sample, the generalizability of our findings is limited. Future research on MT4C should be conducted with larger sample sizes and should include a more random sampling approach. The participants were well-educated and had access to computers; however, 27% of the treatment group did not use MT4C during the 3-month period. Nonuse of Web-based interventions has been reported in another study [48]. Another limitation is that the majority of participants were recruited from Alzheimer's Societies and already had access to resources. There is an additional possibility, even with blinding of the consents, that there was contamination as the treatment group participants may have discussed MT4C with the control group participants. Future research should examine whether users of MT4C who are unable to attend Alzheimer's Society support groups can achieve significant improvements in their mental health. Finally, the limitation regarding participant use of MT4C is of concern. In future research, participants should be instructed more clearly regarding the importance of utilizing MT4C and potentially specific components within; moreover, keeping a track of the time spent on each component can be accomplished through website design strategies.

Conclusion

This study was unique because MT4C is focused on supporting carers of community-living older persons with ADRD and MCC. Furthermore, it is one of very few studies to include costs from a societal perspective. The findings contribute to future research designs for Web-based interventions with carers as well as future research with MT4C.

Acknowledgments

This study is part of a program of research (Aging, Community and Health Research Unit) supported by the Canadian Institutes of Health Research Signature Initiative in Community-Based Primary Healthcare (Funding Reference Number: TTF 128261). We wish to thank all of the study participants and our community partners: Alberta Caregivers Association, Alzheimer Society of Alberta and North West Territories, Alzheimer Society of Brant, Haldimand Norfolk, Hamilton and Halton, Alzheimer Society of Calgary, Alzheimer Society of Ontario, and Covenant Health Alberta. We also thank Tracy Chambers and data collectors: Ann Cotton, Vanessa Hang, Melissa Northwood, Lori Pollard, and Tracy Wildeman.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Letter of information and consent forms.

[PDF File (Adobe PDF File), 131KB - [jmir_v20i6e10484_app1.pdf](#)]

Multimedia Appendix 2

Screenshots MT4C.

[PDF File (Adobe PDF File), 1MB - [jmir_v20i6e10484_app2.pdf](#)]

Multimedia Appendix 3

CONSORT-EHEALTH checklist (V 1.6.1.).

References

1. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. *J Med Internet Res* 2017 Jul 28;19(7):e263 [FREE Full text] [doi: [10.2196/jmir.7564](#)] [Medline: [28754652](#)]
2. Boots LM, de Vugt ME, van Knippenberg RJ, Kempen GI, Verhey FR. A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. *Int J Geriatr Psychiatry* 2014 Apr;29(4):331-344. [doi: [10.1002/gps.4016](#)] [Medline: [23963684](#)]
3. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. Alzheimer's Disease International. World Alzheimer report: The global impact of dementia, an analysis of the prevalence, incidence, costs & trends URL: <https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf> [accessed 2018-05-14] [WebCite Cache ID 6zPonB6Nb]
4. Ploeg J, Matthew-Maich N, Fraser K, Dufour S, McAiney C, Kaasalainen S, et al. Managing multiple chronic conditions in the community: a Canadian qualitative study of the experiences of older adults, family caregivers and healthcare providers. *BMC Geriatr* 2017 Jan 31;17(1):40 [FREE Full text] [doi: [10.1186/s12877-017-0431-6](#)] [Medline: [28143412](#)]
5. Duggleby WD, Swindle J, Peacock S, Ghosh S. A mixed methods study of hope, transitions, and quality of life in family caregivers of persons with Alzheimer's disease. *BMC Geriatr* 2011 Dec 22;11:88 [FREE Full text] [doi: [10.1186/1471-2318-11-88](#)] [Medline: [22192235](#)]
6. Meleis A. *Transitions Theory: Middle Range and Situation Specific Theories in Nursing Research and Practice*. New York, NY: Springer Publishing Company; 2010.
7. Godwin KM, Mills WL, Anderson JA, Kunik ME. Technology-driven interventions for caregivers of persons with dementia: a systematic review. *Am J Alzheimers Dis Other Dement* 2013 May;28(3):216-222. [doi: [10.1177/1533317513481091](#)] [Medline: [23528881](#)]
8. Thompson CA, Spilsbury K, Hall J, Birks Y, Barnes C, Adamson J. Systematic review of information and support interventions for caregivers of people with dementia. *BMC Geriatr* 2007;7:18 [FREE Full text] [doi: [10.1186/1471-2318-7-18](#)] [Medline: [17662119](#)]
9. Topo P. Technology studies to meet the needs of people with dementia and their caregivers: a literature review. *Journal of Applied Gerontology* 2008 Oct 01;28(1):5-37. [doi: [10.1177/0733464808324019](#)]
10. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. *Gerontologist* 2005 Dec;45(6):793-801. [Medline: [16326661](#)]
11. Austrom MG, Geros KN, Hemmerlein K, McGuire SM, Gao S, Brown SA, et al. Use of a multiparty web based videoconference support group for family caregivers: Innovative practice. *Dementia (London)* 2015 Sep;14(5):682-690 [FREE Full text] [doi: [10.1177/1471301214544338](#)] [Medline: [25062788](#)]
12. Gaugler JE, Hobday JV, Robbins JC, Barclay MP. CARES(®) Dementia Care for Families™: effects of online, psychoeducational training on knowledge of person-centered care and satisfaction. *J Gerontol Nurs* 2015 Oct;41(10):18-24. [doi: [10.3928/00989134-20150804-61](#)] [Medline: [26270065](#)]
13. Finkel S, Czaja SJ, Schulz R, Martinovich Z, Harris C, Pezzuto D. E-care: a telecommunications technology intervention for family caregivers of dementia patients. *Am J Geriatr Psychiatry* 2007 May;15(5):443-448. [doi: [10.1097/JGP.0b013e3180437d87](#)] [Medline: [17463195](#)]
14. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. *Aging Ment Health* 2013;17(5):544-554 [FREE Full text] [doi: [10.1080/13607863.2013.775641](#)] [Medline: [23461355](#)]
15. Chiu T, Marziali E, Colantonio A, Carswell A, Gruneir M, Tang M, et al. Internet-based caregiver support for Chinese Canadians taking care of a family member with alzheimer disease and related dementia. *Can J Aging* 2009 Dec;28(4):323-336. [doi: [10.1017/S0714980809990158](#)] [Medline: [19925698](#)]
16. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. *J Med Internet Res* 2015;17(5):e117 [FREE Full text] [doi: [10.2196/jmir.3717](#)] [Medline: [25967983](#)]
17. Lewis ML, Hobday JV, Hepburn KW. Internet-based program for dementia caregivers. *Am J Alzheimers Dis Other Dement* 2010 Dec;25(8):674-679. [doi: [10.1177/1533317510385812](#)] [Medline: [21131674](#)]
18. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 Internet-based intervention programs. *Am J Alzheimers Dis Other Dement* 2011 Feb;26(1):36-43. [doi: [10.1177/1533317510387586](#)] [Medline: [21282276](#)]
19. Pagán-Ortiz ME, Cortés DE, Rudloff N, Weitzman P, Levkoff S. Use of an online community to provide support to caregivers of people with dementia. *J Gerontol Soc Work* 2014;57(6-7):694-709 [FREE Full text] [doi: [10.1080/01634372.2014.901998](#)] [Medline: [24689359](#)]
20. Torkamani M, McDonald L, Saez Aguayo I, Kanios C, Katsanou M, Madeley L, ALADDIN Collaborative Group. A randomized controlled pilot study to evaluate a technology platform for the assisted living of people with dementia and their carers. *J Alzheimers Dis* 2014;41(2):515-523. [doi: [10.3233/JAD-132156](#)] [Medline: [24643137](#)]

21. Ducharme FC, Lévesque LL, Lachance LM, Kergoat M, Legault AJ, Beaudet LM, et al. "Learning to become a family caregiver" efficacy of an intervention program for caregivers following diagnosis of dementia in a relative. *Gerontologist* 2011 Aug;51(4):484-494. [doi: [10.1093/geront/gnr014](https://doi.org/10.1093/geront/gnr014)] [Medline: [21383112](https://pubmed.ncbi.nlm.nih.gov/21383112/)]
22. Duggleby W, Swindle J, Peacock S. Self-administered intervention for caregivers of persons with Alzheimer's disease. *Clin Nurs Res* 2014 Feb;23(1):20-35. [doi: [10.1177/1054773812474299](https://doi.org/10.1177/1054773812474299)] [Medline: [23393291](https://pubmed.ncbi.nlm.nih.gov/23393291/)]
23. Duggleby W. 2018. My Tools 4 Care. URL: <https://www.mytools4care.ca> [WebCite Cache ID 6zURoRYhF]
24. Duggleby W, Ploeg J, McAiney C, Fisher K, Swindle J, Chambers T, et al. Study protocol: pragmatic randomized control trial of an internet-based intervention (My tools 4 care) for family carers. *BMC Geriatr* 2017 Aug 14;17(1):181 [FREE Full text] [doi: [10.1186/s12877-017-0581-6](https://doi.org/10.1186/s12877-017-0581-6)] [Medline: [28806917](https://pubmed.ncbi.nlm.nih.gov/28806917/)]
25. Chan D. 2018. ATMIST. URL: <http://atmist.com/> [accessed 2018-05-18] [WebCite Cache ID 6zUQHcpqr]
26. Duggleby W, Tycholiz J, Holtslander L, Hudson P, Nikolaichuk C, Mirhosseini M, et al. A metasynthesis study of family caregivers' transition experiences caring for community-dwelling persons with advanced cancer at the end of life. *Palliat Med* 2017 Jul;31(7):602-616. [doi: [10.1177/0269216316673548](https://doi.org/10.1177/0269216316673548)] [Medline: [28618898](https://pubmed.ncbi.nlm.nih.gov/28618898/)]
27. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
28. Song M, Sandelowski M, Happ M. Current practices emerging trends in conducting mixed methods intervention studies in the health sciences. In: Tashakkori A, Teddlie C, editors. *Sage Handbook of Mixed Methods in Social and Behavioral Research*. 2nd ed. Thousand Oaks, CA: SAGE; 2010:725-748.
29. Alzheimer's Society of Canada. Alzheimer Society Canada. Toronto, Ontario; 2016. Progression URL: <http://alzheimer.ca/sites/default/files/files/national/progression-series/ps300-01e-2018-overview-final-md.pdf> [accessed 2018-05-14] [WebCite Cache ID 6zR1xfyGE]
30. Markowitz JS, Gutterman EM, Sadik K, Papadopoulos G. Health-related quality of life for caregivers of patients with Alzheimer disease. *Alzheimer Dis Assoc Disord* 2003;17(4):209-214. [Medline: [14657784](https://pubmed.ncbi.nlm.nih.gov/14657784/)]
31. Arango-Lasprilla JC, Lehan T, Drew A, Moreno A, Deng X, Lemos M. Health-related quality of life in caregivers of individuals with dementia from Colombia. *Am J Alzheimers Dis Other Dement* 2010 Nov;25(7):556-561. [doi: [10.1177/1533317510382287](https://doi.org/10.1177/1533317510382287)] [Medline: [20826708](https://pubmed.ncbi.nlm.nih.gov/20826708/)]
32. Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. User's manual for the SF-12v2 Health Survey (with a supplement documenting version 1). Lincoln, RI: QualityMetric Incorporated; 2002.
33. Schwarzer R. General Self-Efficacy Scale (GSE). 2012. URL: <http://userpage.fu-berlin.de/health/selfscal.htm> [accessed 2018-05-14] [WebCite Cache ID 6zR3aPMOg]
34. Herth K. Abbreviated instrument to measure hope: development and psychometric evaluation. *J Adv Nurs* 1992 Oct;17(10):1251-1259. [Medline: [1430629](https://pubmed.ncbi.nlm.nih.gov/1430629/)]
35. Browne G, Roberts J, Gafni A, Byrne C, Weir R, Majumdar B, et al. Economic evaluations of community-based care: lessons from twelve studies in Ontario. *J Eval Clin Pract* 1999 Nov;5(4):367-385. [Medline: [10579701](https://pubmed.ncbi.nlm.nih.gov/10579701/)]
36. Browne G, Roberts J, Byrne C, Gafni A, Weir R, Majumdar B. The costs and effects of addressing the needs of vulnerable populations: results of 10 years of research. *Can J Nurs Res* 2001 Jun;33(1):65-76. [Medline: [11928156](https://pubmed.ncbi.nlm.nih.gov/11928156/)]
37. Markle-Reid M, McAiney C, Forbes D, Thabane L, Gibson M, Browne G, et al. An interprofessional nurse-led mental health promotion intervention for older home care clients with depressive symptoms. *BMC Geriatr* 2014 May 10;14:62 [FREE Full text] [doi: [10.1186/1471-2318-14-62](https://doi.org/10.1186/1471-2318-14-62)] [Medline: [24886344](https://pubmed.ncbi.nlm.nih.gov/24886344/)]
38. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010 Mar;340:c869 [FREE Full text] [Medline: [20332511](https://pubmed.ncbi.nlm.nih.gov/20332511/)]
39. Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol* 2013 Jul 16;13:92 [FREE Full text] [doi: [10.1186/1471-2288-13-92](https://doi.org/10.1186/1471-2288-13-92)] [Medline: [23855337](https://pubmed.ncbi.nlm.nih.gov/23855337/)]
40. Liu Y, De A. Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *Int J Stat Med Res* 2015;4(3):287-295 [FREE Full text] [doi: [10.6000/1929-6029.2015.04.03.7](https://doi.org/10.6000/1929-6029.2015.04.03.7)] [Medline: [27429686](https://pubmed.ncbi.nlm.nih.gov/27429686/)]
41. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2nd edition. Hoboken, NJ: Wiley Publishers; 2011.
42. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ* 2011 Aug;20(8):897-916 [FREE Full text] [doi: [10.1002/hec.1653](https://doi.org/10.1002/hec.1653)] [Medline: [20799344](https://pubmed.ncbi.nlm.nih.gov/20799344/)]
43. Sandelowski M. What's in a name? Qualitative description revisited. *Res Nurs Health* 2010 Feb;33(1):77-84. [doi: [10.1002/nur.20362](https://doi.org/10.1002/nur.20362)] [Medline: [20014004](https://pubmed.ncbi.nlm.nih.gov/20014004/)]
44. Ploeg J, McAiney C, Duggleby W, Chambers T, Lam A, Peacock S, et al. A Web-based intervention to help caregivers of older adults with dementia and multiple chronic conditions: A qualitative study. *J Med Internet Res Aging* 2018;1(1):e2. [doi: [10.2196/aging.8475](https://doi.org/10.2196/aging.8475)]
45. Kaltenbaugh DJ, Klem ML, Hu L, Turi E, Haines AJ, Hagerty Lingler J. Using Web-based interventions to support caregivers of patients with cancer: a systematic review. *Oncol Nurs Forum* 2015 Mar;42(2):156-164. [doi: [10.1188/15.ONF.156-164](https://doi.org/10.1188/15.ONF.156-164)] [Medline: [25806882](https://pubmed.ncbi.nlm.nih.gov/25806882/)]

46. Laver K, Milte R, Dyer S, Crotty M. A systematic review and meta-analysis comparing carer focused and dyadic multicomponent interventions for carers of people with dementia. *J Aging Health* 2016 Jul 25:1308-1349 [FREE Full text] [doi: [10.1177/0898264316660414](https://doi.org/10.1177/0898264316660414)] [Medline: [27458254](https://pubmed.ncbi.nlm.nih.gov/27458254/)]
47. Williams A, Duggleby W, Ploeg J, Markle-Reid M, Moquin H, Yang J. Overcoming recruitment challenges for securing a survey sample of caregivers of community-dwelling older adults with multiple chronic conditions. *JHHR* 2017 Nov 14;1(1):16-24. [doi: [10.14302/issn.2576-9383.jhhr-17-1787](https://doi.org/10.14302/issn.2576-9383.jhhr-17-1787)]
48. Vandelanotte C, Kolt GS, Caperchione CM, Savage TN, Rosenkranz RR, Maeder AJ, et al. Effectiveness of a Web 2.0 intervention to increase physical activity in real-world settings: randomized ecological trial. *J Med Internet Res* 2017 Nov 13;19(11):e390 [FREE Full text] [doi: [10.2196/jmir.8484](https://doi.org/10.2196/jmir.8484)] [Medline: [29133282](https://pubmed.ncbi.nlm.nih.gov/29133282/)]

Abbreviations

ADRD: Alzheimer's disease and related dementias
ANCOVA: analysis of covariance
GEE: general estimating equations
GSES: General Self-Efficacy Scale
HHI: Herth Hope Index
HRQOL: health-related quality of life
HSSUI: Health and Social Services Utilization Inventory
MCC: multiple chronic conditions
MCS: mental component score SF-12
MT4C: My Tools 4 Care
PCS: physical component score SF-12
QOL: quality of life
RCT: randomized controlled trial
SF-12: Short Form-12 item health survey

Edited by G Eysenbach; submitted 24.03.18; peer-reviewed by E Hamad, Q Yuan; comments to author 09.05.18; accepted 19.05.18; published 29.06.18.

Please cite as:

Duggleby W, Ploeg J, McAiney C, Peacock S, Fisher K, Ghosh S, Markle-Reid M, Swindle J, Williams A, Triscott JAC, Forbes D, Jovel Ruiz K

Web-Based Intervention for Family Carers of Persons with Dementia and Multiple Chronic Conditions (My Tools 4 Care): Pragmatic Randomized Controlled Trial

J Med Internet Res 2018;20(6):e10484

URL: <http://www.jmir.org/2018/6/e10484/>

doi:[10.2196/10484](https://doi.org/10.2196/10484)

PMID:[29959111](https://pubmed.ncbi.nlm.nih.gov/29959111/)

©Wendy Duggleby, Jenny Ploeg, Carrie McAiney, Shelley Peacock, Kathryn Fisher, Sunita Ghosh, Maureen Markle-Reid, Jennifer Swindle, Allison Williams, Jean AC Triscott, Dorothy Forbes, Kathya Jovel Ruiz. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.06.2018. 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.

Original Paper

Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study

Akane Sano¹, PhD; Sara Taylor¹, MS; Andrew W McHill^{2,3}, PhD; Andrew JK Phillips^{2,3}, PhD; Laura K Barger^{2,3}, PhD; Elizabeth Klerman^{2,3}, PhD, MD; Rosalind Picard¹, ScD

¹Affective Computing Group, Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States

²Brigham and Women's Hospital, Boston, MA, United States

³Harvard Medical School, Boston, MA, United States

Corresponding Author:

Akane Sano, PhD

Affective Computing Group

Media Lab

Massachusetts Institute of Technology

75 Amherst Street

Cambridge, MA, 02139

United States

Phone: 1 6178999468

Email: akanes@media.mit.edu

Abstract

Background: Wearable and mobile devices that capture multimodal data have the potential to identify risk factors for high stress and poor mental health and to provide information to improve health and well-being.

Objective: We developed new tools that provide objective physiological and behavioral measures using wearable sensors and mobile phones, together with methods that improve their data integrity. The aim of this study was to examine, using machine learning, how accurately these measures could identify conditions of self-reported high stress and poor mental health and which of the underlying modalities and measures were most accurate in identifying those conditions.

Methods: We designed and conducted the 1-month SNAPSHOT study that investigated how daily behaviors and social networks influence self-reported stress, mood, and other health or well-being-related factors. We collected over 145,000 hours of data from 201 college students (age: 18-25 years, male:female=1.8:1) at one university, all recruited within self-identified social groups. Each student filled out standardized pre- and postquestionnaires on stress and mental health; during the month, each student completed twice-daily electronic diaries (e-diaries), wore two wrist-based sensors that recorded continuous physical activity and autonomic physiology, and installed an app on their mobile phone that recorded phone usage and geolocation patterns. We developed tools to make data collection more efficient, including data-check systems for sensor and mobile phone data and an e-diary administrative module for study investigators to locate possible errors in the e-diaries and communicate with participants to correct their entries promptly, which reduced the time taken to clean e-diary data by 69%. We constructed features and applied machine learning to the multimodal data to identify factors associated with self-reported poststudy stress and mental health, including behaviors that can be possibly modified by the individual to improve these measures.

Results: We identified the physiological sensor, phone, mobility, and modifiable behavior features that were best predictors for stress and mental health classification. In general, wearable sensor features showed better classification performance than mobile phone or modifiable behavior features. Wearable sensor features, including skin conductance and temperature, reached 78.3% (148/189) accuracy for classifying students into high or low stress groups and 87% (41/47) accuracy for classifying high or low mental health groups. Modifiable behavior features, including number of naps, studying duration, calls, mobility patterns, and phone-screen-on time, reached 73.5% (139/189) accuracy for stress classification and 79% (37/47) accuracy for mental health classification.

Conclusions: New semiautomated tools improved the efficiency of long-term ambulatory data collection from wearable and mobile devices. Applying machine learning to the resulting data revealed a set of both objective features and modifiable behavioral

features that could classify self-reported high or low stress and mental health groups in a college student population better than previous studies and showed new insights into digital phenotyping.

(*J Med Internet Res* 2018;20(6):e210) doi:[10.2196/jmir.9410](https://doi.org/10.2196/jmir.9410)

KEYWORDS

mobile health; mood; machine learning; wearable electronic devices; smartphone; mobile phone; mental health; psychological stress

Introduction

Background

Recent advances in wearable and mobile technologies have enabled individuals to monitor their daily lives and enabled scientific investigators to passively collect real-time data without disrupting people's habitual routines. Two examples of such devices are wrist-wearable devices that collect activity and other physiological data (eg, activity or sleep; heart rate; skin conductance, SC; blood pressure; and blood sugar level) and mobile phones (eg, smartphones) that monitor location, activity, social interaction over calls and texts (short message service, SMS), app use, screen on or off, and environmental data such as ambient light exposure and humidity.

Leveraging data from wearable and mobile devices to gain meaningful information about human health has been called digital phenotyping [1-3]. Digital phenotyping is defined as the moment-by-moment quantification of the individual-level human phenotype *in situ* using data from personal digital devices. Data from personal digital devices may be used to understand health and behaviors with a goal of preventing or minimizing disorders and diseases. For example, current health status, behavior history, and potential future health trajectories information might help (1) individuals become more aware of their risk profiles and enable them to make better informed decisions and take actions to change their behaviors to reduce potential negative physical and mental outcomes and (2) clinicians monitor changes in their client's or patient's status.

Mobile phones have been used to monitor stress and mental health [4-10]. The pioneering Student Life study that monitored 48 college students across a 10-week term using objective Android mobile phone sensors and usage investigated the relationship between well-being measures such as self-reported stress, depression, flourishing and loneliness, and academic performance [4]. Lower Perceived Stress Scale (PSS) score was correlated with higher conversation frequency during the day (9 AM-6 PM: the time frame participants might be in classes) and the evening (6 PM-0 AM), longer conversation duration during the day, and longer sleep duration. One study that evaluated self-reported depression using mobile phones for 2 weeks (N=28) [8] showed that mobility patterns (ie, regularity in 24-hour mobility patterns, as well as location variance) from Global Positioning System and phone usage features including usage duration and frequency were correlated with depressive symptom severity on a self-reported depression survey, the Patient Health Questionnaire-9 (PHQ-9) [8]. Another mobile phone-based study that lasted 12 weeks (N=73) identified mobile phone features that predicted clinically diagnosed depressed mood with 0.74 area under the curve; these features

including the total count of outgoing calls, the count of unique numbers texted, absolute distance traveled, dynamic variation of the voice, speaking rate, and voice quality [10].

The combination of wearable sensor and mobile phone data has also been used to study self-reported stress in daily life [11-14]. Muaremi et al, using both wearable sensors and mobile phones, developed a way to automate the recognition of self-reported daily stress levels using sleep parameters and 37 physiological responses (including heart rate, heart rate variability (HRV) and SC) from wearable sensors (N=10, 19 days), or mobile phone usage and sleep HRV from wearable sensors (N=35, 4 months). They showed 61% 3-class stress level classification accuracy with a combination of phone usage and sleep HRV features and 73% accuracy using sleep duration, upper body posture, and sleep HRV features [11,12]. Sano et al also investigated 5-day self-reported high or low stress recognition (N=18) and 1-month high or low stress recognition (N=66) using wearable sensor and mobile phone data; they showed 75% and 90% accuracy using leave-one participant-out or 10-fold cross-validation, respectively [13,14].

Objectives

These previous studies focused on only mobile phone usage or on phone usage plus wearable sensor data only during sleep and have not taken advantage of 24/7 multimodal phone + wearable data during wake and sleep to understand behaviors and physiology for long-term study of self-reported stress and mental health. We chose to approach this goal beginning with college students, most of whom report high stress, and some of whom are at risk of low or declining mental health [15,16]. According to the 2017 National College Health Assessment that examined data from 47,821 college students at 92 schools in the United States, more than half of the respondents said that their stress levels were higher than average, more than one-third had difficulty functioning because of depression, and two-thirds said they felt overwhelming anxiety in the last year [15]. Students' high stress and low mental health could negatively impact their academic performance [17]. Moreover, one-tenth of the students had a plan for suicide. Suicide rate is increasing, and suicide is the second leading cause of death for college students [18]. More students are seeking help, and 34% of counseling centers have a treatment waitlist [19]. Under these conditions, development of improved tools for screening, monitoring, and intervening for self-reported stress and poor mental health through wearable sensors and mobile phones in daily life settings will be beneficial. We aim to ultimately detect stress and mental health changes before clinical interventions are required and provide personalized early warnings together with data-driven suggestions of individualized behaviors that might promote better mental health outcomes.

Our SNAPSHOT study was designed to collect and examine rich multimodal information in participants' everyday life using wearable sensors and mobile phones for phenotyping sleep, stress, and mental health, all of which are major health issues in modern society. This paper has three main elements. First, we introduce a methodology and tools to capture long-term, large-scale ambulatory data on physiological and behavioral characteristics using sensors installed in wearable devices and mobile phones. The dataset from the SNAPSHOT study is one of the first large multimodal datasets that contains continuous physiology from a healthy college student population. The dataset currently includes approximately 145,000 hours of data from 201 participants at one university. Second, as real-world ambulatory data are messy, we describe tools we developed and deployed to improve the integrity and quality of the collected data and to reduce the time experimenters spend checking for and fixing errors. Third, we identify objective physiological markers and modifiable behaviors that successfully classify self-reported high or low stress and mental health and examine the separate contributions of wearable sensors and mobile phone data.

Methods

The 1-month SNAPSHOT study is a long-term and large-scale study developed to measure Sleep, Networks, Affect, Performance, Stress, and Health using Objective Techniques. Our aim was to investigate how daily behaviors and social networks influence sleep, self-reported stress, mood, performance, and other well-being-related factors. For each of five Fall and Spring semesters starting in Fall 2013, we collected approximately 1 month of data per person from college students who were socially connected and at a single New England university. Students were only allowed to participate in the study once. There was a total of 201 participants; Fall 2013: 20, Spring 2014: 48, Fall 2014: 46, Spring 2015: 47, Fall 2015: 40; ages 18 to 25 years; 129 male, 72 female; 70 freshman, 49 sophomore, 44 junior, 36 senior, and 2 unreported. The approximately 1 month of data collection was between the start of semester and midterms.

Recruitment

We intentionally recruited college students from a single academic institution who were socially connected because of our interest in how social networks affect sleep and health behaviors. Our definition of socially connected was making a call or SMS at least once a week with each other. Each semester, we recruited groups of at least 5 people who knew each other and interacted socially. We posted our study advertisement to undergraduate students' mailing lists. Potential participants filled out screening questionnaires to determine eligibility. Our exclusion criteria were as follows: (1) non-Android phone users, (2) inability to wear wrist sensors (eg, irritated skin on wrist), (3) pregnant women, (4) travel across more than one time zone 1 week before the study or have plans to travel more than one time zone away during the study, and (5) age <18 years or >60 years. In our study, we targeted only Android phone users because other mobile phones (eg, iPhone) did not allow us to monitor phone usage as needed for this study.

Eligible participants attended information and consent sessions. For each session, we invited approximately 15 participants and explained in detail the study and tasks that participants would perform during the study. After participants gave written informed consent, they completed prestudy questionnaires, started wearing devices, and installed an Android app (described below) on their phone. The study obtained a National Institutes of Health Certificate of Confidentiality so that potentially sensitive information such as drug or alcohol use provided by the participants could not be revealed for legal purposes; this was important protection for the students as the daily diary included requests for such information.

The participants received financial compensation at the end of the study; the amount depended on the number of days they completed diaries, wore the sensors, and completed other protocol tasks.

Study protocols were approved by the Massachusetts Institute of Technology and Partners HealthCare Institutional Review Boards. The study was registered on clinicaltrials.gov (NCT02846077).

Data Collection

All data were deidentified before analysis, although location information could potentially be used to reidentify people. Phone numbers, email addresses, and actual names from the social network surveys were hashed.

Start of the Study Questionnaires

At the start of the study, participants completed the Morningness-Eveningness Questionnaire [20], the Pittsburgh Sleep Quality Index [21], the Myers Brigg Personality test, the Big Five Inventory Personality Test [22], the PSS [23], the 12-Item Short Form Health Survey (SF-12) for physical and mental component summary (MCS) scores [24], and a set of social network surveys assessing with whom participants spent their time to help map their social networks. We also collected age, sex, academic major, and living situation (eg, dorm name and whether single or multiple occupancy room) information.

Ambulatory Monitoring

Wearable Sensors

Participants wore two sensors on their wrists: a Q-sensor (Affectiva, Boston, MA, United States) to measure SC, skin temperature (ST), three-axis *acceleration* (ACC) on their dominant wrist and a Motion Logger (AMI, Ardsley, NY, United States) on their nondominant wrist to measure acceleration and ambient light data. ACC can be used to estimate activity levels and sleep or wake patterns. SC reflects autonomic arousal during the day, providing a stress index during wakefulness; SC increases during sleep are highly likely to occur in either non-rapid eye movement (non-REM) stage 2 sleep or slow-wave sleep (SWS) [25]. The sensor data were logged into the flash memory of the sensors. Participants were instructed to remove sensors only in instances when the sensor could become wet or risked being broken.

Mobile Phone App

We wrote a custom Android phone app based on funf [26] that monitored location, receivers, senders, and timings of calls and SMS text messages, screen on or off timings, and phone app usage. No content of emails, calls, or SMS text messages was recorded. Phone usage was measured for two main reasons. First, phone usage and location data give clues to social interactions. The timing of calls, SMS, and screen on provide an estimate of how often participants interact with their phone during the day and the night, whereas the number of calls, SMS, and the number of people they interact with helps quantify their social interaction. Second, lighting from the interaction with mobile phones or emailing late at night could disturb the biological circadian clock and increase alertness, both of which

can influence sleep patterns [27,28]. We asked our participants not to use third-party messaging apps, if possible, during the study for the last two cohorts.

Twice-Daily Electronic Diaries

Participants completed electronic diaries (e-diaries): upon awakening and at bedtime each day. These diaries contained questions about sleep and wake times; naps; exercise; academic and extracurricular activity times; social interactions; caffeine, alcohol, and drug intake; overall health condition; sleep; mood; and self-reported stress (Figure 1). Participants received emails that included a URL to the morning and evening diaries. They could complete the diaries using computers, tablets, or mobile phones.

Figure 1. An example evening e-diary. For some questions, if yes is chosen, additional questions are presented.

Evening daily diary for 2016-10-06

Did you attend any academic activities today (including classes, e-classes, sections, seminars, labs, study groups)? ☒ Yes ☐ No

How many?

List each: when the activity started and for how long? [put extra events under 'Comments']

Activity 1 start 09:00 AM [Clear entry](#)

For how long?

How many hours did you study by yourself today, not including any of the academic activities in the previous question?

Did you engage in any exercise-based activities today (including sports, gym, cycling, etc.)? ☐ Yes ☒ No

Did you attend any other extracurricular activities today, besides academic activities and exercise-based activities? ☐ Yes ☒ No

For each of the following, indicate how you feel right now by clicking on each line and adjusting the sliders.

Sleepy	<input type="range"/>	Alert
Sad	<input type="range"/>	Happy
Sluggish	<input type="range"/>	Energetic
Sick	<input type="range"/>	Healthy
Stressed Out	<input type="range"/>	Calm Relaxed

In addition, did you have an emotionally charged interaction with someone today? ☒ No
☐ Yes, a memorable positive interaction
☐ Yes, a somewhat negative interaction
☐ Yes, and it was very negative

Please enter any comments for today

Save »

Poststudy Questionnaires and Other Measurements

At the end of the month of intensive data collection:

1. Academic performance as measured by grade point average was self-reported by each participant for the semester previous to the study and the current study semester.
2. Email usage during the experiment (ie, to, from, cc, and time stamps) was collected through the Massachusetts Institute of Technology (MIT) website Immersion [29].
3. On the basis of their phone call, SMS, and email usage objectively measured during the experiment, participants were asked to self-report whether they had positive or neutral or negative interactions with each frequent contact as a whole over the month. Participants also indicated to which category each frequent contact belonged to (ie, family, social, work, others).
4. The PSS, the SF-12, the set of social network surveys, and the State-Trait Anxiety Index [30] were completed.

Data Preprocessing

Ambulatory data measured with wearable sensors, mobile phones, and surveys tend to be noisy. Examples include (1) AM vs PM errors when participants complete survey items about their sleep and activity times; (2) participants forgetting to charge or wear sensors; (3) sensors breaking or the signals becoming noisy; and (4) mobile phone connectivity, hardware sensor functionality, and mobile software updates, which can break and interfere with data integrity. To address these issues, various techniques have been applied, such as data cleaning before data analysis [31]; data quality evaluation [32], detecting faulty data, noise reduction [33], and interpolating faulty or missing values [34,35]. To reduce the occurrence or impact of these issues, additional approaches can be used during ambulatory data collection. For example, during the study, an e-diary system can notify participants about potential inaccurate answers before they submit their answers, and a study investigator can check data quality of incoming data and provide feedback to the participants. For this study, we developed tools for improving the quality of the collected data and for supporting more efficient human checking and correcting of the phone, sensor, and e-diary data.

Preprocessing Twice-Daily Electronic Diaries

We collected a total of 6077 days of e-diary data. In the first year of the SNAPSHOT study, we set up an e-diary system that automatically sent surveys to our participants every morning and evening and then sent reminders if the participants did not complete the surveys within 12 hours. We implemented logic check functions on the system that prompted users to revise their answers if certain types of errors or missing answers were detected (eg, if two activity events overlapped, or if their reported wake time was earlier than their reported bedtime). During this first year, study investigators manually checked participants' answers every 1 to 2 days and emailed them to revise their answers when errors were found.

In year 2 of the study, we installed raster plots that visualize participants' activities over time (Figure 2). These raster plots were displayed to participants after they submitted their answers, allowing users to visually confirm their responses and return to

their survey to correct any errors. These raster plots reduced about half of the daily diary errors. The raster plots also reduced the total average time taken to preprocess 1 month of a participant's e-diary data by 53%: from 145 min (year 1) to 68 min (year 2).

Finally, in year 3, we created and installed an administrative module that includes three components to further improve data validity: a calendar view, interactive checking system, and a summary view. Every day, a study investigator logged into the diary system and saw the calendar view (Multimedia Appendix 1) that showed the number of participants in the study, the number of participants whose morning and evening diaries were checked, the number of unchecked diaries, the number of diaries that needed to be rechecked, and participants' comments. The interactive checking system automatically flagged missing answers in the e-diary and allowed the study investigator to check daily diaries just by flagging sections of the e-diary as error (Figure 3). Emails were automatically sent to participants if there were errors or missing answers. The summary view (Multimedia Appendix 2) showed the daily diary status for each participant in different colors (eg, green-acceptable, red-missing, and pink-error). These plots enabled the study investigator to understand which participants had filled out the daily surveys and which participants they needed to contact (eg, if there were repeated errors or missing entries in the diaries). This module further reduced the total average time taken to preprocess 1 month of a participant's e-diary data from 68 min (year 2) to 45 min (year 3). The combined changes in raster plots and the administrative modules reduced the total average time taken to clean 1 month of one participant's e-diary data by 69%: from 145 min (year 1) to 45 min (year 3). Overall, participants' daily diary completion rates ranged between 92% and 97% with no significant differences across semesters.

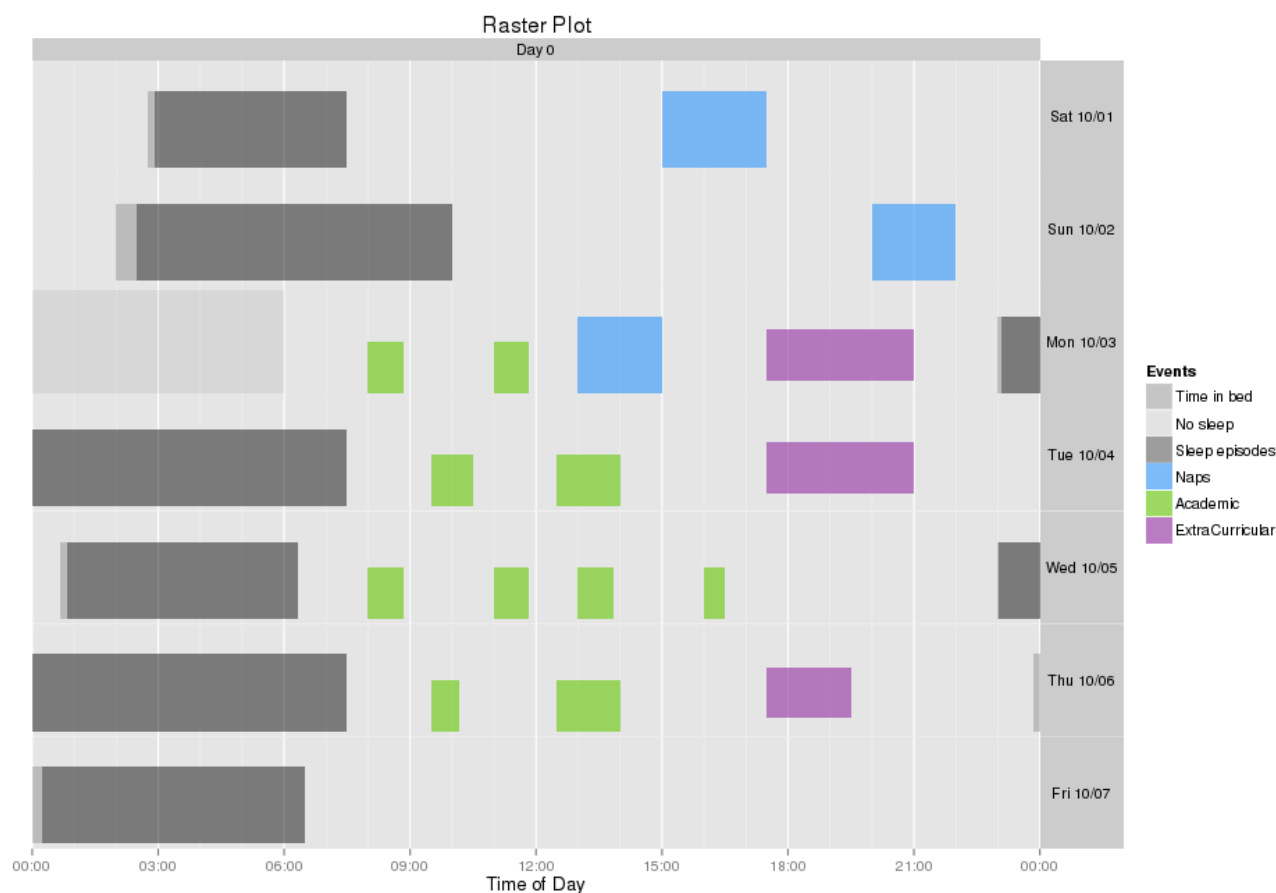
Preprocessing Sensor or Mobile Phone Data

Every week, a study investigator had a face-to-face meeting with each participant to download sensor data and to check if sensors were working correctly, if the participants were wearing them properly, and if sensor electrodes needed replacement. We developed scripts to download the data from sensors and check sensor readings automatically for quality using a previously developed and tested automated classifier [36]. This classifier separated clean epochs and noisy epochs of SC data for further analysis.

We collected 6309 days of Q-sensor data for a total of 125,413 hours. We computed how much data were within a typical range per published guidelines: for SC, 83% were within the range of 0.01 to 30 microS [37-39], and for ST, 99.7% were within the range of 20 to 42 degrees Celsius [40]. In addition, 92% of the collected SC data were classified as clean data using an artifact detection algorithm [36]. Thus, among the collected SC data, 80% of the data were used for further analysis.

Mobile phone data were sent automatically to a server by the custom funf-based app. On the server, another set of scripts that we wrote checked the data quality every day and sent notification to a participant if a problem was found in their data (eg, not receiving phone data for a day). Phone data were collected on 85% of the days.

Figure 2. Plot of daily activity timing (raster plot) with time of day (midnight to midnight) on the y-axis and each day plotted on a separate line. Participants saw this plot after filling out their surveys and before they submitted their answers. Different activities were marked with different colors.



Identifying Risk Factors, Objective Biomarkers, and Modifiable Behavioral Features Related to Stress and Mental Health

We defined high stress and low stress groups based on their poststudy PSS scores (Figure 4). PSS scores range from 0 to 40: higher scores indicate higher perceived stress. A PSS score of 14.2 is the average for the age group of 18 to 29 years, and a score over 16 is considered as high stress and of high health concern [23]. Our participants' average PSS score was 17.1. We used the value of $PSS \geq 16$ to construct the high stress group ($N=109$, top 57.7% [109/189]) and $PSS < 16$ for the low stress group ($N=80$, bottom 42.3% [80/189]). Because we originally had an unbalanced set of data for high stress and for low stress, we first reduced the size of the high stress group by the method of random sampling of its data to equalize the size of the high and low stress classes at $N=80$. Thus, the prior probabilities on both classes were made to be 0.5, so that a random classifier would be expected to attain accuracy of 50%.

We defined high mental health and low mental health groups based on their poststudy MCS from the SF-12 (Figure 4). For the MCS, a value ≥ 50 is considered good mental health [41,42], and 11.8% (23/195) of our population scored ≥ 50 . We therefore extracted the top and bottom 12% to form the two groups: high mental health group ($MCS \geq 50$, top 11.8% [23/195], $N=23$) and low mental health group ($MCS \leq 29.4$, bottom 12.3% [24/195], $N=24$). Thus, the data in the high and low mental health groups

were balanced so that the prior probability of either group would be 0.5, with a random classifier expected to have an accuracy of 50%.

Feature Extraction

To quantify the relative importance of the many measures, we compared the classification performance using the following separate categories of features: (1) Big Five personality + gender, (2) wearable sensors (eg, ST, SC, and ACC), (3) mobile phone (eg, call, SMS, screen on, and location), and (4) objective features (combining wearable sensors and mobile phone metrics). We also separately defined (5) modifiable behaviors as features that can potentially be controlled by participants, such as sleep and activity timing and phone usage; these are important features to measure for future behavioral interventions (Table 1). Note that some features such as phone features and ACC feature are found in more than one of the five categories.

SC was processed first using low-pass filtering (cutoff frequency 0.4 Hz, 32nd order finite impulse response filter). Because there are individual differences in SC amplitude, we extracted features from both unnormalized and normalized SC data based on the maximum and minimum amplitude of each day within each individual. To detect SC peaks, we obtained the first derivative of the low-pass-filtered non-normalized SC data and then determined where the slope exceeded a value of 0.02 μS per second [43]. We detected SC peaks based on those that exceeded this threshold and counted the number of peaks in each 30-second epoch.

Figure 3. Interactive diary check system. The left panel shows a participant's answers. The right panel shows if there are any detected errors or missing entries and enables adding comments. After the study investigator clicked the Save button, the system sent an email to a participant about any missing or erroneous entries if appropriate.

Did you attend any academic activities today (including classes, e-classes, sections, seminars, labs, study groups)? ☒ Yes ☐ No

How many?

List each: when the activity started and for how long? [put extra events under 'Comments']

Activity 1 start [Clear entry](#)

For how long?

How many class deadlines did you have today? (Class deadlines include homework deadlines, projects, quiz, mid-terms and finals)

How many hours did you study by yourself today, not including any of the academic activities in the previous question?

Did you engage in any exercise-based activities today (including sports, gym, cycling, etc.)? ☒ Yes ☐ No

How many times?

List each: when the activity started and for how long [put extra under 'Comments']

Activity 1 start [Clear entry](#)

For how long?

Did you attend any other extracurricular activities today, besides academic activities and exercise-based activities? ☒ Yes ☐ No

How many times?

List each: when the activity started and for how long [put extra under 'Comments']

Activity 1 start [Clear entry](#)

For how long?

Did you miss or were you late for any scheduled events (e.g., academic activities, exercise-based activities, other extracurricular activities) because you overslept today? ☒ Yes ☐ No

When did the event start? [Clear entry](#)

How many total servings of caffeine did you have today?

When is the latest hour you consumed caffeine? [Clear entry](#)

Please refer to this guide to help calculate servings

Cola 12 oz (1 can)	1/2 serving
Tea (1 cup)	1/2 serving
Home brew coffee 8 oz (1 cup)	1 serving
Energy drink 16 oz (1 large can)	2 servings
5 hour energy	2 servings
Large coffee 16 oz (large or grande)	2 servings
Caffeine pill - 100mg	1 serving
Caffeine pill - 200mg	2 servings

Besides caffeine, did you use any other medications, drugs, or alcohol today? ☐ Yes ☒ No

For each of the following, indicate how you feel right now by clicking on each line and adjusting the sliders.

Sleepy Alert

Sad Happy

Sluggish Energetic

Sick Healthy

Stressed Out Calm Relaxed

In addition, did you have an emotionally charged interaction with someone today? ☐ No ☒ Yes, a memorable positive interaction ☐ Yes, a somewhat negative interaction ☐ Yes, and it was very negative

Please enter any comments for today

[Save »](#)

Academic activity ☐ OK ☒ Missing ☐ Error

☐ AMPM ☐ Wrong Date ☐ Overlap

Email content

Exercise ☐ OK ☐ Missing ☒ Error

☒ AMPM ☐ Wrong Date ☐ Overlap

Email content

Extracurricular ☒ OK ☐ Missing ☐ Error

Overslept ☒ OK ☐ Missing ☐ Error

Caffeine or Drugs ☒ OK ☐ Missing ☐ Error

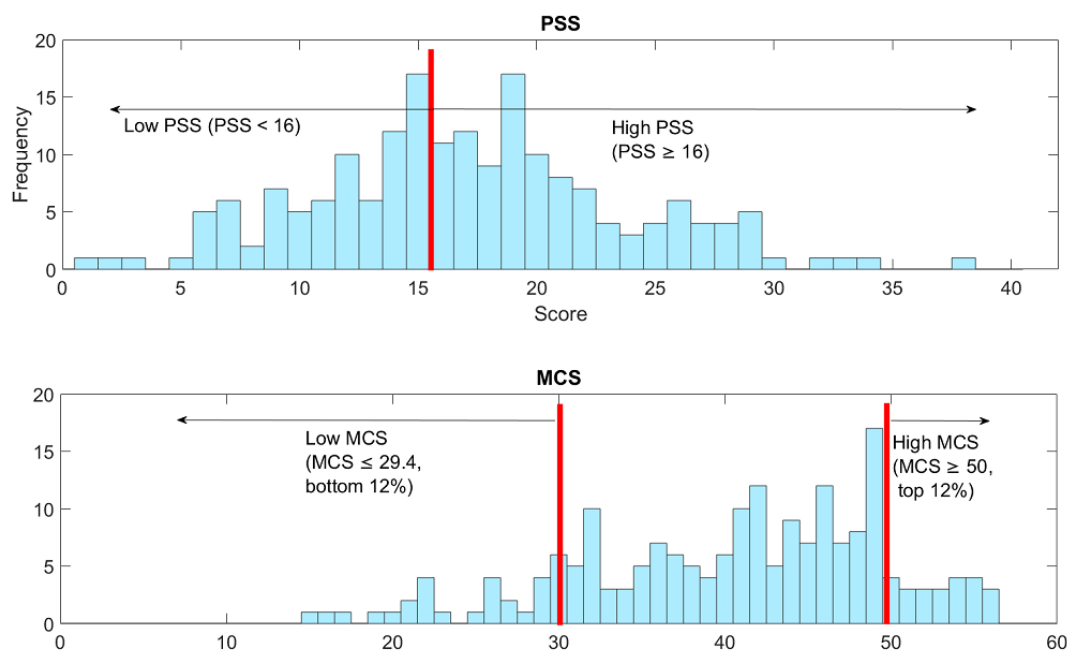
Scales ☒ OK ☐ Missing ☐ Error

Emotional Interaction ☒ OK ☐ Missing ☐ Error

Comments

Admin comments (not sent to participant)

[« Back](#) [Save »](#) [Skip](#)

Figure 4. (1) Distribution of poststudy Perceived Stress Scale (PSS) and (2) Distribution of poststudy mental component summary (MCS) scores.

We used four different times of interest for analyses: day (9 AM–6 PM), night (6 PM–0 AM), late night (0 AM–3 AM), and sleep time (estimated for each individual from actigraphy and daily sleep diaries) as physiological responses, such as SC and ACC during daytime and sleep time have different meanings [44] and late night phone and exercise activities could relate to self-reported stress and mental health [45].

Bedtime and sleep regularity were calculated from the daily sleep diaries, and sleep duration and sleep efficiency were estimated from actigraphy with help of the daily sleep diaries. Sleep regularity was computed because a relationship between irregular sleep and low mental health was found in a previous study using this index [46]. The Sleep Regularity Index (SRI; Figure 5) captures the probability of an individual being in the same state (asleep vs awake) at any two time points 24 hours apart with 1 minute resolution, averaged across the entire study [47], where $s(t)=1$ during wake and $s(t)=-1$ during sleep for each minute. Assume data are collected for $[0, T]$ with T =total number of hours of data and $\tau=24$ hours.

In practice, individuals will only display sleep patterns that range between an SRI of 0 (random) and 100 (periodic: an individual who sleeps and wakes at exactly the same times each day). Values less than 0 are theoretically possible (eg, alternating 24 h of sleep and 24 h of wake) but very unlikely to be observed.

Phone usage and location data can provide information on sociability. We computed the timing and the number of calls,

SMS, and screen on, which provide an estimate of how often participants interact with their phone during the day and the night. Previous studies showed the relationships between long phone usage duration and high stress [45] and long and frequent phone usage and severe depressive symptoms [8]. We also computed the number of people each participant interacted with over calls and SMS to help quantify their social interaction. For mobility features, we computed the distance and radius based on locations to which our participants travelled as these features were shown to be important in previous studies [8,48]. Additionally, because our population spent most of their time on campus or at their residence, we computed whether the day's mobility pattern varied from the typical routine based on a Gaussian mixture model trained for each participant's 1-month mobility patterns [49].

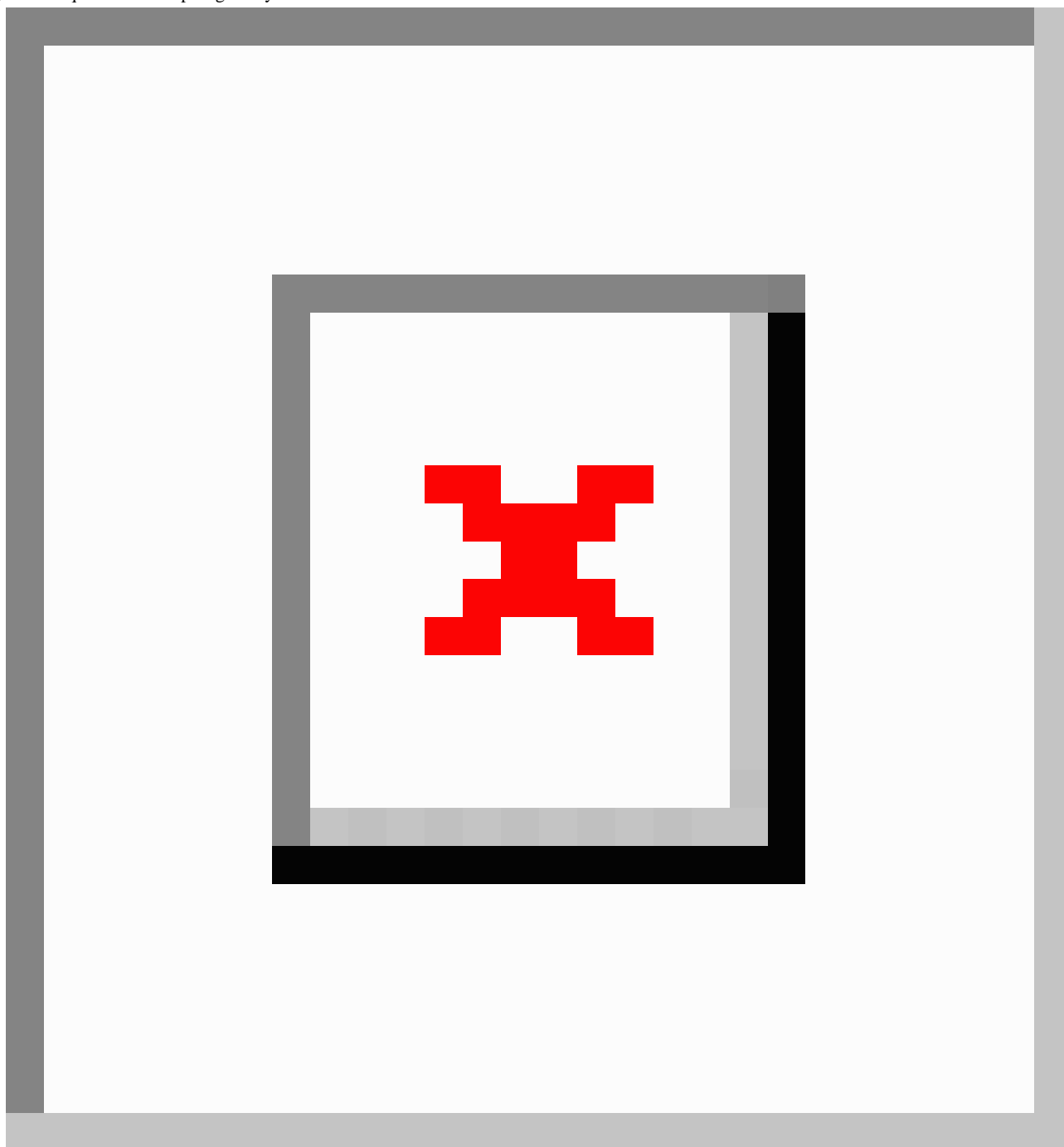
Classification

For classifying high or low stress groups and high or low mental health groups, we compared the methods of least absolute shrinkage and selection operator (LASSO), support vector machine (SVM) with linear kernel classifier, and SVM with radial basis function (RBF) kernel classifier; these algorithms were used in previous related work [8,10]. LASSO is a logistic regression that performs regularization and feature selection by minimizing the least squares objective function with an L1 penalty [50].

Table 1. List of features.

Modality	Features
All	Personality types, gender, diary, sensor, and phone features
Big Five personality types, gender (6 features)	Openness, conscientiousness, extraversion, agreeableness, neuroticism, gender
Sensors (17 features x 4 time frames x 3=204 features)	Mean, median, SD of 0 AM-3 AM, sleep, 9 AM-6 PM, 6 PM-0 AM for SC ^a , ACC ^b , and ST ^c Skin conductance: Area under the curve for 30 s epochs, max, mean, median, and SD of amplitude; mean, median and SD of peaks for 30 s epochs; mean, median, and SD of normalized amplitude Acceleration: total # of zero crossing for 30 s epochs Skin temperature: max, min, mean, median, and SD of temperature
Phone (25 features (call, SMS ^d , screen) x 3 time frames x 3 + 4 features (mobility) x 3 features=237 features)	Mean, median, SD of 0 AM-24 AM, 0 AM-3 AM, 6 PM-0 AM for call, SMS, and screen (not mobility) Call: Mean, median, and SD of duration and time stamp of calls per day; total duration per day, total number per day, and number of unique people per day SMS: Mean, median, and SD of duration and time stamp of SMS per day; total number per day and number of unique people per day Screen: Mean, median, and SD of screen-on duration and screen-on time stamp per day; total duration per day and total number of on or off per day Mobility: Total distance per day, 5-min distance, radius per day, and log likelihood of each day
Objective (441 features)	Phone and sensor features (see above)
Modifiable behaviors (296 features)	Sleep Regularity Index Mean, median, and SD of bedtime and sleep duration Diary features (see below) ACC total # of zero crossing for 30 s epochs Phone features (see above)
Diary (17 x 3=51 features)	Mean, median, SD of sleep or no sleep (pulled an all-nighter; binary valued), pre sleep electronic media interaction (emails, calls, SMS, Skype, chat, and online games; binary valued), pre sleep personal interaction(binary valued), # of naps, nap duration, # of academic activities per day, total academic duration, study duration, # of extracurricular activities, total extracurricular activities, # of exercise, exercise duration, # of caffeinated drink intake, memorable positive interaction(binary valued), somewhat negative interaction (binary valued), very negative interaction(binary valued), last caffeine intake time
Sleep (1 + 3 x 8=25 features)	Sleep Regularity Index Mean, median, and SD of bedtime, sleep duration, sleep efficiency, sleep or no sleep (pulled an all-nighter; binary valued), pre sleep electronic media interaction (emails, calls, SMS, Skype, chat, and online games; binary valued), pre sleep personal interaction (binary valued), # of naps and nap duration

^aSC: skin conductance.^bACC: acceleration.^cST: skin temperature.^dSMS: short message service.

Figure 5. Equation of Sleep Regularity Index.

For training and testing models, we used nested-cross validation. To evaluate model performance, we applied leave-one-cohort-out: training a model with all except one semester cohort's data and testing the model against the left-out cohort's data, repeating this process for the total number of cohorts (ie, 5 times). First we (1) split the data into two datasets: a training set made up of four cohorts and a test set made up of one cohort. We then left the test set out until step (5) or (8) below.

For training the SVM models, we applied sequential forward feature selection to the training data to reduce overfitting and find the best combinations. (2) We applied a *t* test to each feature of the training datasets and selected 100 features with the lowest *P* values for finding features to separate two groups effectively

then (3) applied sequential forward feature selection [51]: applied an SVM RBF classifier with 10-fold cross validation to find the best up to five combinations from these 100 features and optimized hyperparameters (*C* for SVM linear and *C* and *gamma* for SVM RBF). Then, (4) we trained the SVM linear or RBF models with the selected features of the training data, (5) tested the models against the test data, and (6) repeated this process (1-6) five times.

For LASSO, (7) the penalization parameter was determined with the training data by 10-fold cross validation and (8) the trained model was tested using the test data. This process (1, 7, and 8) was repeated five times.

We computed overall accuracy and F1 scores by concatenating the five-cohort predicted output to compare the performance of the models to reduce the bias from splitting [52]. We computed 95% confidence levels using adjusted Wald test [53]. The F1 score is a measure of performance computed using precision (also known as positive predictive value) and recall (also known as sensitivity) as described in Equation 1, where precision is the number of correct positive results divided by the number of all positive results, and recall is the number of correct positive results divided by the number of positive results that should have been returned.

$$(1) F1 = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$$

We also compared the performance of the models using features based on data from the entire 1-month study period with that using features based only on using the data from the week before the PSS and MCS surveys were completed.

We applied *t* tests or Mann-Whitney *U* tests (for non-Gaussian distributions) to examine if the means of the features were statistically different between the high or low PSS groups or the high or low MCS groups. We adjusted for the multiple comparisons using false discovery rate (FDR).

Results

Relationships Among Prestudy and Poststudy Perceived Stress Scores and Mental Component Summary

There were no differences in the poststudy PSS or MCS among the five cohorts; one-way analysis of variance ($P=.20$, $F=1.50$). Students' poststudy scores (both PSS and MCS) were highly correlated with prestudy scores ($r=.59$, $.60$, Pearson correlation). Poststudy PSS scores statistically increased (mean prestudy PSS: 15.0, poststudy PSS: 17.1, paired *t* test, $P<.001$) and MCS scores decreased compared with the prestudy scores (mean prestudy MCS: 44.4, poststudy MCS: 40.4, Wilcoxon signed rank test, $P<.001$). Thus, the students reported worsening stress and mental health over the 1 month of measurement.

The poststudy PSS was inversely correlated with the poststudy MCS ($r=-.71$, Pearson correlation; [Multimedia Appendix 3](#)): (1) 83% (19/23) of the students in the high MCS group belonged to the low PSS group and (2) 88% (21/24) of the students in the low MCS group belonged to the high PSS group. The low MCS group had higher PSS scores than the rest of the students in the high PSS group: low MCS group's average PSS score was 25.2, whereas the rest in the high PSS group's average PSS score was 20.7 ($P<.001$).

Stress and Mental Health Classification

Overall, we found SVM models with the RBF kernel worked better than LASSO and linear SVM models using RBF kernels for all of the metrics ([Figures 6 and 7](#); see [Multimedia Appendix 4](#) for accuracy and F1 scores and [Multimedia Appendices 5 and 6](#) for F1 scores for all results). SVM with the RBF kernel can model more complex decision boundaries. Sensor features showed higher performance than phone features both for PSS and MCS.

We also compared the performance of the SVM RBF models using features from only the last week of the 1-month period to using the features from the entire month. Overall, the performances with the 1 month of features were better (classification accuracy improved by 1-16%) than those using just the last week of features, except in the case of the SVM models using all features.

The accuracy for PSS classification was highest when using all features (82%), followed by when using features from only sensors (78%), only behaviors (74%), only the Big Five (71%), or only objective data (70%). The same rank ordering also held when comparing F1 scores. For MCS, sensor features and objective features showed the highest accuracy (87%), followed by Big Five (85%), behaviors (79%), and all (77%). The ranking of the F1 scores was similar except for all features had a slightly higher F1 than behaviors. The means and SD of the accuracy and F1 scores from leave-one-cohort-out cross validation are presented in [Multimedia Appendix 7](#).

We also tested different cutoffs: (1) instead of PSS cutoffs ≥ 16 for high and <16 for low stress, we used PSS ≥ 14 for high stress group and PSS <14 for low stress, as (as noted above) 14.2 is the reported average for people aged 18 to 29 years [23] and (2) instead of extreme MCS cutoffs (top and bottom 12%), we used MCS \geq median (42.05) for high mental health group and MCS $<$ median for low mental health group). This was done to test if the rankings of performances were sensitive to the exact cutoff values. Sensor and modifiable behavior features worked best with both cutoff values ([Multimedia Appendices 8 and 9](#)). Compared with the extreme MCS cutoffs, the median cutoff showed much lower classification performance (the accuracy decreased by 21 to 6 %).

We summarize the features most commonly selected by the algorithms as useful for high or low PSS detection ([Figure 8](#)) and high or low MCS detection ([Figure 9](#)) using the full 1 month of data. Percentages indicate the percent time these features are selected across 10-fold cross validation over five cohorts and five feature modalities (all, Big Five + gender, sensor, phone, objective and modifiable behavior features).

For PSS classification for self-reported stress, neuroticism and conscientiousness were the most often selected features (90% and 70% of the models, respectively). The high PSS group had higher neuroticism (q [which is the FDR-adjusted P value]=.0004). The high stress group had a larger extracurricular activity duration SD ($q=.04$).

In the MCS classification for self-reported mental health, the low MCS group showed higher neuroticism ($q<.001$) and lower conscientiousness ($q=.04$) than the high MCS group. The low MCS group had naps more frequently (40%; $q=.04$). In the MCS classification models using only the last week of data, the low MCS group showed a lower probability of interacting with electronic media (eg, emails, calls, SMS, Skype, chat, and online games) before sleep (30%; $q=.004$) and lower SD of the number of SC peaks during the time frame of 0 AM to 3 AM (20%; $q=.03$), as well as higher neuroticism ($q<.001$).

The percentages of time each feature was selected for each fold of leave-one-cohort cross validation are presented in [Multimedia Appendices 10](#) and [11](#).

We also tried building models only with sleep features (eg, features in the sleep category and some sleep related features in the survey category). We obtained 72% and 65% accuracy for classifying high or low PSS and high or low MCS. Mean nap duration was the most common feature used for the PSS models (80% of the models), followed by median bed time and

the frequency of pulling all-nighters (60%). The frequency of pulling all-nighters (100% of the models), mean number of naps, sleep duration, and sleep efficiency (60%) were commonly selected features by the MCS classification models. Average sleep duration was not significantly different statistically in the high vs low PSS groups or in the high vs low MCS groups (high PSS: 6 hours 42 min vs low PSS: 6 hours 51 min [$P=.09$], high MCS: 6 hours 40 min, low MCS: 6 hours 34 min [$P=.72$]). Instead, the low MCS group's more frequent napping was one of the most discriminating features.

Figure 6. High or low Perceived Stress Scale (PSS) classification results. Top: comparison of performance using 1 month of data with three machine learning algorithms. Bottom: comparison of performance using 1 month of data vs only the last week of data with support vector machine radial basis function (SVM RBF). Accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once. Error bars indicate the 95% CIs based on adjusted Wald test.

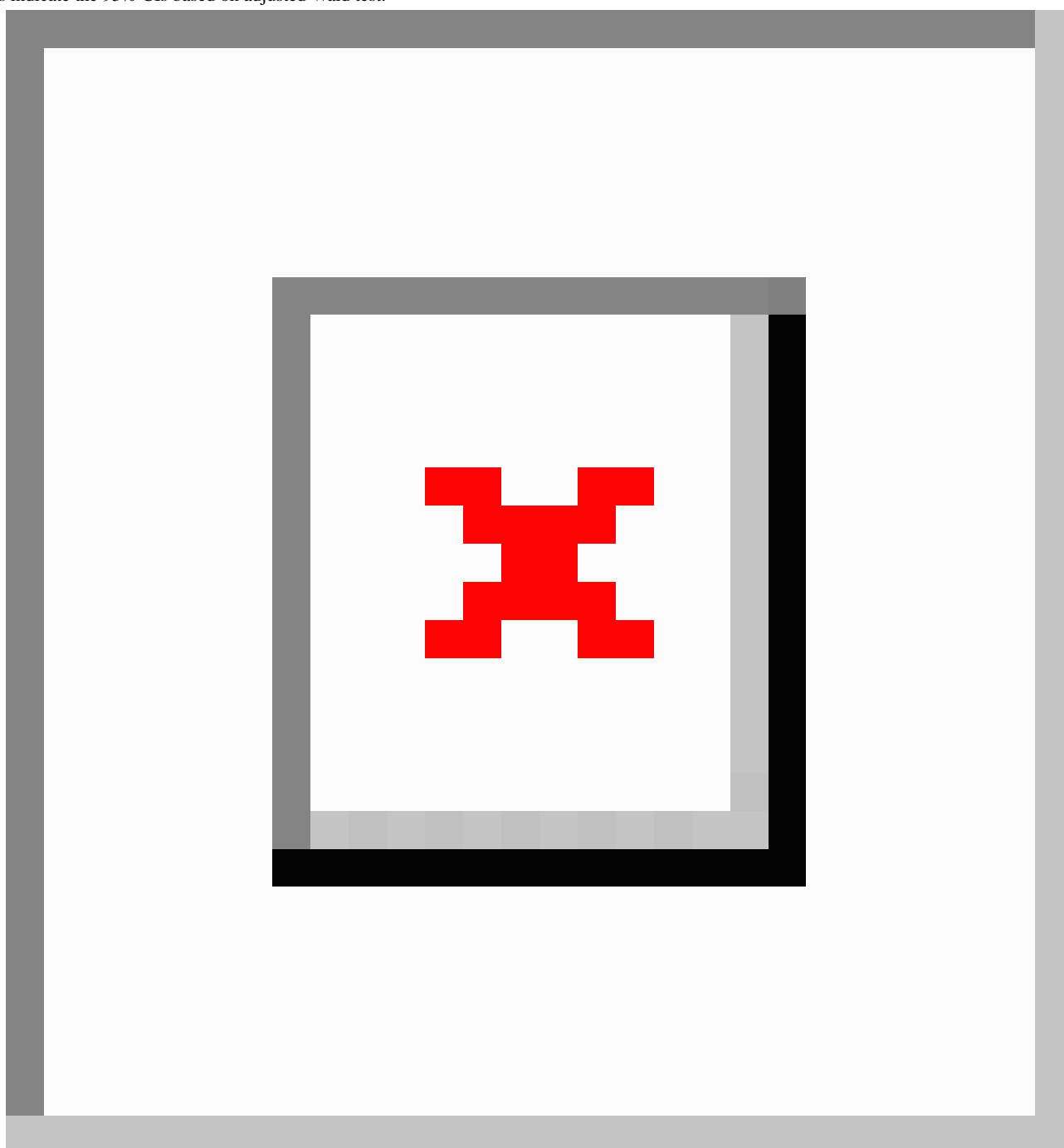


Figure 7. As in Figure 6 with high or low mental component summary score classification results, accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once. Error bars indicate the 95% CIs based on adjusted Wald test.

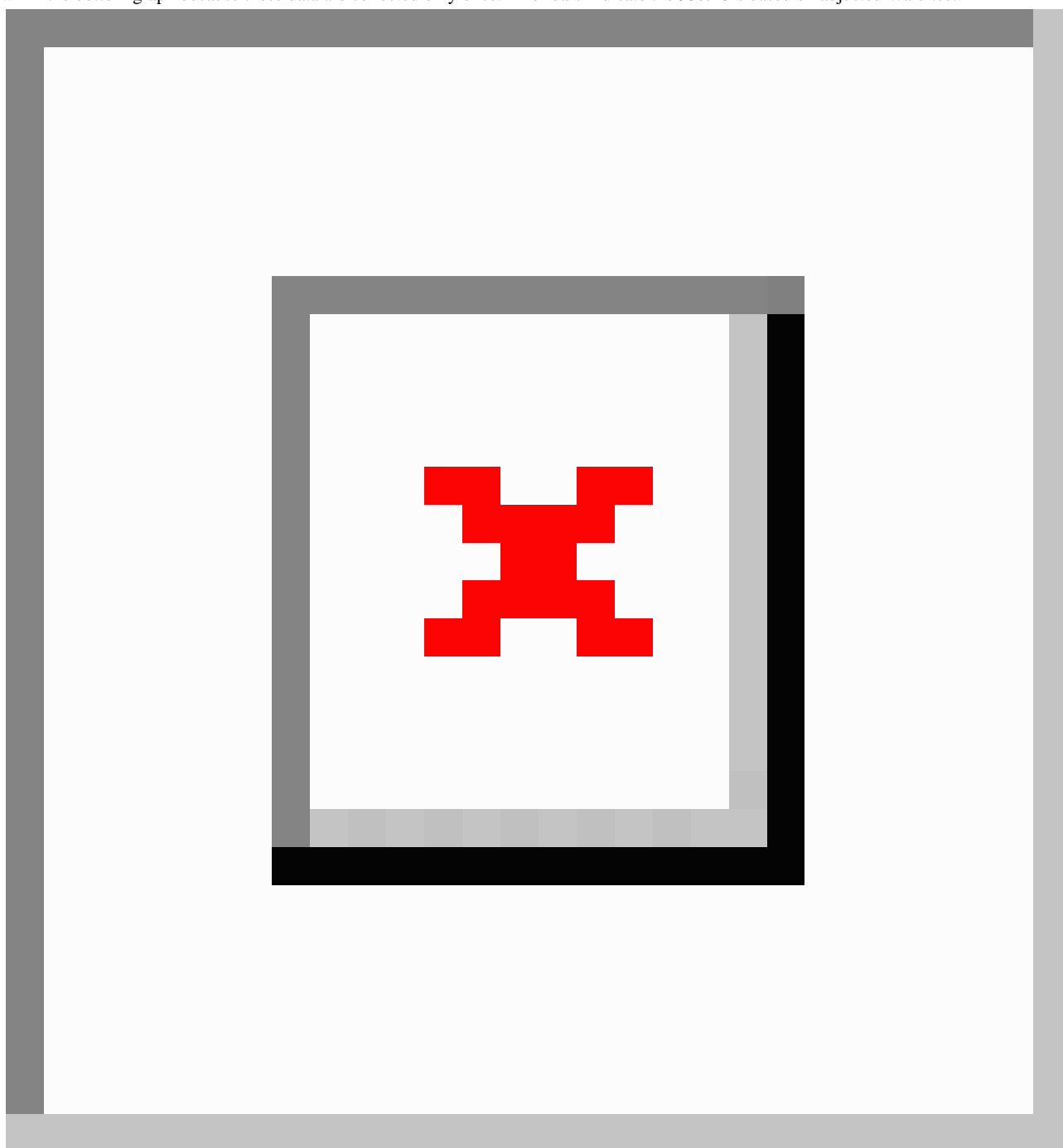


Figure 8. Percentage of time each feature was selected across 10-cross-validation for high or low Perceived Stress Scale (PSS) classification models with 1 month of data.

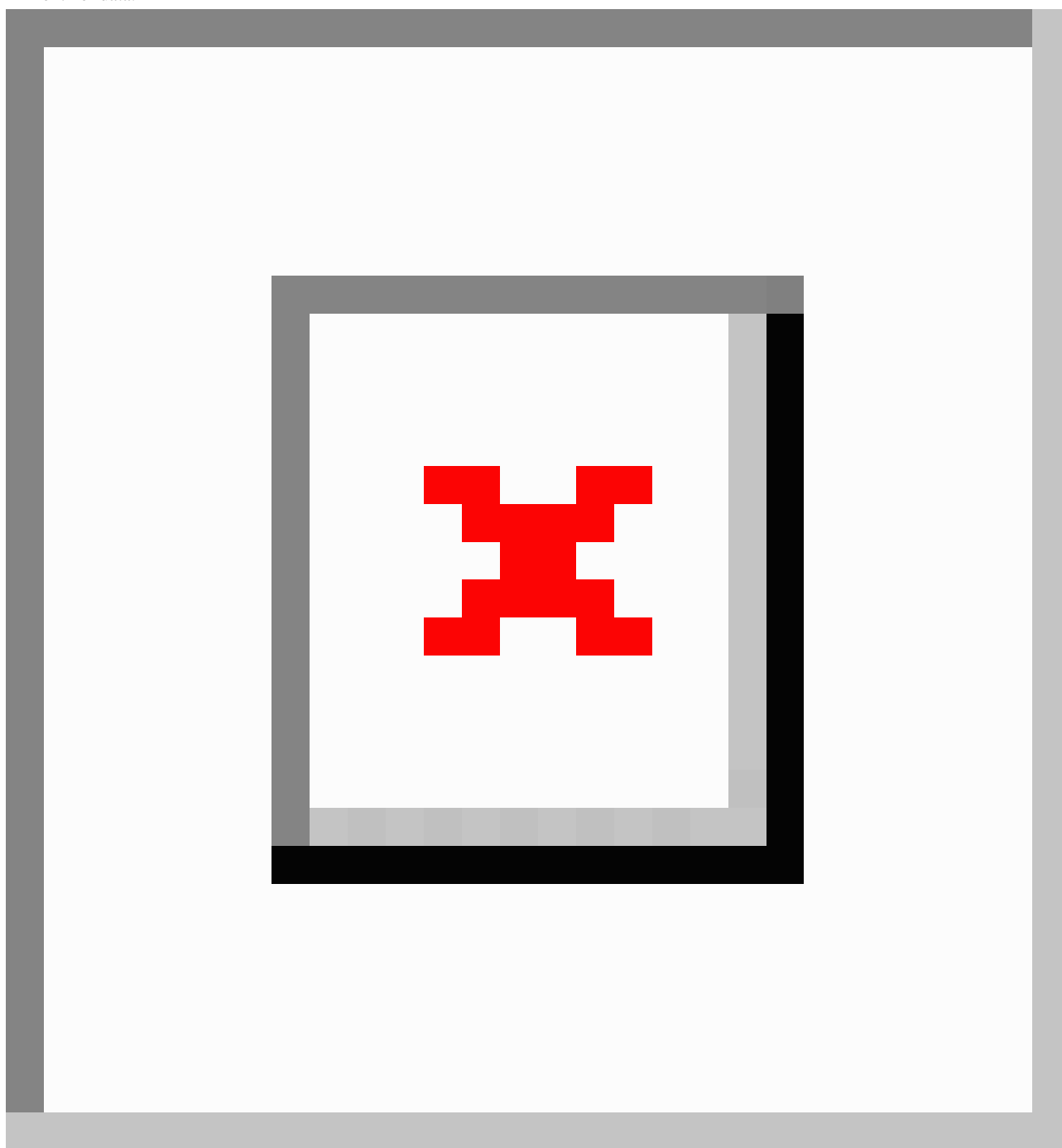
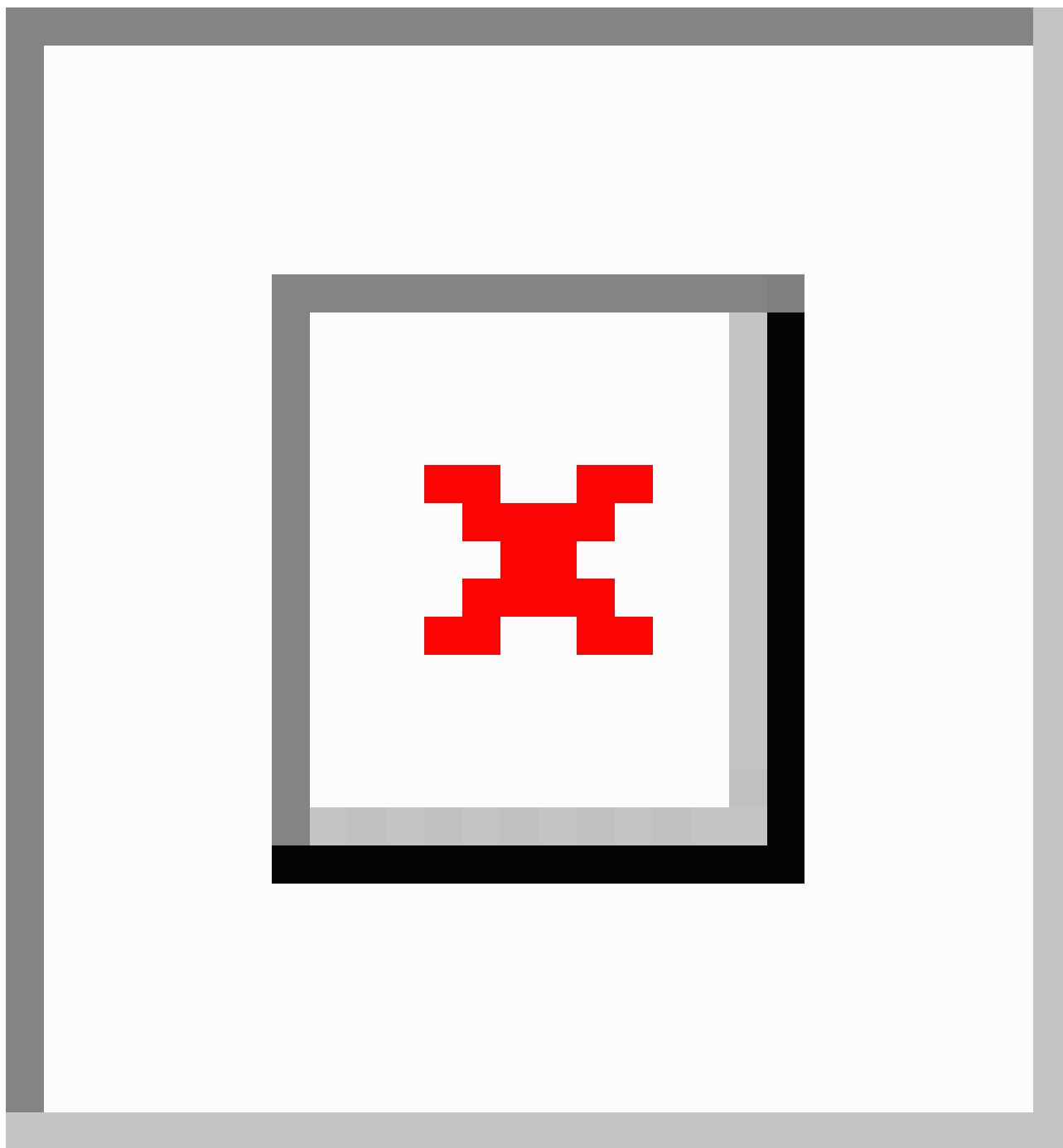


Figure 9. Percentage of times each feature was selected across 10-cross-validation for high or low mental component summary (MCS) classification models with 1 month of data.



Discussion

Principal Findings

In this paper, we developed novel tools to collect and process objective physiological and behavioral measures using online diaries, wearable sensors, and mobile phones. We aimed to investigate how accurately these measures could identify conditions of self-reported high stress and poor mental health and features most accurate in identifying these conditions.

Physiological sensor, phone, and mobility features were the best predictors for distinguishing self-reported high or low stress and mental health. Wearable sensor features, including SC and

ST, reached 79% accuracy for classifying high or low stress groups and 87% accuracy for classifying high or low mental health groups. Modifiable behaviors, including number of naps, studying duration, phone calls (number, time stamp and duration of calls), mobility patterns, and phone-screen-on time, reached 74% accuracy for high or low stress group classification and 78% accuracy for high or low mental health group classification.

Comparison With Prior Work and Interpretations of Our Results

Our analysis showed that relatively high accuracy and F1 scores can be achieved using the leave-one-semester-cohort-out testing of the machine learning classifier for high or low stress

measured by PSS and high or low mental health measured by MCS. Of all the features tested, the sensor features resulted in approximately 14% higher classification accuracies in both PSS and MCS than the phone features. In particular, SC responses during the time frame of 9 AM to 6 PM were one of the best predictors for PSS. SC has been considered as a biomarker for stress [44] because SC quantifies eccrine sweat activity that is controlled by only sympathetic nervous activity. These findings (1) are among the first to show the potential contribution of SC in stress detection using a wrist wearable sensor in a 24/7 daily life setting and (2) agree with previous findings that use a conventional finger SC sensor or a wearable SC sensor in settings where a person is seated, eg, driving a car. For example, Healey et al measured SC, heart rate, HRV, respiration, and electromyogram in Boston drivers and reported that SC was the most associated with stress [54]. Additionally, Hernandez et al discriminated stressful and nonstressful calls at a call center environment using SC features with 78% accuracy [55], and Setz et al automatically classified SC responses from cognitive load and stress with accuracy higher than 80% [56].

As we examined more closely which sensor features were most discriminative, we found that SC responses during the time frame of 0 AM to 3 AM and during sleep were predictors for separating high and low self-reported mental health. Some studies have shown that finger-based SC are reduced for patients with depression measured in a short-term lab study [57-59]. One possible explanation of how low SC responses during sleep could be related to MCS scores is that there is a decrease in SWS in depression [60] and other psychiatric disorders [61], and the largest SC responses during sleep are likely to occur during non-REM stage 2 and SWS [25]. Note that in our data, (1) 0 AM to 3 AM could include both awake and asleep conditions, and if it included sleep, we would expect it to include more SWS being at the start of the night for this cohort; (2) our low mental score groups are based on self-report; and (3) we do not know if any of our participants had clinically defined depression or other psychiatric disorders as that information was not gathered as part of this study.

We found that ST features were also predictors for PSS and MCS. A previous study has shown that acute stress does reduce distal finger ST but does not statistically significantly reduce wrist ST in a laboratory stress test setting [62]. Furthermore, another study showed that ST is one of the strongest discriminants to distinguish sleep and wake states [63]. Another study showed that patients with depression have less rhythmicity in ST [64], which would also be consistent with less regular sleep in depression. To our knowledge, this paper is the first to report that ambulatory wrist ST features are related to self-reported stress.

For phone features, our results showed phone usage time stamp and duration can be predictors for PSS and MCS. These results are consistent with several previous studies. People with a PHQ-9 score higher than 5 showed longer phone usage and higher phone usage frequency than those with a PHQ-9 score lower than 5 in a 2-week study with mobile phones [8]. A questionnaire-based study also showed a relationship between high mobile phone usage, stress, and symptoms of depression [45].

Mobility, specifically travel distance per day and SD of the distance traveled as measured by phone geolocation data, is a predictor both for self-reported stress and mental health. This result agrees with one study [8] reporting that normalized mobility entropy (distribution of frequency of visiting different places) and location variance were negatively correlated with depression symptoms and another study reporting that mobility patterns were highly related to stress level [13]. The relationship between reduced activity levels and mobility patterns and high stress and low mental health has been studied [48,65]. These behavioral markers could be an objective index for monitoring self-reported low mental health. It is possible that encouraging people to move more could be an effective intervention to reduce stress and improve mental health.

Consistent with previous studies [7,13,14], personality types were one of the most influential and statistically significant factors for self-reported stress and mental health in this college population. In the Big Five Inventory Personality Test categories, neuroticism was a predictor of stress [66]. The combination of low extraversion and low conscientiousness or low agreeableness contributed to the high stress group; these directions of the associations in our analysis were consistent with prior work [67]. High neuroticism and low extraversion have previously been associated with low MCS [68].

There is a known association between sleep deficiency and mental health status (eg, [61]). Our results, however, did not show that sleep duration was a strong discriminant feature for self-reported stress and mental health.

Limitations

There are multiple limitations of this study:

1. Selection of a feature as discriminating between two categories does not mean it is an important feature or causative of that behavior.
2. These results do not tell us the causality (eg, does a student sleep later and less regularly because of higher stress or have higher stress because of later or more irregular sleep?).
3. Our participants were limited to Android phone users because we wanted to log detailed phone usage, which is not allowed by other phone systems such as iPhone. As about half of the undergraduate students were Android users on the campus, a selection bias might exist. A previous study showed slight differences in personality types and economic status between iPhone users and Android users [69].
4. A total of 64% of our study population were male participants. It has been reported that females report higher perceived stress levels and more depressive symptoms [70-73], and there are gender differences in psychological and biological stress responses [74]. In our dataset, the ratios of female participants in the high or low PSS and MCS groups were 45% and 20% (high and low PSS) and 22% and 54% (high and low MCS). Modeling stress and mental health differently in males and females might help understand the mechanism. Gender was included as a potential feature in our models and was not selected frequently.

5. Our data come from college students at one New England university over 4 years. The work needs to be applied to other populations to determine generalizability.
6. Our data come from socially connected student groups. We might observe some statistically coherent behaviors in our dataset because of these connections.

Future Work

These new tools and methods can allow multimodal data in daily life to be captured more continuously, with greater accuracy and integrity of the data, and for long-term and at great scale. We are planning to collect a larger amount of data for an even longer time to study long-term behaviors and physiological responses and build predictive models. To do this, we need to build a new system for consenting people in remote locations, fully automate checking their measurement status and data accuracy automatically, and let the participants know about errors so they can fix them to keep study compliance and data accuracy high.

We will continue our data analysis for understanding behaviors, physiological responses, and traits that impact health and well-being. One of our hypotheses is that health-related behaviors will be contagious within social networks and that social network data we obtained from call, SMS, and email data could capture the social contagion quantitatively instead of requiring self-report to capture it. We are also interested in studying how phone usage influences sleep and health and how we can predict stress and mental health using previous behaviors and physiology.

These machine learning models are not limited to modalities and features we measured and computed in this study but can also be used for other modalities such as heart rate and heart rate variability that are controlled by autonomous activities, and other features such as app usage, ambient light, and audio or sentiment-based patterns extracted from text or speech could be added to improve the models. The features and models presented in this paper can be tested in similar multimodal ambulatory datasets collected in other future studies. Tracking stress and mental health conditions would help students better understand their stress and mental health conditions over multiple semesters, as well as help clinicians see how treatment affects students' conditions if they receive treatment.

Conclusions

In this paper, we introduced a methodology and tools we developed to measure ambulatory multimodal data and improve the integrity of collected data to study self-reported stress and mental health in the daily lives of college students. We showed that objective and modifiable behavioral features collected over 1 month can classify these college students as high or low stress based on the PSS and as having high or low mental health based on MCS from SF-12 collected at the end of that month with over 70% accuracy, whereas sensor features alone could classify high or low mental health and achieve over 88% on an F1 score. For classifying high or low stress groups, we found that combining phone and sensor features typically gave the best results over using either modality alone, whereas for classifying high or low mental health groups, the use of wearable sensor features performed comparable to wearable + phone features.

Acknowledgments

The authors are grateful to Mr Conor O'Brien, Mr Justin Buie, Mr Salim Qadri, Ms Natalie Virayan, Mr Michael Shreeve, Mr Omer Zaidi, Ms Natasha Jaques, Mr Weixuan Chen, Ms Asma Ghandeharioun, Mr Daniel Lopez Martinez, Ms Ehimwenma Nosakhare, Ms Amy Zhao Yu, Mr Daniel Smilkov, Ms Jade Philipoom, Ms Yuna Hahn, Ms Sienna Ramos, Ms Ji Hyun Gia Min, Ms Tania Yu, Ms Shirley Chen, Ms Laura Breiman, and Dr Catherine Ricciardi for their tremendous support helping run the SNAPSHOT study and collecting the data, and to Dr Cesar Hidalgo for helping design social network surveys. They also appreciate all the participants and support from the MIT Media Lab Consortium, especially a generous donation by Samsung Electronics, NEC, and funding from NIH grants R01GM105018, R00HL119618 (AJKP), K24HL105664 (EBK), and KL2TR002370, F32DK107146, T32HL007901 (AWM) and Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health Award UL1 TR001102), and financial contributions from Harvard University and its affiliated academic health care centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the National Institutes of Health.

Conflicts of Interest

RP is a cofounder of and shareholder in Affectiva, who commercialized the original sensors used in this study. RP is also a cofounder and shareholder in Empatica, a company that makes wearable sensors that can collect ambulatory data similar to the data collected in this study. EK has consulted for legal firms and for Pfizer Pharmaceuticals.

Multimedia Appendix 1

Calendar view. For each day, the investigator can see how many participants are in the study, how many surveys have been verified, how many need to be re-examined, and participant comments.

[[PNG File, 102KB](#) - [jmir_v20i6e210_app1.png](#)]

Multimedia Appendix 2

Summary view. The investigator can see the daily diary status for each participant in different colors: green: acceptable; red: missing; pink: error; red/white slash: missing but participants still have time to complete the diary.

[[PNG File, 100KB](#) - [jmir_v20i6e210_app2.png](#)]

Multimedia Appendix 3

The relationship between poststudy Perceived Stress Scale (PSS) and poststudy Mental Component Score (MCS) for each participant (blue circle). Chosen cutoffs for high and low PSS and high and low MCS are indicated.

[[PNG File, 57KB](#) - [jmir_v20i6e210_app3.png](#)]

Multimedia Appendix 4

Performance of PSS and MCS classification models with 1 month or last week of data.

[[PDF File \(Adobe PDF File\), 28KB](#) - [jmir_v20i6e210_app4.pdf](#)]

Multimedia Appendix 5

High or low Perceived Stress Scale (PSS) classification results. Top: Comparison of F1 scores for PSS classification with three machine learning algorithms (LASSO, SVM linear, and SVM RBF) on 1 month of different types of data (All, Big Five + Gender, Sensor, Phone, Objective, Behaviors). Bottom: Comparison of F1 scores with SVM RBF machine learning algorithm on 1 month of data versus on only the last week of the same data types. Accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once.

[[PNG File, 21KB](#) - [jmir_v20i6e210_app5.png](#)]

Multimedia Appendix 6

High or low Mental Component Score (MCS) classification results. Top: Comparison of F1 scores for MCS classification with three machine learning algorithms (LASSO, SVM linear, and SVM RBF) on 1 month of different types of data (All, Big Five + Gender, Sensor, Phone, Objective, Behaviors). Bottom: Comparison of F1 scores with SVM RBF machine learning algorithm on 1 month of data versus on only the last week of the same data types. Accuracy scores for Big Five + Gender data are not shown in the bottom graph because these data are collected only once.

[[PNG File, 22KB](#) - [jmir_v20i6e210_app6.png](#)]

Multimedia Appendix 7

Mean and SD of accuracy and F1 scores from leave-one-cohort-out PSS and MCS classification models with 1 month of data and SVM RBF.

[[PDF File \(Adobe PDF File\), 26KB](#) - [jmir_v20i6e210_app7.pdf](#)]

Multimedia Appendix 8

High or low Perceived Stress Scale (PSS) and Mental Component Score (MCS) classification results. Comparison of F1 for PSS and MCS classification scores with SVM RBF and 1 month of different types of data (All, Big Five + Gender, Sensor, Phone, Objective, Behaviors). Cutoff: 14 (the average in the 18-29 years age group) for PSS and 42.05 (median) for MCS.

[[PNG File, 14KB](#) - [jmir_v20i6e210_app8.png](#)]

Multimedia Appendix 9

Performance of PSS and MCS classification models with 1 month of data and SVM RBF. PSS cutoff: 14 (the average in the 18-29 years age group) and MCS cutoff: 42.05 (median).

[[PDF File \(Adobe PDF File\), 16KB](#) - [jmir_v20i6e210_app9.pdf](#)]

Multimedia Appendix 10

Percentages of the number of times each feature was selected for each fold of leave-one-cohort-out cross validation for 1-month PSS models.

[PDF File (Adobe PDF File), 20KB - [jmir_v20i6e210_app10.pdf](#)]

Multimedia Appendix 11

Percentages of the number of times each feature was selected for each fold of leave-one-cohort-out cross validation for 1-month MCS models.

[PDF File (Adobe PDF File), 19KB - [jmir_v20i6e210_app11.pdf](#)]

References

1. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Ment Health* 2016 May 5;3(2):e16 [FREE Full text] [doi: [10.2196/mental.5165](#)] [Medline: [27150677](#)]
2. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol* 2015 May;33(5):462-463. [doi: [10.1038/nbt.3223](#)] [Medline: [25965751](#)]
3. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol* 2017 Jan;15(1):e2001402 [FREE Full text] [doi: [10.1371/journal.pbio.2001402](#)] [Medline: [28081144](#)]
4. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Presented at: UbiComp '14 Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, Washington. [doi: [10.1145/2632048.2632054](#)]
5. Bogomolov A, Lepri B, Pianesi F. Happiness recognition from mobile phone data. 2013 Presented at: International Conference on Social Computing; September 8-14, 2013; Alexandria, VA, USA. [doi: [10.1109/SocialCom.2013.118](#)]
6. Bauer G, Lukowicz P. Can smartphones detect stress-related changes in the behaviour of individuals? 2012 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops; March 19-23, 2012; Lugano, Switzerland. [doi: [10.1109/PerComW.2012.6197525](#)]
7. Bogomolov A, Lepri B, Ferron M, Pianesi F, Pentland A. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. 2014 Presented at: Proceedings of the 22nd ACM international conference on Multimedia; November 3-7, 2014; Orlando, Florida, USA. [doi: [10.1145/2647868.2654933](#)]
8. 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 Jul 15;17(7):e175 [FREE Full text] [doi: [10.2196/jmir.4273](#)] [Medline: [26180009](#)]
9. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *J Med Internet Res* 2017 Mar 3;19(3):e62 [FREE Full text] [doi: [10.2196/jmir.6820](#)] [Medline: [28258049](#)]
10. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. *J Med Internet Res* 2017 Mar 16;19(3):e75 [FREE Full text] [doi: [10.2196/jmir.6678](#)] [Medline: [28302595](#)]
11. Muaremi A, Arnrich B, Tröster G. Towards measuring stress with smartphones and wearable devices during workday and sleep. *Bionanoscience* 2013;3:172-183 [FREE Full text] [doi: [10.1007/s12668-013-0089-2](#)] [Medline: [25530929](#)]
12. Muaremi A, Bexheti A, Gravenhorst F, Arnrich B, Tröster G. Monitoring the Impact of Stress on the Sleep Patterns of Pilgrims using Wearable Sensors. 2014 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); June 1-4, 2014; Valencia, Spain. [doi: [10.1109/BHI.2014.6864335](#)]
13. Sano A, Picard RW. Stress Recognition Using Wearable Sensors and Mobile Phones. 2013 Presented at: Humaine Association Conference on Affective Computing and Intelligent Interaction; September 2-5, 2013; Geneva, Switzerland. [doi: [10.1109/ACII.2013.117](#)]
14. Sano A, Phillips AJ, Yu AZ, McHill AW, Taylor S, Jaques N, et al. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. 2015 Presented at: IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN); June 9-12, 2015; Cambridge, MA, USA. [doi: [10.1109/BSN.2015.7299420](#)]
15. CACUSS. American College Health Association-National College Health Assessment II: Canadian Reference Group Executive Summary Spring 2017 URL: http://www.cacuss.ca/health_data.htm [accessed 2018-05-16] [WebCite Cache ID [6zT9bD7yo](#)]
16. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization world mental health surveys. *Psychol Med* 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: [10.1017/S0033291716001665](#)] [Medline: [27484622](#)]
17. Eisenberg D, Golberstein E, Hunt J. Mental health and academic success in college. *B E J Econom Anal Policy* 2009;9(1):-.

18. The American Association of Suicidology. Suicidology. 2016. College Students & Suicide Fact Sheet: 2016 Fact Sheet URL: <http://www.suicidology.org/Portals/14/Re-Formatted%20College%20Students%20Fact%20Sheet.pdf?ver=2016-11-16-110354-547> [accessed 2018-05-16] [WebCite Cache ID 6zTBIkcVH]
19. The Association for University and College Counseling Center Directors. The association for university and college counseling center directors annual survey URL: http://files.cmcglobal.com/AUCCCD_2013_Monograph_Public.pdf [accessed 2018-05-16] [WebCite Cache ID 6zT9rC5tt]
20. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int J Chronobiol* 1976;4(2):97-110. [Medline: [1027738](#)]
21. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989 May;28(2):193-213. [Medline: [2748771](#)]
22. John O, Srivastava S. Moityca.com.br. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives URL: http://moityca.com.br/pdfs/bigfive_John.pdf [accessed 2018-05-18] [WebCite Cache ID 6zVEjcmQv]
23. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983 Dec;24(4):385-396. [Medline: [6668417](#)]
24. 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](#)]
25. Sano A, Picard RW, Stickgold R. Quantitative analysis of wrist electrodermal activity during sleep. *Int J Psychophysiol* 2014 Dec;94(3):382-389 [FREE Full text] [doi: [10.1016/j.ijpsycho.2014.09.011](#)] [Medline: [25286449](#)]
26. Aharony N, Pan W, Ip C, Khayal I, Pentland A. Social fMRI: investigating and shaping social mechanisms in the real world. *Pervasive Mob Comput* 2011 Dec;7(6):643-659. [doi: [10.1016/j.pmcj.2011.09.004](#)]
27. Cajochen C, Frey S, Anders D, Späti J, Bues M, Pross A, et al. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. *J Appl Physiol* (1985) 2011 May;110(5):1432-1438 [FREE Full text] [doi: [10.1152/japplphysiol.00165.2011](#)] [Medline: [21415172](#)]
28. Chang AM, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proc Natl Acad Sci U S A* 2015 Jan 27;112(4):1232-1237 [FREE Full text] [doi: [10.1073/pnas.1418490112](#)] [Medline: [25535358](#)]
29. Immersion. A people-centric view of your email life URL: <https://immersion.media.mit.edu/> [accessed 2018-05-18] [WebCite Cache ID 6zVK7Rb2d]
30. Spielberger C. Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, CA: Consulting Psychologists Press; 1983.
31. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med* 2005 Oct;2(10):e267 [FREE Full text] [doi: [10.1371/journal.pmed.0020267](#)] [Medline: [16138788](#)]
32. Clifford GD, Lopez D, Li Q, Rezek I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. 2011 Presented at: Computing in Cardiology; September 18-21, 2011; Hangzhou, China.
33. Wood LB, Asada HH. Noise cancellation model validation for reduced motion artifact wearable PPG sensors using MEMS accelerometers. 2006 Presented at: International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 03, 2006; New York, NY, USA. [doi: [10.1109/IEMBS.2006.260359](#)]
34. Engels J. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol* 2003 Oct;56(10):968-976. [doi: [10.1016/S0895-4356\(03\)00170-7](#)]
35. Schmitt P, Mandel J, Guedj M. A comparison of six methods for missing data imputation. *J Biom Biostat* 2015;6:224. [doi: [10.4172/2155-6180.1000224](#)]
36. Taylor S, Jaques N, Chen W, Fedor S, Sano A, Picard R. Automatic identification of artifacts in electrodermal activity data. 2015 Presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 25-29, 2015; Milan, Italy URL: <http://europepmc.org/abstract/MED/26736662> [doi: [10.1109/EMBC.2015.7318762](#)]
37. Doberenz S, Roth WT, Wollburg E, Maslowski NI, Kim S. Methodological considerations in ambulatory skin conductance monitoring. *Int J Psychophysiol* 2011;80(2):87-95. [doi: [10.1016/j.ijpsycho.2011.02.002](#)]
38. Dawson ME, Schell AM, Filion DL. The electrodermal system. In: *Handbook of Psychophysiology*. New York, NY, US: Cambridge University Press; 2007:159-181.
39. Wenzel A. The SAGE Encyclopedia of Abnormal and Clinical Psychology. Thousand Oaks, California: SAGE Publications, Inc; 2017.
40. Texas Instruments. TIDA-00824 Human Skin Temperature Sensing for Wearable Applications Reference Design URL: <http://www.ti.com/lit/ug/tiduay7/tiduay7.pdf> [accessed 2018-05-16] [WebCite Cache ID 6zTAnV7XJ]
41. Bolge SC, Flores NM, Phan JH. The burden of poor mental well-being among patients with type 2 diabetes mellitus: examining health care resource use and work productivity loss. *J Occup Environ Med* 2016 Nov;58(11):1121-1126 [FREE Full text] [Medline: [27820762](#)]
42. Wolff JL, Roter DL. Older adults' mental health function and patient-centered care: does the presence of a family companion help or hinder communication? *J Gen Intern Med* 2012 Jun;27(6):661-668 [FREE Full text] [doi: [10.1007/s11606-011-1957-5](#)] [Medline: [22180197](#)]

43. Boucsein W, Fowles DC, Grimnes S, Ben-Shakhar G, Roth WT, Dawson ME, Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. Publication recommendations for electrodermal measurements. *Psychophysiology* 2012 Aug;49(8):1017-1034. [doi: [10.1111/j.1469-8986.2012.01384.x](https://doi.org/10.1111/j.1469-8986.2012.01384.x)] [Medline: [22680988](https://pubmed.ncbi.nlm.nih.gov/22680988/)]
44. Boucsein W. *Electrodermal Activity*. United States: Springer US; 2012.
45. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. *BMC Public Health* 2011 Jan;11:66 [FREE Full text] [doi: [10.1186/1471-2458-11-66](https://doi.org/10.1186/1471-2458-11-66)] [Medline: [21281471](https://pubmed.ncbi.nlm.nih.gov/21281471/)]
46. Sano A, Phillips A, McHill A, Taylor S, Barger L, Czeisler C, et al. Influence of weekly sleep regularity on self-reported wellbeing. *Sleep* 2017;40(suppl_1):A67-A68. [doi: [10.1093/sleepj/zsx050.181](https://doi.org/10.1093/sleepj/zsx050.181)]
47. Phillips AJK, Clerx WM, O'Brien CS, Sano A, Barger LK, Picard RW, et al. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Sci Rep* 2017 Jun 12;7(1):3216 [FREE Full text] [doi: [10.1038/s41598-017-03171-4](https://doi.org/10.1038/s41598-017-03171-4)] [Medline: [28607474](https://pubmed.ncbi.nlm.nih.gov/28607474/)]
48. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11, 2015; Osaka, Japan. [doi: [10.1145/2750858.2805845](https://doi.org/10.1145/2750858.2805845)]
49. Jaques N, Taylor S, Azaria A, Ghandeharioun A, Sano A, Picard R. Predicting students' happiness from physiology, phone, mobility, and behavioral data. 2015 Presented at: International Conference on Affective Computing and Intelligent Interaction (ACII); September 21-24, 2015; Xi'an, China URL: <http://europepmc.org/abstract/MED/28515966> [doi: [10.1109/ACII.2015.7344575](https://doi.org/10.1109/ACII.2015.7344575)]
50. Tibshirani R. Regression selection shrinkage via the Lasso. *J R Stat Soc Series B Stat Methodol* 2011;73(3):273-282. [doi: [10.1111/j.1467-9868.2011.00771.x](https://doi.org/10.1111/j.1467-9868.2011.00771.x)]
51. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell* 1997 Dec;97(1-2):273-324. [doi: [10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)] [Medline: [356583](https://pubmed.ncbi.nlm.nih.gov/356583/)]
52. Forman G, Scholz M. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *SIGKDD Explor* 2010;12:49-57. [doi: [10.1145/1882471.1882479](https://doi.org/10.1145/1882471.1882479)]
53. Agresti A, Coull BA. Approximate is better than "Exact" for interval estimation of binomial proportions. *Am Stat* 1998 May;52(2):119-126. [doi: [10.1080/00031305.1998.10480550](https://doi.org/10.1080/00031305.1998.10480550)]
54. Healey JA, Picard RW. Affect.media.mit.edu. 2005. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors URL: <https://affect.media.mit.edu/pdfs/05.healey-picard.pdf> [accessed 2018-05-18] [WebCite Cache ID 6zVI8MWT1]
55. Hernandez J, Morris RR, Picard RW. Call Center Stress Recognition with Person-Specific Models. 2011 Presented at: International Conference on Affective Computing and Intelligent Interaction; October 09-12, 2011; Memphis, USA. [doi: [10.1007/978-3-642-24600-5_16](https://doi.org/10.1007/978-3-642-24600-5_16)]
56. Setz C, Arnrich B, Schumm J, La Marca R, Tröster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans Inf Technol Biomed* 2010 Mar;14(2):410-417. [doi: [10.1109/TITB.2009.2036164](https://doi.org/10.1109/TITB.2009.2036164)] [Medline: [19906598](https://pubmed.ncbi.nlm.nih.gov/19906598/)]
57. Ward NG, Doerr HO, Storrie MC. Skin conductance: a potentially sensitive test for depression. *Psychiatry Res* 1983 Dec;10(4):295-302. [Medline: [6583718](https://pubmed.ncbi.nlm.nih.gov/6583718/)]
58. Argyle N. Skin conductance levels in panic disorder and depression. *J Nerv Ment Dis* 1991 Sep;179(9):563-566. [Medline: [1919559](https://pubmed.ncbi.nlm.nih.gov/1919559/)]
59. Rottenberg J. Mood and emotion in major depression. *Curr Dir Psychol Sci* 2005;14(3):167-170. [doi: [10.1111/j.0963-7214.2005.00354.x](https://doi.org/10.1111/j.0963-7214.2005.00354.x)]
60. Nutt D, Wilson S, Paterson L. Sleep disorders as core symptoms of depression. *Dialogues Clin Neurosci* 2008;10(3):329-336 [FREE Full text] [Medline: [18979946](https://pubmed.ncbi.nlm.nih.gov/18979946/)]
61. Abad VC, Guilleminault C. Sleep and psychiatry. *Dialogues Clin Neurosci* 2005;7(4):291-303 [FREE Full text] [Medline: [16416705](https://pubmed.ncbi.nlm.nih.gov/16416705/)]
62. Vinkers CH, Penning R, Hellhammer J, Verster JC, Klaessens JH, Olivier B, et al. The effect of stress on core and peripheral body temperature in humans. *Stress* 2013 Sep;16(5):520-530. [doi: [10.3109/10253890.2013.807243](https://doi.org/10.3109/10253890.2013.807243)] [Medline: [23790072](https://pubmed.ncbi.nlm.nih.gov/23790072/)]
63. Sano A, Picard RW. Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data. *Conf Proc IEEE Eng Med Biol Soc* 2014;2014:930-933 [FREE Full text] [doi: [10.1109/EMBC.2014.6943744](https://doi.org/10.1109/EMBC.2014.6943744)]
64. Barbini B, Benedetti F, Colombo C, Guglielmo E, Campori E, Smeraldi E. Perceived mood and skin body temperature rhythm in depression. *Eur Arch Psychiatry Clin Neurosci* 1998;248(3):157-160. [Medline: [9728735](https://pubmed.ncbi.nlm.nih.gov/9728735/)]
65. Zschucke E, Gaudlitz K, Ströhle A. Exercise and physical activity in mental disorders: clinical and experimental evidence. *J Prev Med Public Health* 2013 Jan;46 Suppl 1:S12-S21 [FREE Full text] [doi: [10.3961/jpmph.2013.46.S.S12](https://doi.org/10.3961/jpmph.2013.46.S.S12)] [Medline: [23412549](https://pubmed.ncbi.nlm.nih.gov/23412549/)]
66. Vollrath M. Personality and stress. *Scand J Psychol* 2001;42(4):335-347. [Medline: [11547909](https://pubmed.ncbi.nlm.nih.gov/11547909/)]
67. Ebstrup JF, Eplov LF, Pisinger C, Jørgensen T. Association between the Five Factor personality traits and perceived stress: is the effect mediated by general self-efficacy? *Anxiety Stress Coping* 2011 Jul;24(4):407-419. [doi: [10.1080/10615806.2010.540012](https://doi.org/10.1080/10615806.2010.540012)] [Medline: [21213153](https://pubmed.ncbi.nlm.nih.gov/21213153/)]

68. van Straten A, Cuijpers P, van Zuuren FJ, Smits N, Donker M. Personality traits and health-related quality of life in patients with mood and anxiety disorders. *Qual Life Res* 2007 Feb;16(1):1-8 [[FREE Full text](#)] [doi: [10.1007/s11136-006-9124-x](https://doi.org/10.1007/s11136-006-9124-x)] [Medline: [17033892](#)]
69. Götz FM, Stieger S, Reips UD. Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PLoS One* 2017 May;12(5):e0176921 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0176921](https://doi.org/10.1371/journal.pone.0176921)] [Medline: [28467473](#)]
70. Mayor E. Gender roles and traits in stress and health. *Front Psychol* 2015 Jun;6:779 [[FREE Full text](#)] [doi: [10.3389/fpsyg.2015.00779](https://doi.org/10.3389/fpsyg.2015.00779)] [Medline: [26106354](#)]
71. Zuckerman DM. Stress, self-esteem, and mental health: How does gender make a difference? *Sex Roles* 1989 Apr;20(7-8):429-444. [doi: [10.1007/BF00288001](https://doi.org/10.1007/BF00288001)]
72. Sigmon ST, Pells JJ, Boulard NE, Whitcomb-Smith S, Edenfield TM, Hermann BA, et al. Gender differences in self-reports of depression: the response bias hypothesis revisited. *Sex Roles* 2005;53(5-6):401-411. [doi: [10.1007/s11199-005-6762-3](https://doi.org/10.1007/s11199-005-6762-3)]
73. Sandanger I, Nygård JF, Sørensen T, Moum T. Is women's mental health more susceptible than men's to the influence of surrounding stress? *Soc Psychiatry Psychiatr Epidemiol* 2004 Mar;39(3):177-184. [doi: [10.1007/s00127-004-0728-6](https://doi.org/10.1007/s00127-004-0728-6)] [Medline: [14999449](#)]
74. Verma R, Balhara YP, Gupta CS. Gender differences in stress response: role of developmental and biological determinants. *Ind Psychiatry J* 2011 Jan;20(1):4-10 [[FREE Full text](#)] [doi: [10.4103/0972-6748.98407](https://doi.org/10.4103/0972-6748.98407)] [Medline: [22969173](#)]

Abbreviations

ACC: acceleration
E-diary: electronic diary
FDR: false discovery rate
HRV: heart rate variability
LASSO: least absolute shrinkage and selection operator
MCS: mental component summary
MIT: Massachusetts Institute of Technology
PHQ-9: patient health questionnaire-9
PSS: perceived stress scores
RBF: radial basis function
REM: rapid eye movement
SC: skin conductance
SF-12: 12-Item Short Form Health Survey
SMS: short message service
SRI: Sleep Regularity Index
ST: skin temperature
SVM: support vector machine
SWS: slow-wave sleep

Edited by G Eysenbach; submitted 13.11.17; peer-reviewed by S Saeb, L Barnes, C Gorrostieta, J Apolinário-Hagen; comments to author 16.12.17; revised version received 24.02.18; accepted 22.04.18; published 08.06.18.

Please cite as:

Sano A, Taylor S, McHill AW, Phillips AJK, Barger LK, Klerman E, Picard R
 Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study
J Med Internet Res 2018;20(6):e210
 URL: <http://www.jmir.org/2018/6/e210/>
 doi: [10.2196/jmir.9410](https://doi.org/10.2196/jmir.9410)
 PMID: [29884610](https://pubmed.ncbi.nlm.nih.gov/29884610/)

©Akane Sano, Sara Taylor, Andrew W McHill, Andrew JK Phillips, Laura K Barger, Elizabeth Klerman, Rosalind Picard. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.06.2018. 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.

Original Paper

Multicomponent mHealth Intervention for Large, Sustained Change in Multiple Diet and Activity Risk Behaviors: The Make Better Choices 2 Randomized Controlled Trial

Bonnie Spring¹, PhD; Christine Pellegrini¹, PhD; H G McFadden¹, BS; Angela Fidler Pfammatter¹, PhD; Tammy K Stump¹, PhD; Juned Siddique¹, DrPH; Abby C King², PhD; Donald Hedeker³, PhD

¹Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

²Health Research and Policy, Stanford University School of Medicine, Stanford, CA, United States

³Department of Public Health Sciences, The University of Chicago, Chicago, IL, United States

Corresponding Author:

Bonnie Spring, PhD

Feinberg School of Medicine, Northwestern University

Department of Preventive Medicine

680 N Lakeshore Drive

Suite 1400

Chicago, IL, 60611

United States

Phone: 1 312 503 6579

Email: bspring@northwestern.edu

Abstract

Background: Prevalent co-occurring poor diet and physical inactivity convey chronic disease risk to the population. Large magnitude behavior change can improve behaviors to recommended levels, but multiple behavior change interventions produce small, poorly maintained effects.

Objective: The Make Better Choices 2 trial tested whether a multicomponent intervention integrating mHealth, modest incentives, and remote coaching could sustainably improve diet and activity.

Methods: Between 2012 and 2014, the 9-month randomized controlled trial enrolled 212 Chicago area adults with low fruit and vegetable and high saturated fat intakes, low moderate to vigorous physical activity (MVPA) and high sedentary leisure screen time. Participants were recruited by advertisements to an open-access website, screened, and randomly assigned to either of two active interventions targeting MVPA simultaneously with, or sequentially after other diet and activity targets (N=84 per intervention) or a stress and sleep contact control intervention (N=44). They used a smartphone app and accelerometer to track targeted behaviors and received personalized remote coaching from trained paraprofessionals. Perfect behavioral adherence was rewarded with an incentive of US \$5 per week for 12 weeks. Diet and activity behaviors were measured at baseline, 3, 6, and 9 months; primary outcome was 9-month diet and activity composite improvement.

Results: Both simultaneous and sequential interventions produced large, sustained improvements exceeding control ($P<.001$), and brought all diet and activity behaviors to guideline levels. At 9 months, the interventions increased fruits and vegetables by 6.5 servings per day (95% CI 6.1–6.8), increased MVPA by 24.7 minutes per day (95% CI 20.0–29.5), decreased sedentary leisure by 170.5 minutes per day (95% CI –183.5 to –157.5), and decreased saturated fat intake by 3.6% (95% CI –4.1 to –3.1). Retention through 9-month follow-up was 82.1%. Self-monitoring decreased from 96.3% of days at baseline to 72.3% at 3 months, 63.5% at 6 months, and 54.6% at 9 months ($P<.001$). Neither attrition nor decline in self-monitoring differed across intervention groups.

Conclusions: Multicomponent mHealth diet and activity intervention involving connected coaching and modest initial performance incentives holds potential to reduce chronic disease risk.

Trial Registration: ClinicalTrials.gov NCT01249989; <https://clinicaltrials.gov/ct2/show/NCT01249989> (Archived by WebCite at <https://clinicaltrials.gov/ct2/show/NCT01249989>).

(*J Med Internet Res* 2018;20(6):e10528) doi:[10.2196/10528](https://doi.org/10.2196/10528)

KEYWORDS

health behavior; risk factors; mobile health; behavioral medicine; randomized controlled trial

Introduction

Public health advocates have long endorsed an approach that prioritizes achieving small risk reductions for most of the population over large improvements for the minority of the population who are at high risk [1-5]. A middle-road between the population and high-risk approaches is emerging from the realization that much of the population lacking biologic cardiometabolic risk factors is not truly at low risk [6,7]; instead, they manifest equally impactful behavioral risk factors that warrant targeting for primordial prevention of disease [8-12].

The average adult reports at least two chronic disease risk behaviors; 25% report three or more; and the magnitude of behavior change needed to bring each risk factor into compliance with public health guidance is typically large [13-16]. Unhealthy diet and activity behaviors are the most prevalent lifestyle risks. Fewer than 15% of US adults eat five or more servings of fruits and vegetables daily; median intake is about half that amount [17]. Only 29% meet dietary guidelines to consume less than 10% calories from saturated fats [18-20]. Half fall short of public health recommendations for moderate to vigorous physical activity (MVPA) [21,22], and more than 50% exceed two hours per day watching television [21,23]. These four behaviors associate separately with heightened risk of cardiovascular disease and cancers, and link with others that predict premature mortality [9,10,24-30].

To date, the diet and activity changes produced by most multiple behavior change interventions have been small and poorly maintained. Conversely, when interventions have produced large initial behavior changes, long term effectiveness has been greater. However, skepticism persists about whether individuals without disease can be motivated to make and maintain large behavioral changes. We hypothesize that community dwelling adults with multiple diet and activity risk behaviors could be activated to achieve and maintain guideline levels of these behaviors by a scalable, multicomponent intervention that integrates mHealth technology, modest incentives, and remote connected coaching.

Inclusion of intervention components was guided by three principles—effectiveness, scalability, and synergy. Telephone coaching was used because the approach has demonstrated effectiveness and greater reach than in-person counseling [31,32]. On the other hand, although remote coaching is more scalable than in-person treatment, it produces smaller behavior changes [32]. Because larger magnitude behavior changes are maintained better [33], modest incentives were used that would maximize initial behavior changes [34,35]. In a prior study, we observed that incentives motivated participants to make changes of greater magnitude than they thought they could accomplish, and after being successful, most tried to maintain gains [36]. A smartphone app and accelerometer were used to provide diet and physical activity feedback synchronously to participants and their coaches, enabling connected, maximally personalized, adaptive coaching.

The Make Better Choices 2 (MBC2) trial aimed to improve upon the most effective intervention from a previous (MBC1) trial that also treated adults with four concurrent diet and activity risk behaviors: low fruit and vegetable intakes, high saturated fat intake, low MVPA, and high sedentary leisure screen time [36]. MBC1 results showed that participants incentivized for 3 weeks to increase fruits and vegetables, while decreasing leisure screen time made large sustained improvements in targeted behaviors as well as in saturated fat, which was untargeted. However, MVPA did not improve. Hence, the MBC2 trial tested whether also targeting MVPA simultaneously with or sequentially after the other diet and activity behaviors could optimize all four diet and activity behaviors, relative to a contact-control intervention.

Methods

Study Design, Population, and Procedures

The MBC2 study [37], was a three-arm prospective randomized controlled trial which compared two sequences of diet and activity intervention to a contact-control intervention. The active interventions targeted MVPA either simultaneously with (simultaneous) or sequentially after (sequential) other diet and activity risk behaviors (fruits and vegetables, sedentary leisure screen time). Saturated fat was not targeted explicitly because findings from the previous MBC1 trial showed that increasing fruits and vegetables and decreasing leisure screen time automatically lowered fat intake, as a tag-along healthy lifestyle improvement. The reduction in fat intake occurred effortlessly (without explicit goal setting), as increased fiber intake crowded out fat intake and as decreased screen time reduced the snacking with which it was usually paired [36]. Because prior findings show that trying to accomplish too many behavior changes simultaneously can be detrimental [38-40], we reduced participant burden by not setting fat goals, or providing app feedback for fat intake, or coaching about fat. Rather, the MBC2 intervention aimed to lower fat intake incidentally by increasing substitute behaviors (fruit and vegetable intake) that could crowd it out and by decreasing complementary behaviors (leisure screen time) that could cue it.

The control intervention addressed stress and sleep. Eligible participants were stratified by gender and individually randomized to a condition using randomly permuted blocks with an allocation ratio of 2:2:1 (simultaneous: sequential: control).

Adults, recruited through subway, bus, flyers, and newspaper advertisements, were referred to a website to complete online screening. Inclusion criteria were willingness to be randomized, age 18-65 years, and meeting all the following: <5 servings of fruits and vegetables per day; ≥8% daily calories from saturated fat; <150 minutes per week MVPA; >120 minutes per week of leisure screen time (ie, television, movies, videogames, recreational internet). These discretionary activities were targeted because they can be decreased without jeopardizing necessary work-related activities. Exclusion criteria were

unstable medical condition (ie, uncontrolled hypertension or diabetes), pregnancy or intent to become pregnant, anorexia, bulimia, binge eating disorder, or weight >350 lb [37].

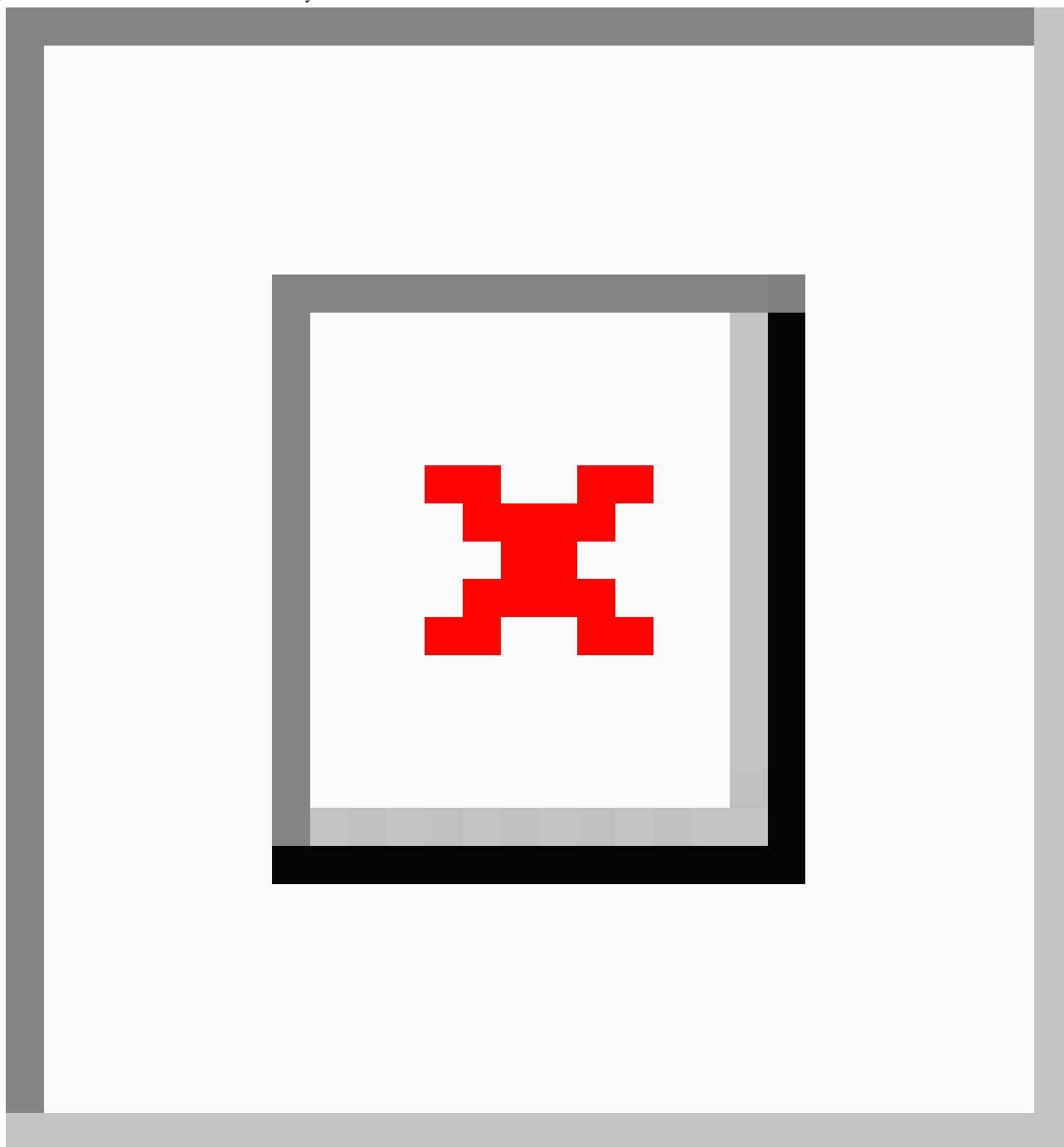
The Northwestern University Institutional Review Board approved all procedures, and the study was conducted in Chicago between July 2012 and July 2014 (see [Multimedia Appendix 2](#) for trial protocol). After screening, eligible candidates attended an in-person session where they discussed the pros and cons of the three treatment options, provided written informed consent, and were loaned a smartphone and accelerometer. They were trained to estimate portion sizes, use the assessment version of a custom-built smartphone app to record behaviors (dietary intake, leisure screen time, stress level, relaxation exercises, and sleep), and wear an accelerometer for a baseline week. Instructions emphasized entering all meals and snacks immediately after eating and using sliders to show accumulated leisure screen time four times daily. Study apps processed dietary data through the integrated CalorieKing food database with all fruits and vegetable items tagged as serving sizes [41]. The Shimmer accelerometer [42], worn in a spibelt around the waist, recorded activity counts and wirelessly transmitted data through Bluetooth to the app, which converted counts to MVPA minutes. Following the in-person session, all eligible participants were sent home with a smartphone, app, and accelerometer to monitor themselves for seven to ten days. The self-monitored data were then assessed at the end of the baseline period and those who exhibited all four diet and activity risk behaviors throughout baseline were randomized.

Participants learned their treatment assignment by downloading one of three custom-built, intervention-specific study apps [37]. They were asked to use the app and accelerometer to

continuously record those behaviors the intervention targeted (ie, dietary intake, sedentary leisure, and MVPA for the simultaneous and aequial conditions; stress, relaxation exercises, and sleep for control) throughout the 9-month study period. Unlike the assessment app, which collected user self-reports and accelerometer data but gave no user feedback, intervention apps provided users with continuously updated feedback about their performance of targeted behaviors relative to goal. The app user interfaces are shown in [Figure 1](#). In addition to giving participants goal attainment feedback for targeted behaviors, apps wirelessly transmitted this information to coaches, who used it to tailor telephone counseling. Sequential and simultaneous apps were similar, except that the physical activity interface for sequential treatment remained inactive until week seven. End goals for the 12-week simultaneous or sequential intervention were: 1) ≤ 90 min per day of sedentary leisure screen time; 2) ≥ 5 servings of fruits and vegetables; and 3) ≥ 150 min per week of MVPA. Those receiving simultaneous treatment were asked to gradually modify all three target behaviors from the outset of the intervention. Those receiving sequential treatment were asked to modify only sedentary leisure screen time and fruit and vegetables for the first 6 weeks. Between weeks 7 and 12, they were asked to maintain goal levels for leisure screen time and fruit and vegetables, while progressively increasing MVPA. Control participants were coached to perform three relaxation exercises per day (a progressive muscle relaxation technique, a mindfulness meditation, and a self-hypnosis technique) [43], and to achieve end goals of ≥ 7.5 hours of sleep per day and a 30% reduction in stress over the 12-week intervention. Participants used a 1 to 10 Subjective Units of Distress Scale (SUDS) to record stress, as shown in [Figure 1](#).

Figure 1. Make Better Choices 2 app user interfaces for (a) receiving behavioral feedback in simultaneous and sequential treatments; (b) reporting sedentary leisure screen time in simultaneous and sequential treatments; (c) receiving feedback in contact control treatment.



Figure 2. Make Better Choices 2 Trial study timeline.

During treatment initiation (weeks 1-12), a trained paraprofessional telephoned each participant weekly for a 15 minute coaching session. Coaches delivered a sequence of online didactic lessons specific to each condition [37] and used motivational interviewing to tailor counseling using data from the participant's app and accelerometer. Coaching call frequency decreased to biweekly in weeks 13-24 and monthly in weeks 25-40, and call duration decreased to 10 minutes. The study timeline is shown in Figure 2. All coaching sessions were audiotaped; a 10% random sample selected quarterly was coded for treatment fidelity by blinded raters. Coaches scoring less than 90% were retrained and certified before making additional calls.

Study Outcomes and Statistical Analysis

The primary trial outcome was the composite diet and activity improvement score measured for 1-week assessment periods at 3, 6, and 9 months when participants wore an accelerometer and used the assessment app to self-monitor their behaviors without receiving any feedback. Secondary outcomes were healthy changes in fruits and vegetables, saturated fat, MVPA, and leisure screen time. Measuring diet and activity by recording on a mobile device has shown acceptable reliability and validity and greater adherence compared to paper reporting or recall [44-46].

Changes in fruit and vegetable intakes, saturated fat intake, sedentary leisure screen time (derived from app self-report),

and MVPA (derived from accelerometry supplemented by app self-report in the event of water activities or battery failure) were aggregated to calculate the MBC composite diet and activity improvement score, weighting each behavior equally [36]. Square root transformation was applied to count data (fruit and vegetables, MVPA, sedentary leisure screen time) and arcsine transformation was applied to the percentage outcome (fat) to better approximate normality.

Behaviors were placed on a common scale by standardizing each one using a modified Z-score (where one U represents a one-SD change) based on the baseline sample means and standard deviations of each behavior; higher values represent greater healthy lifestyle improvement relative to the overall baseline distribution. The MBC composite diet and activity improvement score [36], expressing each participant's overall healthy change across the multiple diet and activity behaviors relative to baseline is calculated as the mean of all four individual Z scores at each time point.

Three components of treatment implementation were assessed: fidelity, receipt, and adherence [47]. Fidelity was considered present for sessions when the coach delivered all required treatment elements correctly (eg, encouraging a control participant to go to bed earlier) and absent if the coach delivered any incorrect treatment element (eg, encouraging a control participant to exercise to be tired at bedtime). Treatment receipt was measured by the percent of completed scheduled coaching calls. Self-monitoring adherence was assessed by the proportion of days that participant used the app to record targeted behaviors for 1-week assessments at baseline, 3, 6, and 9 months. Goal attainment was assessed by the proportion of weeks during treatment initiation when the participant met behavioral goals. To be credited with having met goal and to earn an incentive for the week, participants' average scores for all the three target behaviors needed to meet or exceed goal level.

The average effect size (mean difference in composite diet and activity improvement Z-score divided by common standard deviation) in our previous MBC1 trial equaled 0.46. Based on power calculations, we aimed to recruit 50 control subjects and 100 subjects into each of the two intervention groups, assuming a correlation of 0.50 for the composite Z-scores across time and an attrition rate of 20% at the final time point. We powered the study for an effect size in the range of 0.5 for the first Helmert contrast (H1: simultaneous + sequential vs control) and 0.4 for the second Helmert contrast (H2: simultaneous vs sequential). Bankruptcy of the mobile phone service provider necessitated the return and provision of new mobile phones, creating a budget shortfall that required reducing enrollment from 250 to 212 participants: 84 allocated to the simultaneous intervention; 84 allocated to sequential; and 44 to control. However, because the observed correlation of the composite Z-scores over time was smaller ($r=0.44$) than the predicted 0.50, the study remained sufficiently powered for the posited effect sizes.

Baseline characteristics were compared across groups using analysis of variance (ANOVA) for continuous variables and chi-squared tests for categorical variables. The percentage of coaching calls received and completed in the first versus the

second 6 weeks of treatment was analyzed using repeated measures ANOVA. The percentage of days participants adhered to self-monitoring was measured at baseline, 3-, 6-, and 9-months and analyzed using repeated measures ANOVA. Goal attainment (yes or no), measured every 2 weeks during treatment initiation, was analyzed using mixed-effects logistic regression with time modeled as either the first or the second 6-weeks of initiation. All models of treatment receipt, adherence, and goal attainment included group by time interactions to assess differences between groups at each time point.

Intent-to-treat analyses of primary and secondary endpoints used three-level linear mixed-effects models that treated daily measurements (level one) nested within 1-week assessment periods (level two) nested within subjects (level three). Thus, we analyzed at the daily level and considered the correlation of the daily measurements within weeks and subjects by including random subject intercept and time trends at level three (subjects), and a random intercept at level two (1-week assessment periods). For comparisons across assessment periods, we treated baseline as the reference, and estimated changes at 3-, 6-, and 9-month follow-up. For comparisons between the intervention groups Helmert contrasts were used, in which the first contrast compared the combined simultaneous and sequential groups to control, and the second contrast compared the simultaneous to the sequential group. We also included group by time interactions to assess the degree to which change from baseline varied for either of the Helmert contrasts at each follow-up.

Results

Participants were primarily female (162/212, 76.4%), minority (125/212, 59.0%), college educated (147/212, 69.3%), with a mean age of 40.8 (11.9) years and mean BMI 34.3 (8.8). The groups did not differ in their baseline characteristics (Table 1).

Of study applicants who were web-screened, 18.4% proceeded through in-person screening and on into baseline recording to verify the presence of all four risk behaviors. Most of those listed in Figure 3 as excluded for "Other" reasons only partially completed the web screener. Of candidates who underwent baseline screening, 45.9% (212/462) were randomized. Two candidates randomly assigned to the stress and sleep contact control condition failed to receive the allocated intervention: both withdrew before receiving any treatment.

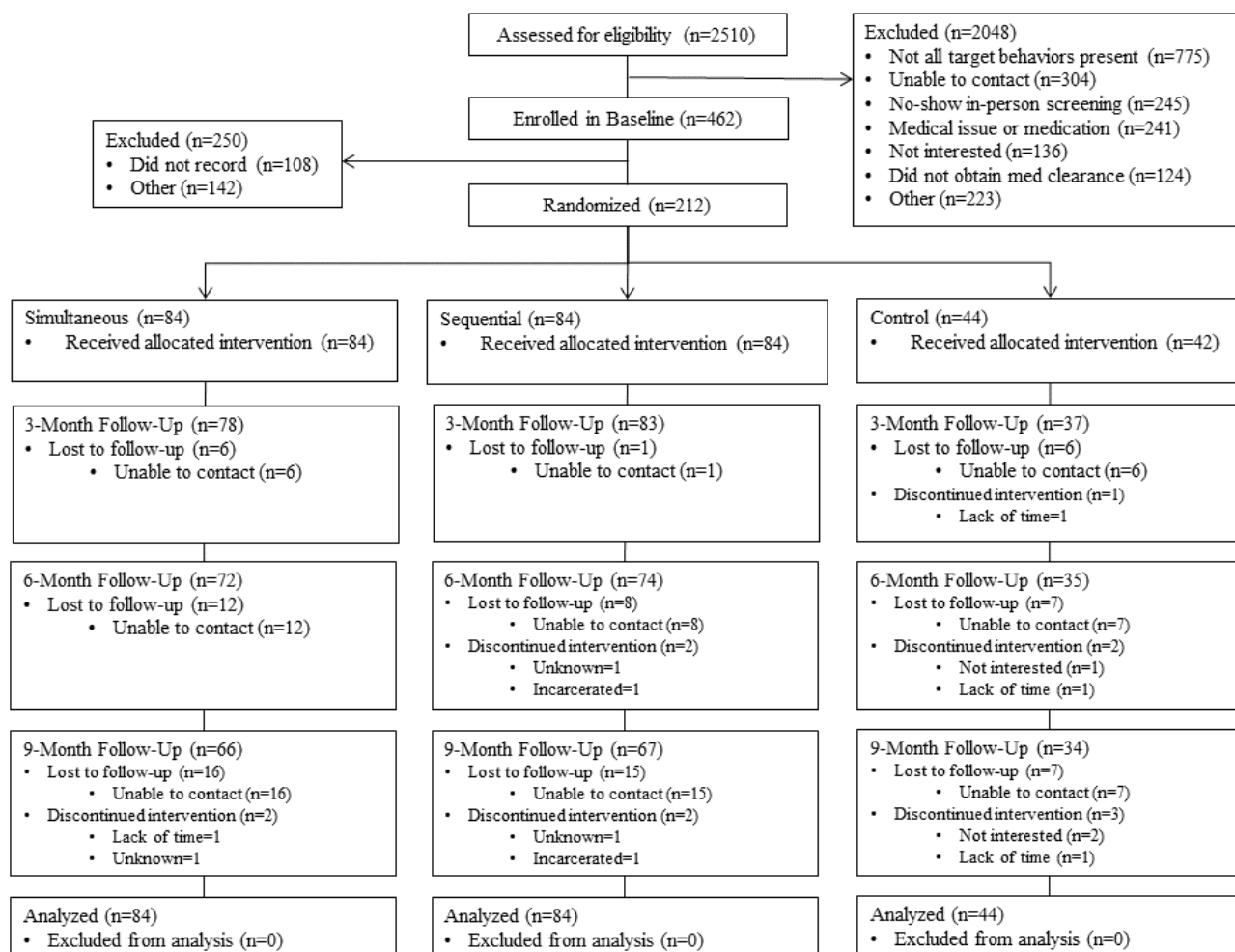
Loss to follow-up was 17.9% (38/212) and not differential across treatments (Figure 3). In the composite Z analysis, 83.5% (177/212) of participants provided a composite Z-score at two or more time points; 68.4% (145/212) provided three or more; and 50.5% (107/212) provided all four time points.

The combined simultaneous and sequential interventions produced sustained improvement, as compared to control, on the composite diet and activity score at 3, 6, and 9 months ($P<.001$; see Table 2 and Figure 4 A). Sequential treatment produced a small, significantly greater composite diet and activity improvement than simultaneous treatment at 6 months ($P=.03$); however, no differences were evident at 3 and 9 months (see Table 2 and Figure 4 B).

Table 1. Participant Baseline Characteristics.

Variable	Total (n=212)	Control (n=44)	Simultaneous (n=84)	Sequential (n=84)	Treatment group differences	
					Test	P value
Age (years), mean (SD)	40.8 (11.9)	40.8 (10.9)	40.7 (11.9)	40.9 (12.5)	F=0.003	.99
Body mass index (kg/m ²), mean (SD)	34.3 (8.8)	36.0(10.1)	33.7(9.0)	33.9(7.9)	F=1.07	.34
Gender, n (%)						
Male	50 (23.6)	11 (25.0)	20 (23.8)	19 (22.6)	$\chi^2=0.95$.95
Female	162 (76.4)	33 (75.0)	64 (76.2)	65 (77.4)		
Race, n (%)						
Caucasian	87 (41.0)	19 (43.2)	28 (33.3)	40 (47.6)	$\chi^2=9.28$.32
Black	99 (46.7)	19 (43.2)	42 (50.0)	38 (45.2)		
Asian	8 (3.8)	2 (4.5)	5 (6.0)	1 (1.2)		
Other or multiple	18 (8.5)	4 (9.1)	9 (10.7)	5 (6.0)		
Ethnicity, n (%)						
Hispanic/Latino	20 (9.8)	6 (14.0)	5 (6.3)	9 (11.0)	$\chi^2=2.04$.36
Not Hispanic/Latino	184 (90.2)	37 (86.0)	74 (93.7)	73 (89.0)		
Education, n (%)						
College degree	147 (69.3)	32 (72.7)	60 (71.4)	55 (65.5)	$\chi^2=1.00$.61
No college degree	65 (30.7)	12 (27.3)	24 (28.6)	29 (34.5)		

Figure 3. CONSORT Flow Diagram of the Make Better Choices 2 Trial: 212 adults were eligible for inclusion, randomly assigned to an intervention condition, and included in analyses.



Figures 4 and 5 show the behavior changes over time of each treatment group. Both simultaneous and sequential interventions improved each of the four behaviors to guideline levels, exceeding control at all three follow-up assessments (See Table 2). At 9 months, intervention increased the two healthful behaviors: (1) fruit and vegetable intake by 6.5 servings per day (95% CI 6.1-6.8), mean difference from control=6.2 servings per day (95% CI 5.5-6.9), and (2) MVPA by 24.7 min per day (95% CI 20.0-29.5), mean difference from control=12.1 min per day (95% CI 5.4-18.9); and decreased the two unhealthful behaviors: (1) sedentary leisure -170.5 min per day (95% CI -183.5 to -157.5), mean difference from control=-137.7 min per day (95% CI -155.9 to -119.5), and (2) saturated fat intake -3.6% (95% CI -4.1 to -3.1), mean difference from control=-3.3% (95% CI -4.3 to -2.2).

In addition to being sustained, the improvements produced by the active interventions were generally large (0.5 to 1.7 standard deviation unit improvements compared to baseline). Although Sequential treatment reduced saturated fat intake more than

simultaneous treatment at 3 months (mean difference -1.1%, 95% CI -1.8 to -0.3) and 6 months (mean difference -1.8%, 95% CI -2.6-1.0, no difference was evident at 9 months, and no other differences between simultaneous and sequential treatments were observed (Table 2).

Treatment fidelity averaged 96.8% across the 2-year study; 3 out of 20 coaches required retraining. Receipt of calls declined from 66.0% during the first half of treatment initiation to 57.7% during the second half ($F[1209]=12.05$, $P<.001$), not differing among treatment groups (61.9%, 95% CI 58.0-65.7; $P=.12$). Self-monitoring decreased from an average of 96.3% at baseline to 72.3% at 3 months, 63.5% at 6 months, and 54.6% at 9 months ($F[3627]=95.0$, $P<.001$), without differences across intervention groups ($P=.41$). Goal attainment was greater for the Active intervention groups (58.8%, 95% CI 52.2% to 65.0%) than control (33.6%, 95% CI 23.1% to 46.0%) during the first half of treatment initiation, ($z=3.46$, $P<.001$), but Active and Control groups did not differ during the last half of treatment (38.3%, 95% CI 32.6% to 44.2%; $z=0.13$, $P=.89$).

Table 2. Differences in Standardized (Z-score) Change from baseline between treatments at follow-up. Italics indicate statistical significance.

Behavioral outcomes	3-Month follow-up, mean (95% CI)	6-Month follow-up, mean (95% CI)	9-Month follow-up, mean (95% CI)
Combined vs control			
Sedentary Leisure Screen Time	<i>0.94 (0.65-1.23)^a</i>	<i>1.30 (0.95-1.64)^a</i>	<i>0.85 (0.42-1.28)^a</i>
Physical Activity	<i>0.66 (0.40-0.91)^a</i>	<i>0.86 (0.56-1.20)^a</i>	<i>0.54 (0.17-0.91)^b</i>
Fruit & Vegetable Intake	<i>1.74 (1.50-1.98)^a</i>	<i>1.72 (1.43-2.00)^a</i>	<i>1.37 (1.02-1.71)^a</i>
Saturated Fat Intake	<i>0.45 (0.19-0.72)^a</i>	<i>0.57 (0.27-0.87)^a</i>	<i>0.66 (0.32-0.99)^a</i>
Composite Diet-Activity Score	<i>0.95 (0.80-1.10)^a</i>	<i>1.16 (0.98-1.34)^a</i>	<i>0.92 (0.69-1.14)^a</i>
Simultaneous vs sequential			
Sedentary Leisure Screen Time	-0.10 (-0.34-0.14)	-0.04 (-0.33-0.25)	0.00 (-0.36-0.36)
Physical Activity	-0.14 (-0.36-0.09)	0.02 (-0.24-0.28)	0.02 (-0.30-0.34)
Fruit & Vegetable Intake	-0.06 (-0.26-0.15)	0.10 (-0.14-0.34)	-0.05 (-0.35-0.24)
Saturated Fat Intake	<i>0.37 (0.14-0.60)^a</i>	<i>0.50 (0.24-0.75)^a</i>	0.15 (-0.13-0.44)
Composite Diet-Activity Score	-0.01 (-0.13-0.12)	<i>0.16 (0.01-0.31)^c</i>	0.06 (-0.13-0.25)

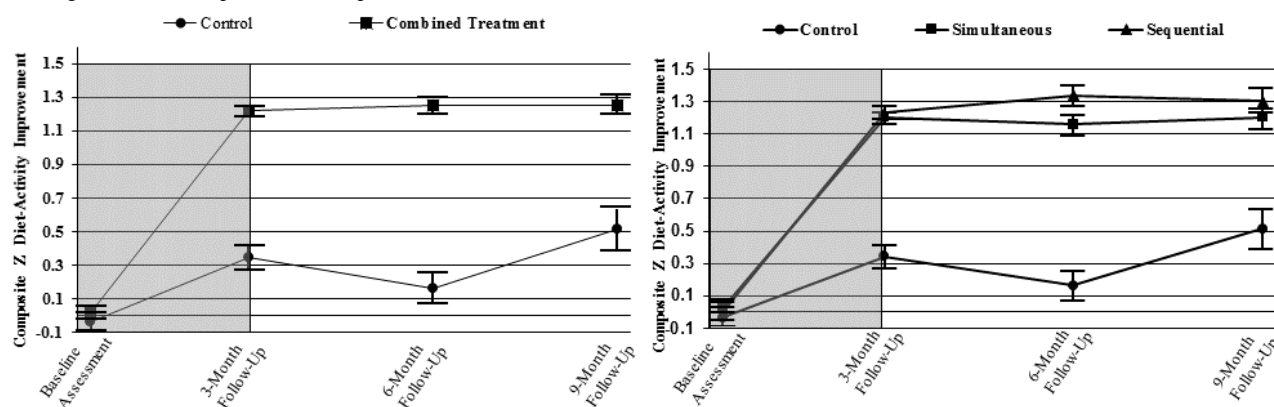
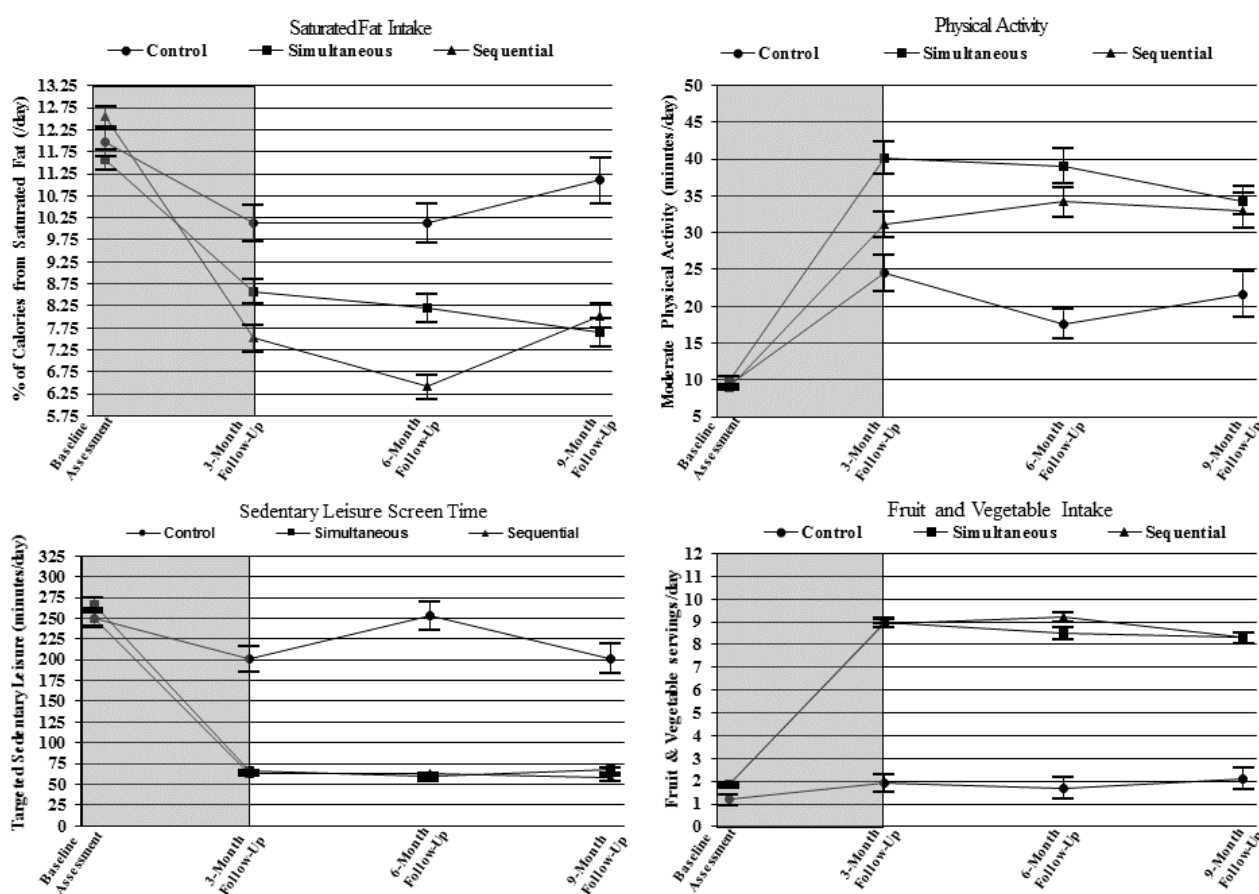
^a $P < .001$ ^b $P < .01$ ^c $P < .05$ **Figure 4.** A: Mean Composite Diet-Activity Improvement Scores over time for combined simultaneous and sequential treatment groups vs control. Combined treatment groups produced greater healthy change at each postbaseline assessment point. B: Mean Composite Diet-Activity Improvement Scores over time for each of the three conditions. Error bars represent 1 SE. Gray background indicates the treatment initiation phase (weeks 0-12); white background, follow-up maintenance phase.

Figure 5. Effects of the 3 intervention conditions on changes over time in each behavior expressed in natural units. Gray background indicates treatment initiation phase (weeks 0-12); white background, follow-up maintenance phase.



Discussion

These findings support the hypothesis that adults with multiple diet and activity risk behaviors can be activated to make and maintain large improvements in diet and activity behaviors by a scalable, multicomponent intervention that integrates mHealth technology, modest incentives, and remote-connected coaching by trained paraprofessionals. Both active MBC2 interventions produced larger sustained improvements in fruit and vegetables, saturated fat, physical activity, and sedentary leisure screen time than those observed in most prior trials, including our MBC1 study [13,36,48-51], such that all four behaviors surpassed guideline recommended levels at the final study follow-up. Also, unlike the prior MBC1 interventions, both active MBC2 interventions yielded a sustained increase in MVPA documented by accelerometry [51,52]. We attribute the greater maintenance of positive diet and activity changes in the MBC2 study to the longer intervention (12 weeks versus three) and continued availability of intervention technology with some minimal coach contact.

To our knowledge, the present study is unique in testing simultaneous vs sequential versions of a multicomponent mHealth intervention to address the four most prevalent, co-occurring diet and activity risk behaviors [38,39,49,53-56]. The findings show comparable improvement from intervening to increase MVPA simultaneously with or sequentially after targeting other diet and activity risk behaviors. Even though sequential treatment produced somewhat greater improvement

than Simultaneous treatment in healthy diet and activity at 6 months, the difference was small in magnitude and not sustained. These findings expand to four the number of co-occurring poor diet and activity habits that can demonstrably be changed simultaneously and support the conclusion that either a simultaneous or a sequential approach to multiple health behavior change can be expected to yield benefit [54,55].

The unusually large, well-maintained diet and activity improvements observed in this trial are likely attributable to the effectiveness of the MBC2 intervention components, including the use of appealing mHealth technology and connective coaching as vehicles to deploy effective behavior change techniques (goal-setting, self-monitoring, feedback, support, accountability). Unlike recent trials that provided digital feedback solely to patients and failed to find benefit from supplying a wearable accelerometer [57,58], MBC2 provided feedback synchronously to both participants and coaches, enabling synergistic benefit through connected counseling. Strong engagement with the intervention and study technology resulted, evidenced by participants self-monitoring on 50% of days even after 9 months.

The MBC2 study's US \$5 per week incentive for participants to meet behavioral goals during treatment initiation apparently had the intended effect of motivating participants to make large improvements. Notably, MBC2 participants made somewhat larger diet and activity improvements than those in the MBC1 study, even though the MBC2 incentive was two-thirds smaller. Moreover, no incentive to sustain healthful diet and activity

changes was operative in either the MBC1 or MBC2 trial; nevertheless, behavioral improvements were maintained. Hence, these findings contradict the worry that use of incentives followed by their discontinuation inevitably undermines behavioral maintenance. Results accord with a growing body of evidence showing sustained improvements after incentives cease [34,35]. Potential scalability of modest incentives is suggested by Centers for Medicare and Medicaid reimbursement of contingency contracting for some habit disorders, and by the growing number of individuals and employers that find incentives for healthy lifestyle change cost effective [59].

Strengths of the study include a strong scientific premise grounded in the high population prevalence of individuals whose multiple diet and activity risk behaviors place them at moderately high chronic disease risk. Other strengths include a rigorous, internally valid clinical trial design comparing two active interventions with a contact-matched control, and strong initial allocation concealment and treatment fidelity procedures. In addition to objective measurement of MVPA, another strength was a 9 month follow up period allowing examination of both initiation and maintenance of behavior change.

The study also had limitations. Three of four behavioral outcomes were assessed exclusively by self-report and could have been subject to demand characteristics. Fruit and vegetable consumption may have been overestimated and time spent in leisure sedentary screen time underestimated. However, although some risk of self-report bias persists, the objective measure of MVPA derived from the accelerometer also showed large, sustained improvement following the active interventions. Fruit and vegetable intake and sedentary leisure showed the largest improvements in this study, as they had in our prior MBC1 study. Notably, in the MBC1 study, improvements in these two behaviors were also unique in being accompanied by increased self-efficacy, suggesting that changing them is both feasible and empowering [60]. Although intervention benefits persisted through nine months, longer duration follow-up remains needed. A lack of sustained superiority of sequential over simultaneous treatment could have been caused by the fact that goal progression was time-dependent, rather than mastery-based. The sequential intervention added a physical activity goal at week 7, regardless of whether participants had achieved mastery of their fruit and vegetable or sedentary leisure screen time goals. It remains possible that sequential treatment could have

increased MVPA even more if the addition of this new target goal had been delayed until initial behavior targets were reached.

Treating multiple risk behaviors simultaneously is inherently more efficient than treating them sequentially, but simultaneous change might be more feasible for some population subgroups than others. In MBC2's diverse, moderately well-educated study sample, we saw no detrimental effects of intervening on multiple diet and activity behaviors all at once, rather than one at a time. However, in a different trial, people with a greater number of risk behaviors were more likely to drop out when treated simultaneously rather than sequentially [38], an effect that could reflect the association between multiplicity of risk behaviors and social disadvantage [61,62]. If overzealous intervention on too many risk behaviors at once disproportionately drives off marginalized, resource-poor subgroups, there is risk that preventive intervention will fail to reach those in greatest need of help. Caution remains warranted before inferring that all subpopulations and contexts can accomplish unlimited behavior changes at once because: 1) simultaneous (vs sequential) intervention has yielded higher relapse and dropout in some trials; 2) the number of behaviors that can be changed at once remains unknown; and 3) some risk behaviors may be disproportionately hard to change concurrently with others [39,56,63-67].

The finding that an integrated multicomponent connective mHealth intervention produced large, sustained changes in multiple diet and activity behaviors over time is encouraging. Smartphones and wearable sensors are becoming increasingly ubiquitous, equipping consumers with real-world tools that have the potential to support healthy lifestyle changes. Interventions like the present one that combine patient facing technology with digitally connected, personalized coaching support have shown promise [36,68,69]. Including behavior change coaching as a service provided by trained paraprofessionals or artificially intelligent agents could soon make technology-supported behavioral interventions a scalable part of the health care system.

A mobile health intervention integrating a smartphone app, accelerometer, modest initial performance incentives, and remote connected coaching can produce large sustained improvements in multiple prevalent diet and activity risk behaviors, whether physical activity is targeted simultaneously with or sequentially after other risk behaviors.

Acknowledgments

This study was funded by National Heart, Lung, and Blood grant HL075451 to BS and by a career development award (CA154862) to JS. TKS acknowledges salary support from NIH/NCI training grant T32 CA193193. Research reported in this publication was supported in part, by the National Cancer Institute's Robert H Lurie Comprehensive Cancer Center, Grant Number P30CA60553 and by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number UL1TR000150. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

We are grateful to the study participants who joined the trial. We also thank the Data and Safety Monitoring Committee and project staff who provided coaching or performed participant training and assessments.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 539KB - jmir_v20i6e10528_app1.pdf](#)]

Multimedia Appendix 2

Published clinical trial protocol.

[[PDF File \(Adobe PDF File\), 388KB - jmir_v20i6e10528_app2.pdf](#)]

References

1. Hill JO, Wyatt HR, Reed GW, Peters JC. Obesity and the environment: where do we go from here? *Science* 2003 Feb 7;299(5608):853-855. [doi: [10.1126/science.1079857](#)] [Medline: [12574618](#)]
2. Rodearmel SJ, Wyatt HR, Stroebele N, Smith SM, Ogden LG, Hill JO. Small changes in dietary sugar and physical activity as an approach to preventing excessive weight gain: the America on the Move family study. *Pediatrics* 2007 Oct;120(4):e869-e879. [doi: [10.1542/peds.2006-2927](#)] [Medline: [17908743](#)]
3. Lutes LD, Daiss SR, Barger SD, Read M, Steinbaugh E, Winett RA. Small changes approach promotes initial and continued weight loss with a phone-based follow-up: 9-month outcomes from ASPIRES II. *Am J Health Promot* 2012;26(4):235-238. [doi: [10.4278/ajhp.090706-QUAN-216](#)] [Medline: [22375574](#)]
4. Eldridge JD, Devine CM, Wethington E, Aceves L, Phillips-Caesar E, Wansink B, et al. Environmental influences on small eating behavior change to promote weight loss among Black and Hispanic populations. *Appetite* 2016 Jan 01;96:129-137 [FREE Full text] [doi: [10.1016/j.appet.2015.09.011](#)] [Medline: [26368577](#)]
5. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001 Jun;30(3):427-432. [Medline: [11416056](#)]
6. Brown WJ, Hockey R, Dobson A. Rose revisited: a "middle road"; prevention strategy to reduce noncommunicable chronic disease risk. *Bull World Health Organ* 2007 Nov;85(11):886-887 [FREE Full text] [Medline: [18038080](#)]
7. Feigin VL, Norrving B, Mensah GA. Primary prevention of cardiovascular disease through population-wide motivational strategies: insights from using smartphones in stroke prevention. *BMJ Glob Health* 2016;2(2):e000306 [FREE Full text] [doi: [10.1136/bmjgh-2017-000306](#)] [Medline: [28589034](#)]
8. Folsom AR, Yatsuya H, Nettleton JA, Lutsey PL, Cushman M, Rosamond WD, ARIC Study Investigators. Community prevalence of ideal cardiovascular health, by the American Heart Association definition, and relationship with cardiovascular disease incidence. *J Am Coll Cardiol* 2011 Apr 19;57(16):1690-1696 [FREE Full text] [doi: [10.1016/j.jacc.2010.11.041](#)] [Medline: [21492767](#)]
9. Lloyd-Jones D, Hong Y, Labarthe D, Mozaffarian D, Appel L, Van HL. Defining and setting national goals for cardiovascular health promotion and disease reduction: The American Heart Association's strategic impact goal through 2020 and beyond. *Circulation* 2010 Feb 02;121(4):586-613.
10. Spring B, Ockene JK, Gidding SS, Mozaffarian D, Moore S, Rosal MC, et al. Better population health through behavior change in adults: a call to action. *Circulation* 2013 Nov 5;128(19):2169-2176 [FREE Full text] [doi: [10.1161/01.cir.0000435173.25936.e1](#)] [Medline: [24100544](#)]
11. Gooding HC, Ning H, Gillman MW, Shay C, Allen N, Goff DC, et al. Application of a Lifestyle-Based Tool to Estimate Premature Cardiovascular Disease Events in Young Adults: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. *JAMA Intern Med* 2017 Sep 01;177(9):1354-1360. [doi: [10.1001/jamainternmed.2017.2922](#)] [Medline: [28715555](#)]
12. Claas SA, Arnett DK. The Role of Healthy Lifestyle in the Primordial Prevention of Cardiovascular Disease. *Curr Cardiol Rep* 2016 Jun;18(6):56. [doi: [10.1007/s11886-016-0728-7](#)] [Medline: [27142061](#)]
13. Meader N, King K, Wright K, Graham HM, Petticrew M, Power C, et al. Multiple Risk Behavior Interventions: Meta-analyses of RCTs. *Am J Prev Med* 2017 Jul;53(1):e19-e30 [FREE Full text] [doi: [10.1016/j.amepre.2017.01.032](#)] [Medline: [28258777](#)]
14. Schuit AJ, van LAJM, Tijhuis M, Ocké M. Clustering of lifestyle risk factors in a general adult population. *Prev Med* 2002 Sep;35(3):219-224. [Medline: [12202063](#)]
15. Chou K. The prevalence and clustering of four major lifestyle risk factors in Hong Kong Chinese older adults. *J Aging Health* 2008 Oct;20(7):788-803. [doi: [10.1177/0898264308321082](#)] [Medline: [18562762](#)]
16. Baruth M, Wilcox S, Egan BM, Dowda M, Laken M, Warren TY. Cardiovascular disease risk factor clustering among African American adults. *Ethn Dis* 2011;21(2):129-134 [FREE Full text] [Medline: [21749014](#)]
17. Lee-Kwan SH, Moore LV, Blanck HM, Harris DM, Galuska D. Disparities in State-Specific Adult Fruit and Vegetable Consumption - United States, 2015. *MMWR Morb Mortal Wkly Rep* 2017 Nov 17;66(45):1241-1247 [FREE Full text] [doi: [10.15585/mmwr.mm6645a1](#)] [Medline: [29145355](#)]

18. US Department of Health and Human Services; US Department of Agriculture. US Dept of Health and Human Services. 2015. Dietary Guidelines for Americans 8th ed URL: <http://www.webcitation.org/6yGHYbgFi>
19. US Department of Health and Human Services, US Department of Agriculture. 2015-2020 Dietary Guidelines for Americans 8th ed. 2015. URL: <http://www.health.gov/DietaryGuidelines> [WebCite Cache ID 6zSW0hAyz]
20. Sacks FM, Lichtenstein AH, Wu JHY, Appel LJ, Creager MA, Kris-Etherton PM, American Heart Association. Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association. *Circulation* 2017 Jul 18;136(3):e1-e23. [doi: [10.1161/CIR.0000000000000510](https://doi.org/10.1161/CIR.0000000000000510)] [Medline: [28620111](https://pubmed.ncbi.nlm.nih.gov/28620111/)]
21. Fedewa SA, Sauer AG, Siegel RL, Jemal A. Prevalence of major risk factors and use of screening tests for cancer in the United States. *Cancer Epidemiol Biomarkers Prev* 2015 Apr;24(4):637-652 [FREE Full text] [doi: [10.1158/1055-9965.EPI-15-0134](https://doi.org/10.1158/1055-9965.EPI-15-0134)] [Medline: [25834147](https://pubmed.ncbi.nlm.nih.gov/25834147/)]
22. Clarke TC, Ward BW, Norris T, Schiller JS. Early release of selected estimates based on data from the National Health Interview Survey, January-September 2016: Lack of health insurance coverage and type of coverage. 2017. URL: <https://stacks.cdc.gov/view/cdc/28996> [accessed 2018-05-16] [WebCite Cache ID 6zSWWfEb6]
23. Bowman SA. Television-viewing characteristics of adults: correlations to eating practices and overweight and health status. *Prev Chronic Dis* 2006 Apr;3(2):A38 [FREE Full text] [Medline: [16539779](https://pubmed.ncbi.nlm.nih.gov/16539779/)]
24. Myint PK, Luben RN, Wareham NJ, Bingham SA, Khaw K. Combined effect of health behaviours and risk of first ever stroke in 20,040 men and women over 11 years' follow-up in Norfolk cohort of European Prospective Investigation of Cancer (EPIC Norfolk): prospective population study. *BMJ* 2009 Feb 19;338:b349 [FREE Full text] [Medline: [19228771](https://pubmed.ncbi.nlm.nih.gov/19228771/)]
25. Mendis S, Davis S, Norrving B. Organizational update: the world health organization global status report on noncommunicable diseases 2014; one more landmark step in the combat against stroke and vascular disease. *Stroke* 2015 May;46(5):e121-e122 [FREE Full text] [doi: [10.1161/STROKEAHA.115.008097](https://doi.org/10.1161/STROKEAHA.115.008097)] [Medline: [25873596](https://pubmed.ncbi.nlm.nih.gov/25873596/)]
26. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Executive Summary: Heart Disease and Stroke Statistics—2015 Update. *Circulation* 2015 Jan 26;131(4):434-441. [doi: [10.1161/CIR.0000000000000157](https://doi.org/10.1161/CIR.0000000000000157)]
27. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *Lancet* 2014 Jul 5;384(9937):45-52. [doi: [10.1016/S0140-6736\(14\)60648-6](https://doi.org/10.1016/S0140-6736(14)60648-6)] [Medline: [24996589](https://pubmed.ncbi.nlm.nih.gov/24996589/)]
28. Arena R, Guazzi M, Lianov L, Whitsel L, Berra K, Lavie C, et al. Healthy lifestyle interventions to combat noncommunicable disease—A novel nonhierarchical connectivity model for key stakeholders: A policy statement from the American Heart Association, European Society of Cardiology, European Association for Cardiovascular Prevention and Rehabilitation, and American College of Preventive Medicine. *Mayo Clin Proc* 2015 Aug;90(8):1082-1103 [FREE Full text] [doi: [10.1016/j.mayocp.2015.05.001](https://doi.org/10.1016/j.mayocp.2015.05.001)] [Medline: [26143646](https://pubmed.ncbi.nlm.nih.gov/26143646/)]
29. Spring B, Moller AC, Colangelo LA, Siddique J, Roehrig M, Daviglius ML, et al. Healthy lifestyle change and subclinical atherosclerosis in young adults: Coronary Artery Risk Development in Young Adults (CARDIA) study. *Circulation* 2014 Jul 1;130(1):10-17. [doi: [10.1161/CIRCULATIONAHA.113.005445](https://doi.org/10.1161/CIRCULATIONAHA.113.005445)] [Medline: [24982115](https://pubmed.ncbi.nlm.nih.gov/24982115/)]
30. Adams ML, Grandpre J, Katz DL, Shenson D. Linear association between number of modifiable risk factors and multiple chronic conditions: Results from the Behavioral Risk Factor Surveillance System. *Prev Med* 2017 Dec;105:169-175. [doi: [10.1016/j.ypmed.2017.09.013](https://doi.org/10.1016/j.ypmed.2017.09.013)] [Medline: [28917949](https://pubmed.ncbi.nlm.nih.gov/28917949/)]
31. Hawkes AL, Patrao TA, Green A, Aitken JF. CanPrevent: a telephone-delivered intervention to reduce multiple behavioural risk factors for colorectal cancer. *BMC Cancer* 2012 Nov 27;12:560 [FREE Full text] [doi: [10.1186/1471-2407-12-560](https://doi.org/10.1186/1471-2407-12-560)] [Medline: [23181756](https://pubmed.ncbi.nlm.nih.gov/23181756/)]
32. Appel LJ, Clark JM, Yeh HC, Wang NY, Coughlin JW, Daumit G, et al. Comparative effectiveness of weight-loss interventions in clinical practice. *New England Journal of Medicine* 2011;365(21):1959-1968. [doi: [10.1186/1471-2407-12-560](https://doi.org/10.1186/1471-2407-12-560)] [Medline: [23181756](https://pubmed.ncbi.nlm.nih.gov/23181756/)]
33. Wing RR, Tate DF, Espeland MA, Lewis CE, LaRose JG, Gorin AA, Study of Novel Approaches to Weight Gain Prevention (SNAP) Research Group. Innovative Self-Regulation Strategies to Reduce Weight Gain in Young Adults: The Study of Novel Approaches to Weight Gain Prevention (SNAP) Randomized Clinical Trial. *JAMA Intern Med* 2016 Dec 01;176(6):755-762 [FREE Full text] [doi: [10.1001/jamainternmed.2016.1236](https://doi.org/10.1001/jamainternmed.2016.1236)] [Medline: [27136493](https://pubmed.ncbi.nlm.nih.gov/27136493/)]
34. Cohen AJ, Richardson CR, Heisler M, Sen A, Murphy EC, Hesterman OB, et al. Increasing Use of a Healthy Food Incentive: A Waiting Room Intervention Among Low-Income Patients. *Am J Prev Med* 2017 Feb;52(2):154-162 [FREE Full text] [doi: [10.1016/j.amepre.2016.11.008](https://doi.org/10.1016/j.amepre.2016.11.008)] [Medline: [28109458](https://pubmed.ncbi.nlm.nih.gov/28109458/)]
35. Goswami I, Urminsky O. The dynamic effect of incentives on postreward task engagement. *J Exp Psychol Gen* 2017 Jan;146(1):1-19. [doi: [10.1037/xge0000206](https://doi.org/10.1037/xge0000206)] [Medline: [28054810](https://pubmed.ncbi.nlm.nih.gov/28054810/)]
36. Spring B, Schneider K, McFadden HG, Vaughn J, Kozak AT, Smith M, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. *Arch Intern Med* 2012 May 28;172(10):789-796 [FREE Full text] [doi: [10.1001/archinternmed.2012.1044](https://doi.org/10.1001/archinternmed.2012.1044)] [Medline: [22636824](https://pubmed.ncbi.nlm.nih.gov/22636824/)]
37. Pellegrini CA, Steglitz J, Johnston W, Warnick J, Adams T, McFadden HG, et al. Design and protocol of a randomized multiple behavior change trial: Make Better Choices 2 (MBC2). *Contemp Clin Trials* 2015 Mar;41:85-92 [FREE Full text] [doi: [10.1016/j.cct.2015.01.009](https://doi.org/10.1016/j.cct.2015.01.009)] [Medline: [25625810](https://pubmed.ncbi.nlm.nih.gov/25625810/)]

38. Schulz DN, Kremers SPJ, Vandelanotte C, van AMJG, Schneider F, Candel MJJM, 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](https://doi.org/10.2196/jmir.3094)] [Medline: [24472854](https://pubmed.ncbi.nlm.nih.gov/24472854/)]
39. Spring B, Pagoto S, Pingitore R, Doran N, Schneider K, Hedeker D. Randomized controlled trial for behavioral smoking and weight control treatment: effect of concurrent versus sequential intervention. *J Consult Clin Psychol* 2004 Oct;72(5):785-796 [FREE Full text] [doi: [10.1037/0022-006X.72.5.785](https://doi.org/10.1037/0022-006X.72.5.785)] [Medline: [15482037](https://pubmed.ncbi.nlm.nih.gov/15482037/)]
40. Bush T, Lovejoy J, Javitz H, Torres A, Wassum K, Tan M, et al. Simultaneous vs. sequential treatment for smoking and weight management in tobacco quitlines: 6 and 12 month outcomes from a randomized trial. *BMC Public Health* 2018:S159 (in press).
41. US Department of Agriculture. All about the vegetable group - What counts as a cup of vegetables. 2016. URL: <http://www.choosemyplate.gov/vegetables> [accessed 2018-05-16] [WebCite Cache ID 6zSZPQd2k]
42. Kennedy J. DCU-based Shimmer signs research deal with Harvard's Wyss Institute. 2016. URL: <https://www.siliconrepublic.com/start-ups/shimmer-harvard-wyss-research> [accessed 2018-05-16] [WebCite Cache ID 6zSZwsdIO]
43. Davis, Eshelman, McKay. The relaxation and stress reduction workbook (4th Edition). New York, NY: New Harbinger Press; 1995.
44. Lieffers JRL, Hanning RM. Dietary assessment and self-monitoring with nutrition applications for mobile devices. *Can J Diet Pract Res* 2012;73(3):e253-e260. [Medline: [22958633](https://pubmed.ncbi.nlm.nih.gov/22958633/)]
45. Long JD, Littlefield LA, Estep G, Martin H, Rogers TJ, Boswell C, et al. Evidence review of technology and dietary assessment. *Worldviews Evid Based Nurs* 2010 Dec;7(4):191-204. [doi: [10.1111/j.1741-6787.2009.00173.x](https://doi.org/10.1111/j.1741-6787.2009.00173.x)] [Medline: [19804586](https://pubmed.ncbi.nlm.nih.gov/19804586/)]
46. Bardus M, Smith JR, Samaha L, Abraham C. Mobile Phone and Web 2.0 Technologies for Weight Management: A Systematic Scoping Review. *J Med Internet Res* 2015;17(11):e259 [FREE Full text] [doi: [10.2196/jmir.5129](https://doi.org/10.2196/jmir.5129)] [Medline: [26573984](https://pubmed.ncbi.nlm.nih.gov/26573984/)]
47. Lichstein K, Riedel B, Grieve R. Fair tests of clinical trials: A treatment implementation model. *Adv Behav Res Ther* 1994;16:1-29.
48. Bhattarai N, Prevost AT, Wright AJ, Charlton J, Rudisill C, Gulliford MC. Effectiveness of interventions to promote healthy diet in primary care: systematic review and meta-analysis of randomised controlled trials. *BMC Public Health* 2013;13:1203 [FREE Full text] [doi: [10.1186/1471-2458-13-1203](https://doi.org/10.1186/1471-2458-13-1203)] [Medline: [24355095](https://pubmed.ncbi.nlm.nih.gov/24355095/)]
49. King AC, Castro CM, Buman MP, Hekler EB, Urizar GG, Ahn DK. Behavioral impacts of sequentially versus simultaneously delivered dietary plus physical activity interventions: the CALM trial. *Ann Behav Med* 2013 Oct;46(2):157-168 [FREE Full text] [doi: [10.1007/s12160-013-9501-y](https://doi.org/10.1007/s12160-013-9501-y)] [Medline: [23609341](https://pubmed.ncbi.nlm.nih.gov/23609341/)]
50. Spittaels H, De BI, Vandelanotte C. Evaluation of a website-delivered computer-tailored intervention for increasing physical activity in the general population. *Prev Med* 2007 Mar;44(3):209-217. [doi: [10.1016/j.ypmed.2006.11.010](https://doi.org/10.1016/j.ypmed.2006.11.010)] [Medline: [17197015](https://pubmed.ncbi.nlm.nih.gov/17197015/)]
51. Vandelanotte C, De BI, Sallis JF, Spittaels H, Brug J. Efficacy of sequential or simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. *Ann Behav Med* 2005 Apr;29(2):138-146. [doi: [10.1207/s15324796abm2902_8](https://doi.org/10.1207/s15324796abm2902_8)] [Medline: [15823787](https://pubmed.ncbi.nlm.nih.gov/15823787/)]
52. 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](https://doi.org/10.1016/j.amepre.2007.05.007)] [Medline: [17888860](https://pubmed.ncbi.nlm.nih.gov/17888860/)]
53. Prochaska JJ, Prochaska JO. A Review of Multiple Health Behavior Change Interventions for Primary Prevention. *American Journal of Lifestyle Medicine* 2011 Jan 07;5(3):208-221. [doi: [10.1177/1559827610391883](https://doi.org/10.1177/1559827610391883)]
54. Hyman DJ, Pavlik VN, Taylor WC, Goodrick GK, Moya L. Simultaneous vs sequential counseling for multiple behavior change. *Arch Intern Med* 2007 Jun 11;167(11):1152-1158. [doi: [10.1001/archinte.167.11.1152](https://doi.org/10.1001/archinte.167.11.1152)] [Medline: [17563023](https://pubmed.ncbi.nlm.nih.gov/17563023/)]
55. Vandelanotte C, De BI, Brug J. Two-year follow-up of sequential and simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. *Ann Behav Med* 2007 Apr;33(2):213-219. [doi: [10.1080/08836610701310086](https://doi.org/10.1080/08836610701310086)] [Medline: [17447874](https://pubmed.ncbi.nlm.nih.gov/17447874/)]
56. Joseph AM, Willenbring ML, Nugent SM, Nelson DB. A randomized trial of concurrent versus delayed smoking intervention for patients in alcohol dependence treatment. *J Stud Alcohol* 2004 Nov;65(6):681-691. [Medline: [15700504](https://pubmed.ncbi.nlm.nih.gov/15700504/)]
57. 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](https://doi.org/10.1001/jama.2016.12858)] [Medline: [27654602](https://pubmed.ncbi.nlm.nih.gov/27654602/)]
58. Thosar SS, Niederhausen M, Lapidus J, Fino NF, Cigarroa J, Minnier J, et al. Self-regulated use of a wearable activity sensor is not associated with improvements in physical activity, cardiometabolic risk or subjective health status. *Br J Sports Med* 2017 Dec 08. [doi: [10.1136/bjsports-2017-098512](https://doi.org/10.1136/bjsports-2017-098512)] [Medline: [29222201](https://pubmed.ncbi.nlm.nih.gov/29222201/)]
59. Royer H, Stehr M, Sydnor J. Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. *American Economic Journal: Applied Economics* 2015 Jul;7(3):51-84 [FREE Full text] [doi: [10.1257/app.20130327.60](https://doi.org/10.1257/app.20130327.60)]

60. Schneider KL, Coons MJ, McFadden HG, Pellegrini CA, DeMott A, Siddique J, et al. Mechanisms of Change in Diet and Activity in the Make Better Choices 1 Trial. *Health Psychol* 2016 Apr 07;35(7):723-732 [FREE Full text] [doi: [10.1037/hea0000333](https://doi.org/10.1037/hea0000333)] [Medline: [27054299](https://pubmed.ncbi.nlm.nih.gov/27054299/)]
61. Pampel FC, Rogers RG. Socioeconomic status, smoking, and health: a test of competing theories of cumulative advantage. *J Health Soc Behav* 2004 Sep;45(3):306-321. [doi: [10.1177/002214650404500305](https://doi.org/10.1177/002214650404500305)] [Medline: [15595509](https://pubmed.ncbi.nlm.nih.gov/15595509/)]
62. Stringhini S, Dugravot A, Kivimaki M, Shipley M, Zins M, Goldberg M, et al. Do different measures of early life socioeconomic circumstances predict adult mortality? Evidence from the British Whitehall II and French GAZEL studies. *J Epidemiol Community Health* 2011 Dec;65(12):1097-1103 [FREE Full text] [doi: [10.1136/jech.2009.102376](https://doi.org/10.1136/jech.2009.102376)] [Medline: [20675701](https://pubmed.ncbi.nlm.nih.gov/20675701/)]
63. Cooney NL, Litt MD, Sevarino KA, Levy L, Kranitz LS, Sackler H, et al. Concurrent alcohol and tobacco treatment: Effect on daily process measures of alcohol relapse risk. *J Consult Clin Psychol* 2015 Apr;83(2):346-358. [doi: [10.1037/a0038633](https://doi.org/10.1037/a0038633)] [Medline: [25622198](https://pubmed.ncbi.nlm.nih.gov/25622198/)]
64. Kahler CW, Metrik J, LaChance HR, Ramsey SE, Abrams DB, Monti PM, et al. Addressing heavy drinking in smoking cessation treatment: a randomized clinical trial. *J Consult Clin Psychol* 2008 Oct;76(5):852-862 [FREE Full text] [doi: [10.1037/a0012717](https://doi.org/10.1037/a0012717)] [Medline: [18837602](https://pubmed.ncbi.nlm.nih.gov/18837602/)]
65. Bush T, Lovejoy J, Javitz H, Torres A, Wassum K, Magnussen B. Implementation of a combined smoking cessation weight management intervention in quitlines. 2016 Presented at: Society for Research on Nicotine Tobacco Annual Meeting; 2016; Chicago, IL.
66. Farley AC, Hajek P, Lycett D, Aveyard P. Interventions for preventing weight gain after smoking cessation. *Cochrane Database Syst Rev* 2012 Jan 18;1:CD006219. [doi: [10.1002/14651858.CD006219.pub3](https://doi.org/10.1002/14651858.CD006219.pub3)] [Medline: [22258966](https://pubmed.ncbi.nlm.nih.gov/22258966/)]
67. Hall SM, Tunstall CD, Vila KL, Duffy J. Weight gain prevention and smoking cessation: cautionary findings. *Am J Public Health* 1992 Jun;82(6):799-803. [Medline: [1585959](https://pubmed.ncbi.nlm.nih.gov/1585959/)]
68. Spring B, Duncan JM, Janke EA, Kozak AT, McFadden HG, DeMott A, et al. Integrating technology into standard weight loss treatment: a randomized controlled trial. *JAMA Intern Med* 2013 Jan 28;173(2):105-111 [FREE Full text] [doi: [10.1001/jamainternmed.2013.1221](https://doi.org/10.1001/jamainternmed.2013.1221)] [Medline: [23229890](https://pubmed.ncbi.nlm.nih.gov/23229890/)]
69. Pellegrini CA, Verba SD, Otto AD, Helsel DL, Davis KK, Jakicic JM. The comparison of a technology-based system and an in-person behavioral weight loss intervention. *Obesity (Silver Spring)* 2012 Feb;20(2):356-363 [FREE Full text] [doi: [10.1038/oby.2011.13](https://doi.org/10.1038/oby.2011.13)] [Medline: [21311506](https://pubmed.ncbi.nlm.nih.gov/21311506/)]

Abbreviations

ANOVA: analysis of variance
MBC1: Make Better Choices 1
MBC2: Make Better Choices 2
MVPA: moderate to vigorous physical activity
SUDS: Subjective Units of Distress Scale

Edited by G Eysenbach; submitted 03.04.18; peer-reviewed by A Staiano, O Giggins; comments to author 23.04.18; revised version received 07.05.18; accepted 08.05.18; published 19.06.18.

Please cite as:

Spring B, Pellegrini C, McFadden HG, Pfammatter AF, Stump TK, Siddique J, King AC, Hedeker D
 Multicomponent mHealth Intervention for Large, Sustained Change in Multiple Diet and Activity Risk Behaviors: The Make Better Choices 2 Randomized Controlled Trial
J Med Internet Res 2018;20(6):e10528
 URL: <http://www.jmir.org/2018/6/e10528/>
 doi: [10.2196/10528](https://doi.org/10.2196/10528)
 PMID: [29921561](https://pubmed.ncbi.nlm.nih.gov/29921561/)

©Bonnie Spring, Christine Pellegrini, H G McFadden, Angela Fidler Pfammatter, Tammy K Stump, Juned Siddique, Abby C King, Donald Hedeker. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 19.06.2018. 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.

Original Paper

What People “Like”: Analysis of Social Media Strategies Used by Food Industry Brands, Lifestyle Brands, and Health Promotion Organizations on Facebook and Instagram

Karen Michelle Klassen¹, RD, PhD; Emily S Borleis¹, BSc; Linda Brennan², PhD; Mike Reid², PhD; Tracy A McCaffrey¹, PhD; Megan SC Lim³, PhD

¹Monash University, Notting Hill, Australia

²RMIT University, Melbourne, Australia

³Burnet Institute, Melbourne, Australia

Corresponding Author:

Tracy A McCaffrey, PhD

Monash University

264 Ferntree Gully Road

Level 1

Notting Hill, 3168

Australia

Phone: 61 399024270

Email: tracy.mccaffrey@monash.edu

Abstract

Background: Health campaigns have struggled to gain traction with young adults using social media, even though more than 80% of young adults are using social media at least once per day. Many food industry and lifestyle brands have been successful in achieving high levels of user engagement and promoting their messages; therefore, there may be lessons to be learned by examining the successful strategies commercial brands employ.

Objective: This study aims to identify and quantify social media strategies used by the food industry and lifestyle brands, and health promotion organizations across the social networking sites Facebook and Instagram.

Methods: The six most engaging posts from the 10 most popular food industry and lifestyle brands and six health promotion organizations were included in this study. A coding framework was developed to categorize social media strategies, and engagement metrics were collected. Exploratory linear regression models were used to examine associations between strategies used and interactions on Facebook and Instagram.

Results: Posts from Facebook (143/227, 63.0%) and Instagram (84/227, 37.0%) were included. Photos (64%) and videos (34%) were used to enhance most posts. Different strategies were most effective for Facebook and Instagram. Strategies associated with higher Facebook interactions included links to purchasable items ($\beta=0.81$, 95% CI 0.50 to 1.13, $P<.001$) featuring body image messages compared with food content ($\beta=1.96$, 95% CI 1.29 to 2.64, $P<.001$), and where the content induced positive emotions ($\beta=0.31$, 95% CI 0.04 to 0.57, $P=.02$). Facebook interactions were negatively associated with using pop culture ($\beta=-0.67$, 95% CI -0.99 to -0.34 , $P<.001$), storytelling ($\beta=-0.86$, 95% CI -1.29 to -0.43 , $P<.001$) or visually appealing graphics ($\beta=-0.53$, 95% CI -0.78 to -0.28 , $P<.001$) in their posts compared with other strategies. Posting relatable content was negatively associated with interactions on Facebook ($\beta=-0.29$, 95% CI -0.53 to -0.06 , $P=.01$), but positively associated on Instagram ($\beta=0.50$, 95% CI 0.05 to 0.95, $P=.03$). Instagram interactions were negatively associated with weight loss ($\beta=-1.45$, 95% CI -2.69 to -0.21 , $P=.02$) and other content ($\beta=-0.81$, 95% CI -1.57 to -0.06 , $P=.04$) compared with food content.

Conclusions: Health promotion professionals and organizations can improve engagement using positive messaging and tailoring posts appropriate for different social media channels.

(*J Med Internet Res* 2018;20(6):e10227) doi:[10.2196/10227](https://doi.org/10.2196/10227)

KEYWORDS

nutrition; social media; Facebook; Instagram; health promotion

Introduction

Background

Social media is used almost ubiquitously, especially by young adults, with Facebook being the most common platform [1,2]. More than 80% of Australian young adults are accessing social media platforms at least once per day, particularly during key times of day associated with choosing foods, such as first thing in the morning and at lunch [3]. Since November 2007, Facebook has allowed brands and companies to create profiles [4], and since then, brands have been successful in using the features offered by social media to communicate with the public. Users are not only willingly engaging with brands, but also disseminating brand content to their circle of friends, thereby increasing brand reach [4].

The food industry has been very successful in marketing via social media [5] often winning prizes for marketing innovation [6]. Social media feeds can include both posts from brands individuals have chosen to follow, as well as advertisements. Food industry brands particularly target adolescents and young adults on social media and being exposed to these advertisements appears to influence their attitudes and intentions [7] and may influence their behavior as a result [8]. Young adults have a higher intake of sugar-sweetened beverages and are more likely to consume fast foods compared with other age groups [9]. It is possible that constant exposure to messages from food and beverage brands via social media increases or reinforces these unhealthy eating patterns.

Social media is also being widely used for health promotion [10-12] by health professionals, researchers, government, and non-governmental organizations. Identifying a target audience and tailoring a suitable health message remains important for health promotion design for both traditional and social media campaigns [13]. Benefits of using social media for health promotion are increased reach and interaction [10]; despite this, health promotion campaigns run via social media have traditionally struggled to reach and engage with large numbers of people [14].

In previous research, we found that similar marketing strategies were being used by alcohol brands and health promotion agencies to engage users. Such strategies included posting visually attractive content and linking posts to consumption cultures [14]. Other alcohol and sexual health research also found that consistency of posting and interaction between brands and users are associated with increased success [15,16].

Social media personalities, or “Influencers”, have more recently been identified as being a strategic and powerful avenue for product promotion [17]. Social media Influencers can be defined as individuals or groups of individuals who can shape attitudes and behaviors through online channels [18]. What makes these Influencers so successful appears to be their capacity to engage with users and develop a level of trust [17,19]. According to Nielsen’s 2012 report “Global Trust in Advertising and Brand Messages”, 92% of survey participants stated that they would trust word-of-mouth recommendations above any other form of advertising. Social media Influencers have overtaken

traditional celebrities in their ability to influence purchasing behavior, as users find them more credible and relatable [20]. However, traditional celebrity figures still appear to have a strong influence on lifestyle behaviors and some are even seen as “experts” in these areas [21]. Companies are also using Influencers and celebrities to enhance their brand and take inspiration from strategies used by social media Influencers in their own campaign materials [17,22,23].

Research examining social media strategies used in food and nutrition-related communication is currently lacking. No previous studies have compared user engagement on Facebook and Instagram for food brands, lifestyle brands, and health promotion organizations. In identifying the most successful strategies regarding engagement of users with a post, it will be possible to make recommendations for the improvement of nutrition-related health promotion using social media. A glossary of terms has been provided in [Multimedia Appendix 1](#).

The aim of this study was to analyze the content of popular posts made by food brands, lifestyle brands and health promotion organizations using social media (Facebook and Instagram) and to identify strategies associated with engagement.

Methods

Design

This study was a retrospective content analysis that used a mixed-methods exploratory design [24] to analyze content from public posts made on Facebook and Instagram by food industry brands, lifestyle brands and health promotion organizations and examined associations between content strategies and engagement.

Inclusion criteria

Social media pages or profiles (brands have “pages” on Facebook and “profiles” on Instagram; for consistency, herein referred to as “pages”) were selected for inclusion if they posted nutrition or food content and were active at the time of data collection. The top food industry brands, lifestyle brands (defined as any individual or non-food industry brand that creates content on social media that includes food or nutrition information and can include Influencers who have created a brand) and health promotion organizations in Australia were determined by the number of Australian fans (when a user follows a page, the user becomes a fan of the page) on Facebook. Brands could be global or Australian as long as they had a social media presence in Australia. As this was an exploratory study, 10 lifestyle and food industry brands and only six health promotion organizations were included due to the small number of organizations running nutrition-related health promotion campaigns on social media. This number was chosen to provide sufficient data to undertake exploratory analyses. The top 10 food industry and top 10 lifestyle brands were identified through Socialbakers [25]. For food industry brands, the following filters were selected: Facebook, brands/Fast-moving consumer goods (FMCG) Food/All FMCG Food, Australia (total fans). For lifestyle brands, the filters included: Facebook, celebrities, Australia (total fans); Facebook, entertainment, online show,

Australia (total fans); and Facebook, community, lifestyle, Australia (total fans).

The top health promotion organizations (n=6) were identified using the filters: Facebook, Society/All Society, Australia (total fans). Additionally, an online search of Australian health promotion organizations was conducted, and all partners of the Communicating Health project and organizations were considered for inclusion if they ran a nutrition and health-related campaign on social media in the past five years.

Data Collection

Data were collected for lifestyle and food industry brands from Facebook and Instagram between August and September 2017. Socialbakers Suite [25] monitors and collates data on the activity of millions of brands on social media and reports statistics by country and brand. Socialbakers provided the top six most engaging posts for both Facebook and Instagram during a 30-day period; therefore, these top six posts for each platform were used for all included pages. Some brands did not have an Australian Instagram page; therefore, the number of Instagram posts evaluated was less than on Facebook. If less than six posts were available during the 30-day period, data from all available posts were included.

Data were collected for health promotion organizations from Facebook and Instagram by searching Facebook and Instagram using Google Chrome web browser through searching for the selected campaign hashtag; sorting posts by “top posts” and “organization name” and choosing the top six posts. The most well-known and recent campaign for each health promotion organization was identified through online and literature searches.

Coding framework

A coding framework identified strategies used in posts by examining the qualitative data collected from the post content (including text, videos, and photographs). The framework was constructed by combining both deductive and inductive strategies (Multimedia Appendix 2). The deductive category development was based on prior research [14], everyday knowledge and logic. The inductive category development was informed by the content of the included posts themselves using open coding, a technique from grounded theory [26].

The coding process was iterative and continued to develop with previous posts being revisited throughout. Coding was done by three researchers: (1) KK, a nutrition professional and research fellow, (2) EB a nutrition science undergraduate student, and (3) ACY, a medical undergraduate student). Any differences in coding between researchers were discussed and a final decision was made by KK.

Engagement metrics

Quantitative data collected included social media engagement. Facebook engagement was measured in the following way: (1) reactions: when a user expresses their reaction to a post by clicking either “like,” “love,” “haha,” “wow,” “sad,” or “angry;”

(2) comments: when a user leaves a comment or replies to the post; and (3) shares: when a user “shares” (also referred to as tagging) or reposts the post [27-29]. Interactions are the sum of the number of comments, shares and all reactions. Instagram engagement was measured as follows: (1) likes: when a user clicks “like”, which suggests that a post has resonated with a user in some way [30], and (2) comments. Total Facebook interactions per post were calculated by summing of the number of comments, shares and all reactions (like, love, haha, wow, sad, and angry). Total Instagram interactions per post were calculated by summing the number of comments and likes.

Statistical Analysis

Categories developed from the coding framework were transformed into quantitative categorical data. Descriptive statistics provide characteristics for each organization and coding categories. Each coding category was evaluated based on frequency within and between the categories, and similar categories were combined for analysis.

Statistical models constructed were exploratory and inductive due to the paucity of similar research in this area; therefore, a backward, stepwise approach was used. Engagement metric data were positively skewed, therefore log-transformed for analysis. Univariable linear regression models were constructed to determine the variables to be included in the final multivariable regression models. Variables with a *P* value of <.200 in the univariable linear regression models were considered for inclusion in the multivariable models. All categories identified in the coding framework were considered for inclusion in the models.

Multivariable linear regression models were constructed to explore associations of the post content analysis with the engagement measures (ie, the dependent variables): Facebook and Instagram interactions. Models were tested for heteroscedasticity and normality of residuals, and extreme outliers were removed from the models.

Statistical analyses were conducted using Stata (Version 12, College Station, TX, USA).

Ethics

Ethics approval was received by Monash University Human Research Ethics committee (project 11945). Data presented are anonymized to protect the identity of brands or organizations included in the study.

Results

A total of 227 posts from health promotion organizations (34/227, 15.0%), food industry brands (79/227, 34.8%) and lifestyle brands (114/227, 50.2%) were analyzed from Facebook and Instagram. Regarding engagement metrics, included health promotion organizations had fewer fans on both Facebook and Instagram than food industry brands and lifestyle brands (Table 1). Health promotion organizations had less of a presence on Instagram, with fewer fans than on Facebook (Table 1).

Table 1. Brand engagement metrics.

Engagement metric	Health promotion organizations	Food industry brands	Lifestyle brands
Facebook pages included, n	6	10	10
Instagram profiles included, n	0	7	10
Facebook fans, median (25 th ; 75 th percentiles)	21,784 (9,896; 56,939)	1,033,517 (804,210; 32,039,808)	1,590,354 (1,142,469; 10,625,219)
Facebook fans in Australia, median (25 th ; 75 th percentiles)	20,119 (9,896; 51,155)	807,185 (647,250; 908,238)	365,111 (128,804; 531,601)
Facebook page likes, median (25 th ; 75 th percentiles)	22,034 (9,926; 59,309)	12,999,467 (829,335; 32,046,105)	1,622,860 (1,152,053; 10,912,143)
Instagram fans, median (25 th ; 75 th percentiles)	4,842 (2,732; 7,836) ^a	87,917 (36,001; 146,825)	328,509 (25,453; 2,140,075)
Facebook posts included in analysis, n	34	51	58
Instagram posts included in analysis, n	0	28	56
Total posts included in analysis, n	34	79	114
Interactions per Facebook post, median (25 th ; 75 th percentiles)	41 (30; 96)	2,484 (377; 6,219)	3,766 (1,205; 33,825)
Interactions per Instagram post, median (25 th ; 75 th percentiles)	— ^b	493 (267; 1,417)	8,530 (115; 53, 708)

^aData available for n=4 organizations only.

^bData not available.

Lifestyle brands had higher engagement with posts on both Facebook and Instagram when compared with both food industry and health promotion organizations (Table 1). Posts on Facebook had more engagement with a median (25th; 75thpercentiles) of 1,763 interactions (165; 7,374) than those on Instagram with 1,582 interactions (211; 18,414).

The proportion of posts using different engagement strategies is shown in Multimedia Appendix 3. For all categories except relationship building and format, the strategies used were diverse across the different organizations. Most posts used photographs (145/227, 64%), with only a few using only text (5/227, 2%). Health promotion organizations used more ‘prompting engagement’ strategies, links to health information, featured fruits, vegetables, and grains, had a more serious tone, used hashtags, had more real-world tie-ins than the other organizations and were the only organization type to present statistics or facts in their posts. Lifestyle brands and health promotion organizations induced more positive emotions than food industry; food industry and lifestyle brands had more links to purchasable items; food industry had the highest product promotion and did not feature people in most of their posts; lifestyle brands were the only group to talk about body image and weight loss and had the most interactions per 1000 fans for both Facebook and Instagram.

Two multivariable regression models were constructed to explore the associations between social media strategies and engagement (measured by interactions) on Facebook and

Instagram (Table 2). Facebook interactions were positively associated with lifestyle brands compared with health promotion organizations, including links to purchasable items (beta=0.81, 95% CI 0.50 to 1.13, $P<.001$), featuring body image messages compared with food content (beta=1.96, 95% CI 1.29 to 2.64, $P<.001$), posting videos compared with photos (beta=0.33, 95% CI 0.11 to 0.54, $P=.004$) and where the content induced positive emotions (beta=0.31, 95% CI 0.04 to 0.57, $P=.02$).

Facebook interactions were negatively associated with using pop culture (beta=−0.67, 95% CI −0.99 to −0.34, $P<.001$), story-telling (beta=−0.86, 95% CI −1.29 to −0.43, $P<.001$) or visually appealing graphics (beta=−0.53, 95% CI −0.78 to −0.28, $P<.001$) in their posts compared with other strategies, featuring weight loss compared with food content (beta=−1.06, 95% CI −1.76 to −0.37, $P=.003$), featuring people (beta=−0.42, 95% CI −0.71 to −0.13, $P=.005$), including links to health information (beta=−0.47, 95% CI −0.83 to −0.10, $P=.01$), posting relatable content (beta=−0.29, 95% CI −0.53 to −0.06, $P=.01$) and paying to promote posts (beta=−0.30, 95% CI −0.56 to −0.04, $P=.03$).

Instagram interactions were positively associated with including links to purchasable items (beta=1.32, 95% CI 0.77 to 1.88, $P<.001$) and posting relatable content (beta=0.50, 95% CI 0.05 to 0.95, $P=.03$).

Instagram interactions were negatively associated with weight loss (beta=−1.45, 95% CI −2.69 to −0.21, $P=.02$) and other content (beta=−0.81, 95% CI −1.57 to −0.06, $P=.04$) compared with food content, and with using hashtags.

Table 2. Multivariable Linear Regression Models of Facebook and Instagram interactions.

Variables in model	Facebook interactions, log (10)		Instagram interactions, log (10)	
	Standardized beta (95% CI)	P value	Standardized beta (95% CI)	P value
Organization type				
Health promotion organization	Ref ^a	N/A ^b	N/A	N/A
Food industry	0.45 (–0.15 to 1.05)	.14	Ref	N/A
Lifestyle brands	1.42 (0.96 to 1.88)	<.001	0.30 (–0.14 to 0.75)	.18
Strategies used				
Other strategies	Ref	N/A	N/A	N/A
Pop culture	–0.67 (–0.99 to –0.34)	<.001	N/A	N/A
Story-telling	–0.86 (–1.29 to –0.43)	<.001	N/A	N/A
Visually appealing	–0.53 (–0.78 to –0.28)	<.001	N/A	N/A
Links to purchasable items	0.81 (0.50 to 1.13)	<.001	1.32 (0.77 to 1.88)	<.001
Post content				
Food content	Ref	N/A	Ref	N/A
Body image content	1.96 (1.29 to 2.64)	<.001	–0.30 (–1.21 to 0.62)	.52
Weight loss content	–1.06 (–1.76 to –0.37)	.003	–1.45 (–2.69 to –0.21)	.02
Other content	–0.15 (–0.46 to 0.16)	.34	–0.81 (–1.57 to –0.06)	.04
Posts that featured people	–0.42 (–0.71 to –0.13)	.005	N/A	N/A
Links to health information	–0.47 (–0.83 to –0.10)	.01	0.29 (–0.28 to 0.87)	.31
Post format				
Photo	Ref	N/A	N/A	N/A
Video	0.33 (0.11 to 0.54)	.004	N/A	N/A
Text	–0.04 (–0.57 to 0.49)	.88	N/A	N/A
Relatable content	–0.29 (–0.53 to –0.06)	.01	0.50 (0.05 to 0.95)	.03
Positive emotion	0.31 (0.04 to 0.57)	.02	0.47 (–0.10 to 1.04)	.11
Promoted post	–0.30 (–0.56 to –0.04)	.03	N/A	N/A
Uses hashtags	–0.25 (–0.50 to 0.00)	.05	–0.55 (–0.91 to –0.19)	.003
Optimistic tone of post	–0.22 (–0.49 to 0.05)	.11	N/A	N/A
Number of posts included in model	141	N/A	84	N/A
Adjusted R ²	78.2%	N/A	60.2%	N/A

^aRef: reference category for multivariable linear regression.^bN/A: not applicable.

Discussion

Principal Findings

This is the first study to identify and quantify the social media strategies utilized by food industry brands, lifestyle brands and health promotion organizations across Facebook and Instagram, to our knowledge. Each organization type used different social media strategies to engage users. The food industry brands attempted to induce appetite and encourage users to eat; health promotion organizations frequently provided statistics and facts and used a more serious tone, while lifestyle brands were positive and relatable. Health promotion organizations were not chosen based on the 10 most popular organizations but were

limited to those organizations with lifestyle-related campaigns. Therefore, while health promotion organizations had substantially fewer followers and post interactions, direct comparisons between the number of fans of health promotion organizations and other brands are not applicable.

Links to purchasable items were used by both food and lifestyle brands and were consistently associated with more interactions on both Facebook and Instagram. This is not surprising considering consumers who are fans of such pages are often seeking new products or versions or products and may be engaged in online purchasing or exploration of available offers. Schultz et al [31] found that promotional posts were negatively associated with Facebook post likes, but positively associated

with Facebook shares. However, we did not do separate analyses for each interaction type. Many people hate being exposed to advertisements and try to block it from their social media feeds [32], while others do not notice the advertisements to which they are exposed [33]. Food industry brands included in this study had the highest number of Facebook fans but the lowest level of engagement per post per 1000 fans compared with other organization types. Although fewer Australians claim to be following brand pages on social media than in previous years, those who were following brands said they were doing so to receive discounts (54%) and to receive free items acquired from giveaways (48%) [3]. Of the food industry posts analyzed in the current study, 35% were advertising discounts or giveaways, which is similar to other findings [4]. These results suggest that users are engaging with food industry brands for their own gain, financial or otherwise.

Positive emotion-inducing strategies were associated with more interactions on Facebook and Instagram and using an optimistic tone was associated with more interactions on Facebook. Emotion plays a role in the attention and attraction the user experiences towards a post [34]. Participants who experience positive emotions when viewing a post on social media are far more likely to engage with that post than those who do not experience positive emotions [35].

Models for the two social media channels explored, Facebook and Instagram, included different strategies that were statistically significant. For example, posts classified as “relatable” were negatively associated with Facebook interactions but positively associated with Instagram interactions. Posts were classified as “relatable” if they encouraged feelings of friendship between the poster and fan or if the post contained content that is “relatable” to the user. Examples of this included providing practical advice that would apply to their audience or talking about issues they think are important or interesting to their fans (eg, debating the pros and cons of Hawaiian pizza, or what to eat for breakfast). Each social media channel has different features, is used differently by users and therefore strategies should be tailored for each channel [36]. Since its development in 2010, Instagram has become one of the most popular photo sharing applications worldwide [37] and particularly since the introduction of Instagram Stories and Instagram Live allows for immediate engagement with users. Instagram facilitates parasocial interactions, where imaginary social relationships and interpersonal interactions between the lifestyle personality and the social media user occur [38]. These relationships and interactions can be developed by using some of the strategies observed in the current study: relatable content, use of personal stories, and positive emotion and tone. This form of interaction may help to explain high levels of engagement seen on such posts; users are developing connections to the personality and may treat them similarly, to how they would treat a friend on social media.

Other differences between strategies associated with the different social media channels examined included post format (video versus photographs) and including body image content. Videos and body image content were statistically significantly positively associated with Facebook interactions, but not associated with Instagram interactions. These results emphasize the importance

of tailoring messages to suit both the social media platform used and the desired outcome.

How Can Health Promotion Organizations Enhance Their Social Media Strategies?

Health promotion organizations had less of an overall presence on Instagram. Heldman et al [39] discuss how despite many organizations developing a social media presence, they often lack in the more social aspects. Rather than adapting to this online form of communication, organizations appear to be continuing with many of the strategies used in traditional health promotion [40]. What is evident from our analyses is that different organization types used different strategies in their posts. The use of facts/statistics and less frequent use of an optimistic tone and using real-world tie-ins by health promotion organizations were the biggest differences between the groups. The health promotion organizations analyzed in the current study developed posts that were more serious in tone and often relied on statistics and facts to communicate their intended message. The information provided in the posts was important and frequently linked to further information; however, these posts had minimal engagement from fans. Taken together the engagement metrics and strategies presented in Table 1 and Multimedia Appendix 3 illustrate how building relationships with user/fans/followers are advantageous for lifestyle brands. As health promotion organizations, we should be cognizant of this approach to engage our target audience.

In the past, many health promotion organizations have developed campaigns employing fear as a way to induce behavior change [41]. For some aspects of health promotion (eg, anti-smoking campaigns) there may be room for both approaches, but it remains to be elucidated whether or not this is suitable in organizations communicating food and nutrition messages as this strategy was not used by other organizations in this study.

The coding framework developed in this study can be used as a guide by health organizations who are planning social media campaigns to target young adults.

Limitations

Limitations of this study included the short time frame of data collected. Evaluating a longer period could identify seasonal differences and improve strategies for creating messages for holidays or special events, and we plan to continue to develop the coding framework and monitor future posts to address this limitation. We included posts that did not contain food or health-related information which limited our ability to examine strategies that are particularly effective for food-related messages. We chose to include posts that were not directly about food or nutrition as long as they were from a brand (or profile) that posted about food or nutrition. The diversity of content from these brands, particularly the lifestyle brands demonstrates the importance of relationship building and indicates that people are getting food and nutrition-related information from pages posting diverse content. Furthermore, although interactions provide good measures of user engagement, the level to which users take on and use this information cannot be determined from these metrics alone. Although traditional social media engagement statistics (reactions, comments, shares, and

interactions) indicate the number of users that interact with social media posts, they do not indicate any resulting behavior change nor those who view, process and interact with the content "offline" (eg, the lurkers [42]), nor do they indicate behavior change as a result of interacting with the posts [7,43]. Further research could analyze the content of the comments on posts to gauge the quality of interaction between the page owner and follower and to gain insight into users' intention to behave.

Conclusions

This unique, exploratory study examined "real-life" social media posts with a sample size sufficient to create a coding framework and to create exploratory models.

Social media content should be tailored to suit not only the target audience but also the social media channel being used and the desired engagement. Health promotion practitioners and organizations can learn from other types of brands and consider using few statistics and more positive content to relay healthy eating messages.

Acknowledgments

Thanks to Anchiun Yeoh for assistance with coding. Communicating Health is funded by the National Health and Medical Research Council (Grant number: GNT1115496) with additional investigators Catherine Lombard, Helen Truby, and Claire Palermo.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Glossary of terms.

[PDF File (Adobe PDF File), 40KB - [jmir_v20i6e10227_app1.pdf](#)]

Multimedia Appendix 2

Coding framework.

[PDF File (Adobe PDF File), 643KB - [jmir_v20i6e10227_app2.pdf](#)]

Multimedia Appendix 3

Coding framework characteristics or strategies used for social media posts.

[PDF File (Adobe PDF File), 57KB - [jmir_v20i6e10227_app3.pdf](#)]

References

1. Greenwood S, Perrin A, Duggan M. Social Media Update. 2016. 2016 URL: <http://www.pewinternet.org/2016/11/11/social-media-update-2016/> [accessed 2018-01-18] [WebCite Cache ID 6wZWRzQdT]
2. Sensis. Sensis Social Media Report. 2016. 2016 URL: https://www.sensis.com.au/asset/PDFdirectory/Sensis_Social_Media_Report_2016 [accessed 2018-04-06] [WebCite Cache ID 6ySxFS58N]
3. Sensis. Sensis Social Media Report. 2017. 2017 URL: <https://www.sensis.com.au/asset/PDFdirectory/Sensis-Social-Media-Report-2017.pdf> [accessed 2018-04-06] [WebCite Cache ID 6ySxcTNA9]
4. Freeman B, Kelly B, Baur L, Chapman K, Chapman S, Gill T, et al. Digital junk: food and beverage marketing on Facebook. *Am J Public Health* 2014 Dec;104(12):e56-e64. [doi: [10.2105/AJPH.2014.302167](https://doi.org/10.2105/AJPH.2014.302167)] [Medline: [25322294](https://pubmed.ncbi.nlm.nih.gov/25322294/)]
5. Freeman B, Kelly B, Vandevijvere S, Baur L. Young adults: beloved by food and drink marketers and forgotten by public health? *Health Promot Int* 2016 Dec;31(4):954-961. [doi: [10.1093/heapro/dav081](https://doi.org/10.1093/heapro/dav081)] [Medline: [26276799](https://pubmed.ncbi.nlm.nih.gov/26276799/)]
6. Colchester J, Hargreaves T. WARC Trends. KFC: Dirty Louisiana ? Don't make dirty good, make clean bad URL: https://www.warc.com/content/article/warc-awards/kfc_dirty_louisiana_%E2%80%93_dont_make_dirty_good,_make_clean_bad/110994 [WebCite Cache ID 6ySxtrnxy]
7. Buchanan L, Kelly B, Yeatman H. Exposure to digital marketing enhances young adults' interest in energy drinks: An exploratory investigation. *PLoS One* 2017;12(2):e0171226 [FREE Full text] [doi: [10.1371/journal.pone.0171226](https://doi.org/10.1371/journal.pone.0171226)] [Medline: [28152016](https://pubmed.ncbi.nlm.nih.gov/28152016/)]
8. Fishbein M. The role of theory in HIV prevention. *AIDS Care* 2000 Jun;12(3):273-278. [doi: [10.1080/09540120050042918](https://doi.org/10.1080/09540120050042918)] [Medline: [10928203](https://pubmed.ncbi.nlm.nih.gov/10928203/)]
9. Australian Bureau of Statistics. Australian health survey: First results Australia 2014-2015. 2015. URL: <http://www.abs.gov.au/ausstats/abs@.nsf/mf/4364.0.55.001> [WebCite Cache ID 6yRRMwQvR]

10. 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](https://doi.org/10.2196/jmir.1933)] [Medline: [23615206](https://pubmed.ncbi.nlm.nih.gov/23615206/)]
11. Levac J, O'Sullivan T. Social Media and its Use in Health Promotion. *Revue interdisciplinaire des sciences de la santé - Interdisciplinary Journal of Health Sciences* 2010 Feb 10;1(1):47 [FREE Full text] [doi: [10.18192/riss-ijhs.v1i1.1534](https://doi.org/10.18192/riss-ijhs.v1i1.1534)]
12. Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The Role of Social Network Technologies in Online Health Promotion: A Narrative Review of Theoretical and Empirical Factors Influencing Intervention Effectiveness. *J Med Internet Res* 2015;17(6):e141 [FREE Full text] [doi: [10.2196/jmir.3662](https://doi.org/10.2196/jmir.3662)] [Medline: [26068087](https://pubmed.ncbi.nlm.nih.gov/26068087/)]
13. 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](https://doi.org/10.1177/1524839911405850)] [Medline: [21558472](https://pubmed.ncbi.nlm.nih.gov/21558472/)]
14. Lim MS, Hare JD, Carrotte ER, Dietze PM. An investigation of strategies used in alcohol brand marketing and alcohol-related health promotion on Facebook. *DIGITAL HEALTH* 2016 Apr 29;2:205520761664730. [doi: [10.1177/2055207616647305](https://doi.org/10.1177/2055207616647305)]
15. Burton S, Dadich A, Soboleva A. Competing Voices: Marketing and Counter-Marketing Alcohol on Twitter. *Journal of Nonprofit & Public Sector Marketing* 2013 Apr;25(2):186-209. [doi: [10.1080/10495142.2013.787836](https://doi.org/10.1080/10495142.2013.787836)]
16. Veale HJ, Sacks-Davis R, Weaver ER, Pedrana AE, Stoové MA, Hellard ME. The use of social networking platforms for sexual health promotion: identifying key strategies for successful user engagement. *BMC Public Health* 2015;15(1):85 [FREE Full text] [doi: [10.1186/s12889-015-1396-z](https://doi.org/10.1186/s12889-015-1396-z)] [Medline: [25884461](https://pubmed.ncbi.nlm.nih.gov/25884461/)]
17. Uzunoğlu E, Misci Kip S. Brand communication through digital influencers: Leveraging blogger engagement. *International Journal of Information Management* 2014 Oct;34(5):592-602. [doi: [10.1016/j.ijinfomgt.2014.04.007](https://doi.org/10.1016/j.ijinfomgt.2014.04.007)]
18. Freberg K, Graham K, McGaughey K, Freberg L. Who are the social media influencers? A study of public perceptions of personality. *Public Relations Review* 2011 Mar;37(1):90-92 [FREE Full text] [doi: [10.1016/j.pubrev.2010.11.001](https://doi.org/10.1016/j.pubrev.2010.11.001)] [Medline: [25904163](https://pubmed.ncbi.nlm.nih.gov/25904163/)]
19. Liu S, Jiang C, Lin Z, Ding Y, Duan R, Xu Z. Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach. *Information Sciences* 2015 Jun;306:34-52 [FREE Full text] [doi: [10.1016/j.ins.2015.01.034](https://doi.org/10.1016/j.ins.2015.01.034)]
20. Djafarova E, Rushworth C. Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior* 2017 Mar;68:1-7. [doi: [10.1016/j.chb.2016.11.009](https://doi.org/10.1016/j.chb.2016.11.009)]
21. Hoffman SJ, Tan C. Biological, psychological and social processes that explain celebrities' influence on patients' health-related behaviors. *Arch Public Health* 2015;73(1):3 [FREE Full text] [doi: [10.1186/2049-3258-73-3](https://doi.org/10.1186/2049-3258-73-3)] [Medline: [25973193](https://pubmed.ncbi.nlm.nih.gov/25973193/)]
22. Chapman S. Does celebrity involvement in public health campaigns deliver long term benefit? Yes. *BMJ* 2012 Sep 25;345:e6364. [Medline: [23015036](https://pubmed.ncbi.nlm.nih.gov/23015036/)]
23. Rayner G. Does celebrity involvement in public health campaigns deliver long term benefit? No. *BMJ* 2012 Sep 25;345:e6362. [Medline: [23015035](https://pubmed.ncbi.nlm.nih.gov/23015035/)]
24. Creswell J. *Research design: Qualitative, quantitative and mixed methods approaches*. Thousand Oaks, California: Sage Publications; 2003:0761924426.
25. Socialbakers. Statistics. 2017 URL: <https://www.socialbakers.com/> [accessed 2018-01-18] [WebCite Cache ID 6wZW5sPqI]
26. Glaser BG, Strauss AL. *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine; 1967.
27. Facebook BRC. Reactions. 2016 URL: <https://en.facebookbrand.com/assets/reactions> [WebCite Cache ID 6wZecN86W]
28. Simply Measured. Facebook Comment. URL: <http://simplymeasured.com/definition/facebook-comment/#sm.0000k04rm7dpsdfbw9c1ec85xha7f> [accessed 2018-01-18] [WebCite Cache ID 6wZezldXF]
29. Simply Measured. Facebook Share. URL: <http://simplymeasured.com/definition/facebook-share/#sm.0000k04rm7dpsdfbw9c1ec85xha7f> [accessed 2018-01-18] [WebCite Cache ID 6wZevjpKb]
30. Simply Measured. Instagram Like. URL: <https://simplymeasured.com/definition/instagram-like/#sm.0000k04rm7dpsdfbw9c1ec85xha7f> [accessed 2018-02-26] [WebCite Cache ID 6xVTsmZTM]
31. Schultz CD. Proposing to your fans: Which brand post characteristics drive consumer engagement activities on social media brand pages? *Electronic Commerce Research and Applications* 2017 Nov;26:23-34. [doi: [10.1016/j.elerap.2017.09.005](https://doi.org/10.1016/j.elerap.2017.09.005)]
32. Crozier M. Engaging young people online. 2016. URL: <http://www.bangthetable.com/engaging-young-people/> [accessed 2018-01-19] [WebCite Cache ID 6wZgoIJMz]
33. Media Dynamics Incorporated. Adults Spend Almost 10 Hours Per Day With The Media, But Note Only 150 Ads. 2014. URL: <https://www.mediadynamicsinc.com/uploads/files/PR092214-Note-only-150-Ads-2mk.pdf> [accessed 2018-01-19] [WebCite Cache ID 6wZgKDNlv]
34. Stieglitz S, Dang-Xuan L. Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems* 2013 Jun 7;29(4):217-248. [doi: [10.2753/MIS0742-1222290408](https://doi.org/10.2753/MIS0742-1222290408)]
35. Smith S. Conceptualising and Evaluating Experiences with Brands on Facebook. *International Journal of Market Research* 2013 May;55(3):357-374. [doi: [10.2501/IJMR-2013-034](https://doi.org/10.2501/IJMR-2013-034)]
36. Coelho RLF, Oliveira DSD, Almeida MISD. Does social media matter for post typology? Impact of post content on Facebook and Instagram metrics. *Online Information Review* 2016 Aug 08;40(4):458-471. [doi: [10.1108/oir-06-2015-0176](https://doi.org/10.1108/oir-06-2015-0176)]
37. Lee E, Lee J, Moon JH, Sung Y. Pictures Speak Louder than Words: Motivations for Using Instagram. *Cyberpsychol Behav Soc Netw* 2015 Sep;18(9):552-556. [doi: [10.1089/cyber.2015.0157](https://doi.org/10.1089/cyber.2015.0157)] [Medline: [26348817](https://pubmed.ncbi.nlm.nih.gov/26348817/)]

38. Ward J. A Content Analysis of Celebrity Instagram Posts and Parasocial Interaction. *Elon Journal of Undergraduate Research in Communications* 2016;7(1):1 [FREE Full text]
39. Heldman AB, Schindelar J, Weaver JB. Social Media Engagement and Public Health Communication: Implications for Public Health Organizations Being Truly “Social”. *Public Health Rev* 2013;35(1):13. [doi: [10.1007/BF03391698](https://doi.org/10.1007/BF03391698)]
40. Neiger BL, Thackeray R, Van Wagensen SA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. *Health Promot Pract* 2012 Mar;13(2):159-164. [doi: [10.1177/1524839911433467](https://doi.org/10.1177/1524839911433467)] [Medline: [22382491](https://pubmed.ncbi.nlm.nih.gov/22382491/)]
41. Soames JRF. Effective and ineffective use of fear in health promotion campaigns. *Am J Public Health* 1988 Feb;78(2):163-167. [Medline: [3276236](https://pubmed.ncbi.nlm.nih.gov/3276236/)]
42. Merchant G, Weibel N, Patrick K, Fowler JH, Norman GJ, Gupta A, et al. Click “like” to change your behavior: a mixed methods study of college students' exposure to and engagement with Facebook content designed for weight loss. *J Med Internet Res* 2014 Jun 24;16(6):e158 [FREE Full text] [doi: [10.2196/jmir.3267](https://doi.org/10.2196/jmir.3267)] [Medline: [24964294](https://pubmed.ncbi.nlm.nih.gov/24964294/)]
43. Freeman B, Potente S, Rock V, McIver J. Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers? *Public Health Res Pract* 2015 Mar 30;25(2):e2521517 [FREE Full text] [doi: [10.17061/phrp2521517](https://doi.org/10.17061/phrp2521517)] [Medline: [25848735](https://pubmed.ncbi.nlm.nih.gov/25848735/)]

Abbreviations

AFG: Australian Food Guide

AGtHE: Australian Guide to Healthy Eating

FMCG: Fast-moving consumer goods

Edited by G Eysenbach; submitted 27.02.18; peer-reviewed by P Charney, N Allen; comments to author 28.03.18; revised version received 23.04.18; accepted 24.04.18; published 14.06.18.

Please cite as:

Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MSC

What People “Like”: Analysis of Social Media Strategies Used by Food Industry Brands, Lifestyle Brands, and Health Promotion Organizations on Facebook and Instagram

J Med Internet Res 2018;20(6):e10227

URL: <http://www.jmir.org/2018/6/e10227/>

doi: [10.2196/10227](https://doi.org/10.2196/10227)

PMID: [29903694](https://pubmed.ncbi.nlm.nih.gov/29903694/)

©Karen Michelle Klassen, Emily S Borleis, Linda Brennan, Mike Reid, Tracy A McCaffrey, Megan SC Lim. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 14.06.2018. 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.

Original Paper

Social Drinking on Social Media: Content Analysis of the Social Aspects of Alcohol-Related Posts on Facebook and Instagram

Hanneke Hendriks¹, PhD; Bas Van den Putte^{1,2}, PhD; Winifred A Gebhardt³, PhD; Megan A Moreno⁴, PhD

¹Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

²Trimbos Institute, Netherlands Institute for Mental Health and Addiction, Utrecht, Netherlands

³Health, Medical and Neuropsychology Unit, Institute of Psychology, Leiden University, Leiden, Netherlands

⁴Department of Pediatrics, University of Wisconsin Madison, Madison, WI, United States

Corresponding Author:

Hanneke Hendriks, PhD

Amsterdam School of Communication Research

University of Amsterdam

Nieuwe Achtergracht 166

Amsterdam,

Netherlands

Phone: 31 64 710 9985

Email: h.hendriks@uva.nl

Abstract

Background: Alcohol is often consumed in social contexts. An emerging social context in which alcohol is becoming increasingly apparent is social media. More and more young people display alcohol-related posts on social networking sites such as Facebook and Instagram.

Objective: Considering the importance of the social aspects of alcohol consumption and social media use, this study investigated the social content of alcohol posts (ie, the evaluative social context and presence of people) and social processes (ie, the posting of and reactions to posts) involved with alcohol posts on social networking sites.

Methods: Participants (N=192; mean age 20.64, SD 4.68 years, 132 women and 54 men) gave researchers access to their Facebook and/or Instagram profiles, and an extensive content analysis of these profiles was conducted. Coders were trained and then coded all screenshotted timelines in terms of evaluative social context, presence of people, and reactions to post.

Results: Alcohol posts of youth frequently depict alcohol in a positive social context (425/438, 97.0%) and display people holding drinks (277/412, 67.2%). In addition, alcohol posts were more often placed on participants' timelines by others (tagging; 238/439, 54.2%) than posted by participants themselves (201/439, 45.8%). Furthermore, it was revealed that such social posts received more likes (mean 35.50, SD 26.39) and comments than nonsocial posts (no people visible; mean 10.34, SD 13.19, $P<.001$).

Conclusions: In terms of content and processes, alcohol posts on social media are social in nature and a part of young people's everyday social lives. Interventions aiming to decrease alcohol posts should therefore focus on the broad social context of individuals in which posting about alcohol takes place. Potential intervention strategies could involve making young people aware that when they post about social gatherings in which alcohol is visible and tag others, it may have unintended negative consequences and should be avoided.

(*J Med Internet Res* 2018;20(6):e226) doi:[10.2196/jmir.9355](https://doi.org/10.2196/jmir.9355)

KEYWORDS

social media; alcohol drinking; social interaction

Introduction

Alcohol and Youth

Alcohol consumption and binge drinking among adolescents and young adults have been consistently linked with negative

consequences such as accidents, harassment, vandalism, sexual abuse, impaired brain development, and disease [1-3]. Alcohol abuse is therefore regarded as a major cause of preventable death and morbidity [4]. Many young people, however, consume alcohol frequently and often excessively. Recent studies in

Europe and the United States showed that 35% of youth who consume alcohol had at least 1 binge drinking episode (ie, drinking 5 or more drinks on 1 occasion [5]) in the past month [5-7]. This high prevalence of alcohol use and the multitude of negative consequences increase the need to gain more insight into the context in which this behavior takes place. Research has shown that alcohol is often consumed in social contexts [8]. An emerging social context in which alcohol is becoming increasingly apparent is that of social media. Young people increasingly display alcohol-related posts on social networking sites such as Facebook and Instagram [9]. This study investigated those alcohol posts on social media, and how social those posts actually were, by conducting a content analysis of alcohol posts on Instagram and Facebook among 192 adolescents and young adults.

Social Context of Alcohol

Young people often consume alcohol in social contexts such as dinners and parties, and alcohol consumption is often referred to as a social activity [8,10,11]. Research has shown that alcohol consumption plays a large part in young people's social lives and social identity exploration [12]. Furthermore, many studies have shown that social norms are strongly related to adolescent alcohol use, especially when these norms originate from peers [2]. The importance of the social environment for alcohol consumption is also apparent from the vast amount of research demonstrating the effects of peer influence and social networks on alcohol consumption [13-16]. Rosenquist et al [17], for example, showed that when close friends within a person's social network drink alcohol, this greatly increases the chance that the person drinks alcohol as well. Taken together, alcohol use should be seen as part of the social context in which it is so strongly embedded.

Alcohol and Social Media

Recent technological changes have drastically changed the way in which young people shape their social lives, and new social contexts have emerged as a consequence. Adolescents and young adults increasingly spend their time communicating with others in online settings [18,19]. Social networking sites such as Facebook and Instagram play an important role in young people's daily lives, accounting for a large portion of their time [20-22]. Research has shown that these social networking sites are frequently used by young people to communicate about alcohol, and multiple studies have demonstrated that many adolescents and young adults (percentages vary between 36% and 96%) display alcohol-related posts on Facebook (eg, photos in which young people hold alcoholic beverages [9,23-24]).

Alarmingly, studies have found positive associations between alcohol posts on social media and self-reported alcohol consumption. For example, Moreno et al [25] found that posting about alcohol use on Facebook was associated with increased offline drinking behavior. Similarly, Boyle et al [26] observed that exposure to alcohol posts on Facebook, Instagram, and Snapchat predicted more alcohol consumption 6 months later (see also Geusens and Beullens [27]). Thus, there is evidence suggesting a relation between both posting and being exposed to alcohol posts on social media and actual drinking behavior.

A few studies have examined the content of alcohol posts in more detail, usually by employing a clinical framework to evaluate individuals' alcohol posts. These studies have found that alcohol posts suggesting problematic alcohol use are more exceptional than posts showing regular use (ie, posts about alcohol but not about problem drinking or intoxication) but are more predictive of alcohol abuse [9,23-24,28-29]. However, these studies focus on the individual rather than the individual in a social context, and in-depth knowledge on the social aspects of these posts is lacking. Considering the importance of social aspects for both alcohol consumption and social media use, this study aims to provide insight into the social content of alcohol posts (ie, the evaluative social context and the presence of people) and the social processes (ie, the posting of and reactions to posts) involved with alcohol posts on social networking sites.

Social Content of Alcohol Posts

Alcohol consumption and social media use are inherently social [10]. Although some studies have emphasized the social nature of alcohol posts [12,30], few studies to date have focused on the exact social content displayed in these posts. Having more insight on the extent to which and how the content of alcohol posts are social can greatly increase our understanding of alcohol-related social media use and provide valuable information for intervention strategies (eg, whether an individualistic or social intervention strategy would work better). In this study, we examined 2 social aspects of alcohol-related content: the social evaluative context toward alcohol (ie, whether alcohol is portrayed in a negative, neutral, or positive context) and the presence of people in a post (eg, whether alcohol posts show a close-up of a cocktail, display 1 person drinking wine, or depict groups of people at a party drinking beers). With regard to the social evaluative context, research has shown that the valence of social interactions about alcohol can determine conversation effects [31,32]. For example, Hendriks et al [33] showed that when interpersonal interactions are positive toward alcohol, this can lead to increased alcohol consumption, and when conversations are mainly negative toward alcohol, this can reduce drinking behaviors. It is therefore important to illuminate the social evaluative context of alcohol-related content on social media. In this study, we investigated whether alcohol posts on social media show a negative social context (eg, someone looking disapprovingly at a drunk person) or positive social context (eg, people toasting and laughing), in line with research by Beullens and Schepers [23].

With regard to the presence of people in a post, it has been found that people are easily persuaded to engage in a specific behavior merely by observing other people portraying that behavior. Examples of such studies range from recent experiments showing that healthy social norms in health campaigns (eg, many people in a health ad displaying healthy conduct) can encourage the uptake of healthy behaviors [34-36] to classic studies by Asch [37] demonstrating the powerful impact of other people on participants' willingness to engage in similar behaviors. In addition, there are indications that the social conformity effects found by Asch [37] are a function of group size [38]. These findings are in line with the idea of a basic human need to belong—people have a strong need to fit in with a group [39]. Given this evidence, we investigated the extent

to which people are displayed in alcohol posts. We pose the following research questions about social content (RQ1):

RQ1a: Do alcohol posts reflect a negative or positive social evaluative context?

RQ1b: To what extent are people present in alcohol posts?

Social Processes Involved With Alcohol Posts

In addition to the social content of alcohol posts, it is also important to understand the social processes involved with the posting of alcohol-related content on social media because this can further enhance our understanding of the social nature of alcohol posts. Two factors related to social processes are explored in this study: how alcohol posts get posted and how people respond to these posts. First, it is unclear whether people actively post about alcohol themselves or whether this is part of a social process in which they are tagged (mentioned) in posts by others. Whether people actively post about alcohol or are passively tagged in alcohol posts is important because research has suggested that active (ie, talking) versus passive (ie, listening) interpersonal communication can lead to different effects of this communication. For example, Janis and King [40] asked people to either deliver a speech advocating for a certain issue or listen to that same speech and showed that people who talked were more persuaded by the speech than people who listened. Who posts alcohol posts is relevant to understand for practical reasons because knowing this can provide important information for future interventions aiming to decrease the posting of such content or its negative impact on the individual. If people are mostly tagged in posts by others this requires a different intervention strategy (by encouraging tagged people to ask to be removed or not included in the alcohol post) than if people post about alcohol themselves (by directly discouraging people to post about alcohol).

Second, it is not yet clear whether and how negatively or positively others respond to alcohol posts on social media. Do alcohol posts receive likes and are comments supportive of the posts? Research has revealed that approval of a behavior (ie, a supportive injunctive norm) encourages the behavior [34,36]. Likes (and supportive comments) on social media can illustrate such a social norm and are therefore important factors potentially determining a post's influence. This was suggested by Alhabash et al [41], who investigated the effects of alcohol posts and found that posts with many likes had especially strong persuasive effects. It is therefore important to provide insight into the reactions that alcohol posts trigger.

An additional relevant question addressed in this study is whether reactions to alcohol posts depend on the social content of the post. For example, do alcohol posts with people in them get more likes than posts with no people in them, and do posts with a positive social evaluative context receive more supportive comments than less positive social posts? This seems likely given the abovementioned studies on the importance of social norms and the need to belong [34,39]; alcohol posts showing people may portray alcohol-supportive norms, especially when people are holding alcoholic beverages, that could cause viewers to behave in line with their need for belongingness and approve

of such norms by giving supportive reactions. By liking social alcohol posts, people can express their feelings of friendship and need to belong to the people depicted in the pictures. We pose the following research questions about social processes related to alcohol posts (RQ2):

RQ2a: Who posts alcohol posts (do participants themselves post or are participants tagged)?

RQ2b: What are the responses of others to alcohol posts?

RQ2c: Do responses to alcohol posts depend on whether the content of the post is social?

In sum, this study aimed to provide insight into the social content of alcohol posts and the social processes involved with alcohol posts on social networking sites. This aim was addressed through a content analysis of alcohol posts on Facebook and Instagram.

Methods

Participants and Design

This study was part of a larger data collection during which 561 participants filled out a questionnaire regarding social media use (questionnaire data are not the focus of this study). For the purposes of this paper, participants were asked whether we could access (friend) their Instagram and Facebook profiles, and 214 of these participants gave their consent to do so. Of these participants, 22 were excluded from later analyses because they did not fall into the intended age category (12 to 30 years), resulting in a total of 192 participants to be analyzed (mean 20.64, SD 4.68 years, 132 women and 54 men). Due to technical problems, 6 participants could not be successfully linked to their questionnaire data. Their profiles could be coded, however, and are included in the analyses. Of the participants (N=192), 106 had only a Facebook profile, 15 had only an Instagram profile, and 70 had both.

Procedure

The participants who agreed to give access to their profiles were asked to accept a friend request from a research profile on Facebook and/or Instagram, allowing all participant posts to be accessed by the researchers. Participants were informed that all data would be stored anonymously; all personal information (ie, names and faces) would be removed from the posts so that these could not be traced back to the participants. Screenshots were made of all timelines on profiles for the period of the previous year (April 2015 to April 2016), after which friendships on Facebook and Instagram were cancelled and the research profile was deleted. For participants younger than 18 years, consent was required from the parents as well as the adolescents. This study was approved by the university's ethics committee.

Content Analysis

Coder Training

Two coders were trained in 3 sessions led by the first author, during which several example profiles were coded and inconsistencies were discussed. After the training was finished, 10% of the profiles were coded in order to assess coder

reliability, after which all profiles were coded. Coders' agreement was acceptable ($\kappa=.683-.912$). Please note that we could not calculate agreement for variables with a low n (eg, posts with a negative social evaluative context or Instagram posts).

Coding Procedure

A codebook described the coding process. Coders were asked to scroll down the timeline for the past year and to look at each post. Once an alcohol post was identified, several variables were recorded in Excel (Microsoft Corp). The profiles were coded by 1 of 2 coders. When coders were unsure how to code a post, they discussed it with the first author after which a choice for coding was made.

Coding Variables

Occurrence and Frequency

An alcohol post was defined as "a post about alcohol or in which alcohol is visible." Coders were asked to take the whole post (ie, the photo/video including headings/texts) into account. Coders coded whether there were any alcohol posts visible (occurrence: no/yes) on the profile, and if so, how many they identified on the profile (frequency). If no alcohol posts were present, no further information for that profile was needed. If an alcohol post was identified, however, the coder was asked to code the additional variables (see below). Most Facebook alcohol posts (410/442, 92.8%) consisted of a photo accompanied by a caption; 7.2% (32/442) consisted solely out of text.

Social Evaluative Context

Coders were asked to describe whether the context was negative, neutral, or positive toward alcohol. Social evaluative context was based on Beullens and Schepers [23]; in line with their coding book, we coded whether an alcohol post showed a negative context (eg, someone looking disapprovingly at a drunk person), a neutral context (eg, no explicit judgment or emotion is shown), or a positive context (eg, people laughing and toasting with alcoholic drinks). The context could also be inferred from a caption (eg, when a photo showed a close-up of a dinner table including a glass of wine with the caption "Having a lovely time!"). Text-only posts were coded in a similar manner: negative if alcohol was described with negative words ("drank too much; headache!"), positive if alcohol was described with positive words ("I'm looking forward to boozing tonight!").

People Present in Post

Given research [34-39] showing that the presence of people can affect persuasion, we coded whether people were present in the post (no people visible/tagged [text-only posts], participant only, others only, or participant with others). We also coded whether someone was holding an alcoholic beverage, and if so, who (no one, participant, others). The reason we measured the latter variable is because holding a beverage can be seen as a clear indication of descriptive norms (person is drinking) as well as injunctive norms (person must like drinking [42]). Figures 1 and 2 show examples of how a post would be coded in terms of evaluative context and presence of people.

Placer of Post

We coded whether the post was placed by the participant or by others (with participant tagged) based on previous studies suggesting different effects of active versus passive interpersonal communication [40].

Reactions to Posts

We coded how many likes and comments the post received and whether the comments were mainly negative ("Pathetic!" or "You're not looking too well"), neutral ("Where is this?" or "What are you drinking?"), or positive ("Nice shot!" or "Looking good!") toward the participants and/or post. Valence of comments was adjusted from Beullens and Schepers [23].

Data Analysis

To investigate the social content (RQ1a/RQ1b) and processes (RQ2a/RQ2b) involved with alcohol posts we first described the frequencies and descriptives of the coded variables. In order to address whether the likes and comments on alcohol posts depend on the content of the posts (RQ2c), we conducted analyses of variance with the social content of the post as independent variable (presence of people, holding of beverages, and placer of post) and likes, comments, and valence of comments as dependent variables.

Because n was too low, we did not investigate the influence of a negative (2 Facebook posts) versus positive (11 Facebook posts) context on reactions, and we decided not to analyze Instagram posts (posted by only 24 participants). We consider differences with $P<.01$ as significant in order to compensate for multiple comparisons [43].

Figure 1. Example of a post that would be coded as having a positive evaluative context, showing the participant with others, and showing the participant holding a drink.



Figure 2. Example of a post that would be coded as having a positive evaluative context and no people visible. The caption translates as "Lovely relaxing time with a book and drink!".



Results

Occurrence and Frequency of Alcohol Posts

Results showed that 51.1% (90/176) of participants with a Facebook profile had at least 1 alcohol post on their profile during the past recorded year. On Instagram, 28% (24/85) of participants had a profile with at least 1 alcohol post. The participants who had an alcohol post on their Facebook profile posted on average 5 alcohol posts during the past recorded year (mean 5.02, SD 6.32). On Instagram, among the participants who posted an alcohol post, the average number of alcohol posts was 4 (mean 3.91, SD 4.21).

The following section focuses solely on the alcohol-related posts by participants on Facebook ($n=90$) or on Instagram ($n=24$). Participants with Facebook or Instagram profiles who posted about alcohol (20 men, 73 women, mean age 21.92, SD 3.94 years) differed slightly from those who did not post about alcohol (34 men, 59 women, mean age 19.35, SD 5.01 years) in terms of age ($F_{1,184}=15.12$, $P<.001$) but not significantly in terms of gender ($\chi^2_1=5.1$, $P=.02$). In total, we analyzed 442 alcohol posts on Facebook and 94 alcohol posts on Instagram.

Please note that the total number of posts described sometimes varies—the total number of posts mentioned at social evaluative context is 438, whereas the total number of posts mentioned at placer of post is 439 because coders sometimes had to code a variable as missing (eg, when the context of the post could not be inferred or when the valence of comments could not be coded because there were none). The percentages shown are based on the total number of posts that were coded for that variable.

RQ1: Social Content

Social Evaluative Context

The majority of the alcohol posts on Facebook depicted alcohol in a positive context (425/438, 97.0%), a few showed a neutral context (11/438, 2.5%), and almost none depicted a negative context (2/438, 0.5%). On Instagram, these percentages were similar (see [Table 1](#)).

Presence of People

Most Facebook alcohol posts (277/412, 67.2%) displayed a social context with other people visible in the posts, but 16.3% (67/412) of alcohol posts did not show any people. Similar patterns were seen on Instagram ([Table 1](#)). Beverages were held by participants as opposed to appearing on a table. Thus, on both platforms, most alcohol posts displayed a social context with the participant and other people shown.

RQ2: Social Processes

Placer of Post

Alcohol posts were often placed by participants themselves on Facebook (201/439, 45.8%), but, more frequently, the posts were placed by others (238/439, 54.2%) with the participant tagged ([Table 2](#)). Tagging was not yet possible on Instagram in 2015.

Reactions to Alcohol Posts

On average, an alcohol post on Facebook received 29 likes (mean 29.67, SD 26.15) and 3 comments (mean 3.41, SD 5.36). Almost all alcohol posts on Facebook got at least 1 like (421/442, 95.9%), and most got at least 1 comment (309/442, 71.9%). This was similar on Instagram (see [Table 2](#)).

Although not all alcohol posts received comments; of those that did, the responses were mostly positive (243/297, 81.8%) toward the post and/or participant (neutral: 49/297, 16.5%; negative: 5/297, 1.7%). The pattern was similar on Instagram (see [Table 2](#)).

RQ2c: Do the Responses to Alcohol Posts Depend on the Content of the Post?

Presence of People

First, there was a significant effect of presence of people (no one vs only participant vs only others vs participant together with others) on the number of likes ($F_{3,406}=18.96$, $P<.001$). Posts without any people present in them yielded significantly fewer likes (mean 10.34, SD 13.19) than posts in which the participant was shown alone (mean 30.63, SD 21.21, $P=.002$), posts in which only others were displayed (mean 25.71, SD 28.98, $P=.001$), or posts in which others were shown together with the participant (mean 35.50, SD 26.39, $P<.001$). Posts in which others were shown together with the participant yielded more likes than posts with only others ($P=.01$). Posts without any people present resulted in significantly fewer comments (mean 2.22, SD 3.20) than posts displaying others without the participant (mean 4.47, SD 7.13, $P=.01$). Posts in which the participant was alone (mean 3.68, SD 5.25) did not differ from the other posts in terms of comments ($P>.22$). The valence of comments was not influenced by presence of people ($F_{3,267}=0.47$, $P=.70$). Thus, social posts yielded more responses and more positive reactions than posts without people in them.

Holding of Beverages

A significant effect of holding of beverages (no one versus participant versus others) on likes was revealed ($F_{2,322}=14.36$, $P<.001$). That is, posts without someone explicitly holding an alcoholic beverage resulted in significantly fewer likes (mean 19.20, SD 23.86) than posts in which the participant (mean 30.94, SD 22.97, $P=.001$) and/or others (mean 37.34, SD 29.94, $P<.001$) were holding drinks in their hands. There were no significant effects on the number or valence of comments, all $F<0.92$, all $P>.40$.

Placer of Post

A significant effect of placer of post (participant vs others) on likes was found ($F_{1,434}=8.50$, $P=.004$). That is, posts placed by others yielded significantly more likes (mean 33.02, SD 25.05) than posts placed by the participants themselves (mean 25.77, SD 26.80). There were no significant effects on the number or valence of comments (all $F<1.55$, all $P>.21$). Thus, in general, more social posts (eg, in which others are present, which are posted by others, and in which people hold drinks in their hands) resulted in more reactions, in particular more likes. For an overview of the effects of social content on likes, see [Table 3](#).

Table 1. Social content of alcohol-related posts on Facebook and Instagram.

Variable	Facebook (n=442), n (%)	Instagram (n=94), n (%)
Social evaluative context		
Negative	2 (0.5)	0 (0)
Neutral	11 (2.5)	0 (0)
Positive	425 (96.1)	94 (100)
People present		
No one	67 (16.3)	28 (29.8)
Only participant	19 (4.6)	10 (10.6)
Participant and others	277 (67.2)	49 (52.1)
Only others	49 (11.9)	7 (7.4)
Beverages in hand		
No one	116 (35.4)	25 (43.1)
Participant	98 (29.9)	19 (32.8)
Others	114 (34.8)	14 (21.4)

Table 2. Social processes involved with the posting of alcohol-related posts on Facebook and Instagram.

Variable	Facebook (n=442)	Instagram (n=94)
Placement of post, n (%)		
Participant	201 (45.8)	94 (100)
Others	238 (54.2)	—
Valence of post reactions, n (%)		
Negative comments	5 (1.7)	1 (2.3)
Neutral comments	49 (16.5)	6 (13.6)
Positive comments	243 (81.8)	37 (84.1)
Number of post reactions, mean (SD)		
Number of likes	29.67 (26.15)	43.12 (52.11)
Number of comments	3.41 (5.36)	3.82 (3.01)

Table 3. Significant differences in likes of alcohol-related posts on Facebook based on social content.

Variable	Number of likes on Facebook	
	Mean (SD)	Median
People present		
No one	10.34 (13.19)	5.00
Only participant	30.63 (21.21)	27.00
Participant with others	35.50 (26.39)	32.00
Only others	25.71 (28.98)	18.00
Beverages in hand		
No one	19.20 (23.86)	11.00
Participant	30.94 (22.97)	25.00
Others	37.34 (29.94)	34.00
Placement of post		
Participant	25.77 (26.80)	18.00
Others	33.02 (25.05)	28.50

Discussion

Principal Findings

The goal of this research was to investigate the social content (ie, evaluative social context and presence of people) and social processes (ie, posting of and reactions to posts) involved with alcohol posts on social networking sites. Two main findings were revealed: in terms of content and processes, alcohol posts on social media are social in nature and these social aspects are related to reactions—the more social elements to the post, the more (positive) reactions the post receives.

The first main finding that alcohol posts are social in terms of content was illustrated in 2 ways. First, alcohol posts showed alcohol in a positive social context, with people approving of the drinking behaviors pictured in the post (eg, by laughing or toasting). This is in line with Beullens and Schepers [23], who found that the context of alcohol consumption on social media is often positive. This approval by others of alcohol consumption in alcohol posts is alarming, as ample studies show that a positive injunctive social norm (the perception that others approve of this conduct) leads to increased drinking behaviors [35,36,42]. Alcohol posts may thus, by enhancing positive social norms, lead to more excessive alcohol consumption. This idea was also suggested in a recent study about cigars and cigarillo images on Instagram in which it was observed that these images are relatively common and could add to the normalization of tobacco in everyday life [44]. More research is needed to investigate this potential mediating effect of social media posts on norms and subsequent drinking behavior.

In addition, alcohol posts are social in the sense that most reflect a social setting (eg, dinners or parties) that display (groups of) people. These groups of people can potentially further increase the effect of the portrayed positive social norms by showing that many approve of drinking alcohol and display this behavior themselves (eg, by holding alcoholic beverages) thereby enhancing descriptive norms (ie, the perception that many others engage in this behavior [42]). This suggests a strong impact of alcohol posts, especially when many people appear in the image. Furthermore, it is possible that positive injunctive norms are especially triggered by a positive social context in alcohol posts and that positive descriptive norms are more strongly affected by whether people are holding these beverages in the picture. Given the importance of social norms for alcohol consumption, this role of social norms within the effects of alcohol posts should be investigated in more detail.

Not only are alcohol posts social in terms of content, they are also social in terms of the processes involved with posting them. That is, alcohol posts receive many likes and comments, and these comments are mostly positive about the post. This is consistent with our findings regarding the social evaluative context: people from the social network appear strong in their approval of alcohol posts. This is important to keep in mind when addressing the issue of alcohol posts; they seem to be posted in a very proalcohol context. Furthermore, this study is the first to reveal that tagging plays an important role in the posting of alcohol-related posts. Many participants had alcohol posts on their profile that they did not post; they had simply

been tagged in these posts. This tagging has very relevant implications for interventions aiming to decrease alcohol posts. An individualistic approach (eg, intervention planners asking people to not post an alcohol post on their profile) may not be sufficient. Thus, reducing these tagged posts may require a different strategy by taking into account the social environment (eg, intervention planners asking people to deny tagging, or by asking people not to tag others in such posts).

The second main finding is that social aspects of alcohol posts are related to the responses to these posts. Posts that displayed people, posts in which others were holding drinks, and posts posted by others yielded more likes and often more comments than posts without people in them, posts in which no one was holding a drink, or posts posted by the participants themselves. This further strengthens the idea that alcohol posts are part of a social process, in which the social aspects common to alcohol posts trigger social interactions that show further appreciation of the post and help spread the message even further. Because often-liked posts are featured higher in Facebook's newsfeed, this can further increase the chance the post will get liked, commented on, or shared.

An explanation for why social posts triggered more, and more positive, interpersonal communication could be that human beings have a strong need to connect with and belong to groups [39]. Many studies have shown the preference that humans have for social stimuli (eg, faces) over nonsocial stimuli (eg, geometrical shapes [45,46]). That social posts get liked more seems therefore to be in line with this fundamental human need for belongingness. The question, however, is how generic or specific this effect is. Do all posts with people in them or posted by others receive more likes and comments? Or is this only, or especially, the case with alcohol consumption, a very social behavior? As this study does not compare alcohol posts with posts regarding other unhealthy behaviors, this remains a question. Research exists, however, showing that young people also post about other health behaviors such as physical activity, snacking, smoking, marijuana use, waterpipe use, and sexual behaviors [44,47-51]. Although they do not always depict groups of people, some of these posts elicit a lot of responses when posted in like-minded fora, such as posts related to "fitspiration" (ie, a recent social media trend designed to motivate people to eat healthily and exercise [52]). Thus, it appears that social aspects of a post are not the only factor influencing post responses. Some recent research suggests that certain posts (such as certain types of active selfies [53]) trigger more responses by increasing narrative involvement. It is possible that the social setting often visible in alcohol posts may also increase involvement in the narrative (eg, being part of a dinner or party) and may therefore lead to more likes and comments. The processes through which social alcohol posts lead to more reactions and whether these processes are distinct for alcohol posts is an important avenue for future studies.

Taking these findings together suggests that alcohol posts on social media are a part of young people's everyday social lives, in which drinking at dinners or parties and posting and tagging about these social events go hand in hand. This normalcy of alcohol posts not only strengthens the idea in young people's minds that alcohol is normal and a part of daily life [54,55] but

also increases the idea that a lot of people are positive about alcohol and consume alcohol regularly. This can increase pluralistic ignorance (ie, people incorrectly assuming that many people engage in a specific behavior such as alcohol abuse [56]) and lead to more drinking behaviors (in line with Moreno et al [25] and Boyle et al [26]). Although a lot of research exists showing that alcohol use in traditional media (movies or commercials [57-59]) can increase alcohol use of viewers, recent evidence suggests that alcohol use on social media has even stronger effects because it is more closely linked to descriptive and injunctive norms that consequently leads to a stronger impact on drinking behaviors [60]. The role of alcohol posts on social media should not be underestimated and should be incorporated in interventions that aim to decrease excessive alcohol use.

Limitations and Future Research

Some limitations should be noted. A complication of interpreting the effects of social posts is that the audience size of posts may differ between social and nonsocial posts (more people are tagged in social posts, thereby increasing the number of people who see the post), potentially explaining part of the relationship between social posts and increased responses. However, the fact that not only the number of people but also whether someone was holding the alcoholic beverage influenced the number of likes suggests that something alcohol-specific may further increase the impact of social features on likes. To answer the questions whether and why social alcohol posts lead to more reactions with more certainty, future studies are necessary comparing social versus nonsocial alcohol posts and contrasting those with similar neutral posts and posts about other unhealthy behaviors (eg, smoking or snacking).

Another limitation lies in the fact that we coded the alcohol posts from an outsider viewpoint—coders looked at an alcohol post and decided whether it was social (if people were visible). However, another worthwhile method may be to ask participants to describe and interpret alcohol posts (in line with Hebden et al [61], for example) to see whether any differences arise between our coding and participant interpretations. For example,

it is possible that although the alcohol post in Figure 2 was coded as having no person visible, when participants see this post they may infer that the poster is probably not alone while enjoying this drink. Such interpretations are an interesting avenue for future research.

Conclusions and Potential Implications for Interventions

This study investigated the social content and social processes involved with alcohol posts on social networking sites. Findings revealed that in terms of content and processes, alcohol posts on social media are social in nature. Furthermore, these social aspects are related to responses to these posts: the more social elements in the post, the more (positive) reactions (eg, likes and comments). Taken together, these findings suggest that alcohol posts on social media are a part of young people's everyday social lives, and interventions aiming to decrease alcohol posts should include a focus on the individual in a social (networking) context.

Potential implications for interventions arise from this study. First, the comparison between the participants who did and did not post about alcohol showed that, as a group, more older young adult women post alcohol content on social media. This may illustrate an important target group for future interventions addressing the posting of alcohol-related posts. Second, our findings support viewing the posting of alcohol posts as a social behavior. Therefore, attempts to reduce this behavior should not take an individualistic approach but should focus on the individual in a social context. It may be worthwhile to make people aware that they post about social gatherings in which alcohol is visible and tag others in these posts and such posts may have unintended negative consequences and should be avoided by not posting these pictures or by hiding alcoholic beverages when a photo is taken. Adolescents and young adults can be motivated to not allow tags in such posts and to stimulate others to not post alcohol posts or tag them in it. Whether and which of these tactics will be successful is an important avenue for further research.

Acknowledgments

Data collection was facilitated by Leiden University. This work was supported by a Veni grant (451-15-022) from the Netherlands Organization for Scientific Research awarded to the first author (HH).

Conflicts of Interest

None declared.

References

1. Hughes K, Anderson Z, Morleo M, Bellis MA. Alcohol, nightlife and violence: the relative contributions of drinking before and during nights out to negative health and criminal justice outcomes. *Addiction* 2008 Jan;103(1):60-65. [doi: [10.1111/j.1360-0443.2007.02030.x](https://doi.org/10.1111/j.1360-0443.2007.02030.x)] [Medline: [17996008](https://pubmed.ncbi.nlm.nih.gov/17996008/)]
2. Perkins HW. Surveying the damage: a review of research on consequences of alcohol misuse in college populations. *J Stud Alcohol Suppl* 2002 Mar(14):91-100. [Medline: [12022733](https://pubmed.ncbi.nlm.nih.gov/12022733/)]
3. Rehm J, Room R, Graham K, Monteiro M, Gmel G, Semplos CT. The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease: an overview. *Addiction* 2003 Sep;98(9):1209-1228. [Medline: [12930209](https://pubmed.ncbi.nlm.nih.gov/12930209/)]

4. Cherpitel CJ. Alcohol and injuries: a review of international emergency room studies since 1995. *Drug Alcohol Rev* 2007 Mar;26(2):201-214. [doi: [10.1080/09595230601146686](https://doi.org/10.1080/09595230601146686)] [Medline: [17364856](https://pubmed.ncbi.nlm.nih.gov/17364856/)]
5. Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula MK, Marks JS. Binge drinking among US adults. *JAMA* 2003 Jan 1;289(1):70-75. [Medline: [12503979](https://pubmed.ncbi.nlm.nih.gov/12503979/)]
6. Bräker AB, Soellner R. Alcohol drinking cultures of European adolescents. *Eur J Public Health* 2016 Aug;26(4):581-586. [doi: [10.1093/eurpub/ckw033](https://doi.org/10.1093/eurpub/ckw033)] [Medline: [27037330](https://pubmed.ncbi.nlm.nih.gov/27037330/)]
7. EMCDDA-ESPAD Joint Publication. ESPAD Report 2015: results from the European School Survey Project on Alcohol and Other Drugs. Luxembourg: Publications Office of the European Union; 2015. URL: http://www.espad.org/sites/espad.org/files/ESPAD_report_2015.pdf [accessed 2018-05-30] [WebCite Cache ID 6zmvQBf1G]
8. Beck KH, Arria AM, Caldeira KM, Vincent KB, O'Grady KE, Wish ED. Social context of drinking and alcohol problems among college students. *Am J Health Behav* 2008;32(4):420-430 [FREE Full text] [doi: [10.5555/ajhb.2008.32.4.420](https://doi.org/10.5555/ajhb.2008.32.4.420)] [Medline: [18092902](https://pubmed.ncbi.nlm.nih.gov/18092902/)]
9. Moreno MA, Christakis DA, Egan KG, Brockman LN, Becker T. Associations between displayed alcohol references on Facebook and problem drinking among college students. *Arch Pediatr Adolesc Med* 2012 Feb;166(2):157-163 [FREE Full text] [doi: [10.1001/archpediatrics.2011.180](https://doi.org/10.1001/archpediatrics.2011.180)] [Medline: [21969360](https://pubmed.ncbi.nlm.nih.gov/21969360/)]
10. Christiansen M, Vik PW, Jarchow A. College student heavy drinking in social contexts versus alone. *Addict Behav* 2002;27(3):393-404. [Medline: [12118627](https://pubmed.ncbi.nlm.nih.gov/12118627/)]
11. Pavis S, Cunningham-Burley S, Amos A. Alcohol consumption and young people: exploring meaning and social context. *Health Educ Res* 1997;12:311-322 [FREE Full text]
12. Ridout B, Campbell A, Ellis L. "Off your Face(book)": alcohol in online social identity construction and its relation to problem drinking in university students. *Drug Alcohol Rev* 2012 Jan;31(1):20-26. [doi: [10.1111/j.1465-3362.2010.00277.x](https://doi.org/10.1111/j.1465-3362.2010.00277.x)] [Medline: [21355935](https://pubmed.ncbi.nlm.nih.gov/21355935/)]
13. Bot SM, Engels RCME, Knibbe RA, Meeus WHJ. Friend's drinking behaviour and adolescent alcohol consumption: the moderating role of friendship characteristics. *Addict Behav* 2005 Jun;30(5):929-947. [doi: [10.1016/j.addbeh.2004.09.012](https://doi.org/10.1016/j.addbeh.2004.09.012)] [Medline: [15893090](https://pubmed.ncbi.nlm.nih.gov/15893090/)]
14. Goodwin L. Explaining alcohol consumption and related experiences among fraternity and sorority members. *J College Student Developm* Sep 1989;30(5):448-458.
15. Maxwell KA. Friends: the role of peer influence across adolescent risk behaviors. *J Youth Adolesc* Aug 2002;31(4):267-277. [doi: [10.1023/A:1015493316865](https://doi.org/10.1023/A:1015493316865)]
16. Urberg KA, Luo Q, Pilgrim C, Degirmencioglu SM. A two-stage model of peer influence in adolescent substance use: individual and relationship-specific differences in susceptibility to influence. *Addict Behav* 2003 Sep;28(7):1243-1256. [Medline: [12915166](https://pubmed.ncbi.nlm.nih.gov/12915166/)]
17. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. *Ann Intern Med* 2010 Apr 06;152(7):426-433 [FREE Full text] [doi: [10.7326/0003-4819-152-7-201004060-00007](https://doi.org/10.7326/0003-4819-152-7-201004060-00007)] [Medline: [20368648](https://pubmed.ncbi.nlm.nih.gov/20368648/)]
18. Baym N, Zhang Y, Lin MC. Social interactions across media: interpersonal communication on the internet, telephone and face-to-face. *New Media Soc* 2004 Jun 30;6(3):299-318. [doi: [10.1177/1461444804041438](https://doi.org/10.1177/1461444804041438)]
19. Moreno MA, D'Angelo J, Whitehill J. Social media and alcohol: summary of research, intervention ideas and future study directions. *Media Commun* Jun 2016;4(3):50-59. [doi: [10.17645/mac.v4i3.529](https://doi.org/10.17645/mac.v4i3.529)]
20. Muise A, Christofides E, Desmarais S. More information than you ever wanted: does Facebook bring out the green-eyed monster of jealousy? *Cyberpsychol Behav* 2009 Aug;12(4):441-444. [doi: [10.1089/cpb.2008.0263](https://doi.org/10.1089/cpb.2008.0263)] [Medline: [19366318](https://pubmed.ncbi.nlm.nih.gov/19366318/)]
21. Tong ST, Van der Heide B, Langwell L, Walther JB. Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook. *J Comput Mediat Commun* Apr 2008;13(3):531-549. [doi: [10.1111/j.1083-6101.2008.00409.x](https://doi.org/10.1111/j.1083-6101.2008.00409.x)]
22. Carotte ER, Prichard I, Lim MSC. "Fitspiration" on social media: a content analysis of gendered images. *J Med Internet Res* 2017 Mar 29;19(3):e95 [FREE Full text] [doi: [10.2196/jmir.6368](https://doi.org/10.2196/jmir.6368)] [Medline: [28356239](https://pubmed.ncbi.nlm.nih.gov/28356239/)]
23. Beullens K, Schepers A. Display of alcohol use on Facebook: a content analysis. *Cyberpsychol Behav Soc Netw* 2013 Jul;16(7):497-503. [doi: [10.1089/cyber.2013.0044](https://doi.org/10.1089/cyber.2013.0044)] [Medline: [23617225](https://pubmed.ncbi.nlm.nih.gov/23617225/)]
24. van Hoof JJ, Bekkers J, van Vuuren M. Son, you're smoking on Facebook! College students' disclosures on social networking sites as indicators of real-life risk behaviors. *Comput Hum Behav* 2014 May;34:249-257. [doi: [10.1016/j.chb.2014.02.008](https://doi.org/10.1016/j.chb.2014.02.008)]
25. Moreno MA, Cox ED, Young HN, Haaland W. Underage college students' alcohol displays on Facebook and real-time alcohol behaviors. *J Adolesc Health* 2015 Jun;56(6):646-651 [FREE Full text] [doi: [10.1016/j.jadohealth.2015.02.020](https://doi.org/10.1016/j.jadohealth.2015.02.020)] [Medline: [26003580](https://pubmed.ncbi.nlm.nih.gov/26003580/)]
26. Boyle SC, LaBrie JW, Froidevaux NM, Witkovic YD. Different digital paths to the keg? How exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students. *Addict Behav* 2016 Jun;57:21-29. [doi: [10.1016/j.addbeh.2016.01.011](https://doi.org/10.1016/j.addbeh.2016.01.011)] [Medline: [26835604](https://pubmed.ncbi.nlm.nih.gov/26835604/)]
27. Geusens F, Beullens K. The association between social networking sites and alcohol abuse Among Belgian Adolescents. *J Media Psychol* 2016 Dec 30:1-10. [doi: [10.1027/1864-1105/a000196](https://doi.org/10.1027/1864-1105/a000196)]

28. Hendriks H, Gebhardt WA, van den Putte B. Alcohol-related posts from young people on social networking sites: content and motivations. *Cyberpsychol Behav Soc Netw* 2017 Jul;20(7):428-435. [doi: [10.1089/cyber.2016.0640](https://doi.org/10.1089/cyber.2016.0640)] [Medline: [28650223](https://pubmed.ncbi.nlm.nih.gov/28650223/)]
29. Moreno MA, Briner LR, Williams A, Brockman L, Walker L, Christakis DA. A content analysis of displayed alcohol references on a social networking web site. *J Adolesc Health* 2010 Aug;47(2):168-175 [FREE Full text] [doi: [10.1016/j.jadohealth.2010.01.001](https://doi.org/10.1016/j.jadohealth.2010.01.001)] [Medline: [20638009](https://pubmed.ncbi.nlm.nih.gov/20638009/)]
30. Moewaka Barnes H, McCreanor T, Goodwin I, Lyons A, Griffin C, Hutton F. Alcohol and social media: drinking and drunkenness while online. *Crit Public Health* 2016;26(1):62-76. [doi: [10.1080/09581596.2015.1058921](https://doi.org/10.1080/09581596.2015.1058921)]
31. Dunlop SM, Kashima Y, Wakefield M. Predictors and consequences of conversations about health promoting media messages. *Communication Monographs* 2010 Dec;77(4):518-539. [doi: [10.1080/03637751.2010.502537](https://doi.org/10.1080/03637751.2010.502537)]
32. Hendriks H, de Bruijn GJ, van den Putte B. Talking about alcohol consumption: health campaigns, conversational valence, and binge drinking intentions. *Br J Health Psychol* 2012 Nov;17(4):843-853. [doi: [10.1111/j.2044-8287.2012.02080.x](https://doi.org/10.1111/j.2044-8287.2012.02080.x)] [Medline: [22646524](https://pubmed.ncbi.nlm.nih.gov/22646524/)]
33. Hendriks H, van den Putte B, de Bruijn GJ. Changing the conversation: the influence of emotions on conversational valence and alcohol consumption. *Prev Sci* 2014 Oct;15(5):684-693. [doi: [10.1007/s11121-013-0418-2](https://doi.org/10.1007/s11121-013-0418-2)] [Medline: [23812888](https://pubmed.ncbi.nlm.nih.gov/23812888/)]
34. Cialdini RB, Trost MR. Social influence: social norms, conformity and compliance. In: Gilbert DT, Fiske ST, Lindzey G, editors. *The Handbook of Social Psychology*. New York: McGraw-Hill; 1999:151-192.
35. Haines M, Spear SF. Changing the perception of the norm: a strategy to decrease binge drinking among college students. *J Am Coll Health* 1996 Nov;45(3):134-140. [doi: [10.1080/07448481.1996.9936873](https://doi.org/10.1080/07448481.1996.9936873)] [Medline: [8952206](https://pubmed.ncbi.nlm.nih.gov/8952206/)]
36. Mollen S, Rimal RN, Ruiter RAC, Kok G. Healthy and unhealthy social norms and food selection. Findings from a field-experiment. *Appetite* 2013 Jun;65:83-89 [FREE Full text] [doi: [10.1016/j.appet.2013.01.020](https://doi.org/10.1016/j.appet.2013.01.020)] [Medline: [23402712](https://pubmed.ncbi.nlm.nih.gov/23402712/)]
37. Asch SE. Studies of independence and conformity: a minority of one against a unanimous majority. *Psychol Mon Gen Applied* 1956;70(9):1-70. [doi: [10.1037/h0093718](https://doi.org/10.1037/h0093718)]
38. Gerard HB, Wilhelmy RA, Conolley ES. Conformity and group size. *J Pers Soc Psychol* 1968 Jan;8(1):79-82. [Medline: [4866675](https://pubmed.ncbi.nlm.nih.gov/4866675/)]
39. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull* 1995 May;117(3):497-529. [Medline: [7777651](https://pubmed.ncbi.nlm.nih.gov/7777651/)]
40. Janis IL, King BT. The influence of role playing on opinion change. *J Abnorm Psychol* 1954 Apr;49(2):211-218. [Medline: [13151774](https://pubmed.ncbi.nlm.nih.gov/13151774/)]
41. Alhabash S, McAlister AR, Quilliam ET, Richards JI, Lou C. Alcohol's getting a bit more social: when alcohol marketing messages on Facebook increase young adults' intentions to imbibe. *Mass Commun Soc* 2015 Dec 20;18(3):350-375. [doi: [10.1080/15205436.2014.945651](https://doi.org/10.1080/15205436.2014.945651)]
42. Cialdini RB, Kallgren CA, Reno RR. A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. *Adv Exp Soc Psychol* 1991;24:201-234. [doi: [10.1016/S0065-2601\(08\)60330-5](https://doi.org/10.1016/S0065-2601(08)60330-5)]
43. Lang TA, Secic M. *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers*. Philadelphia: ACP Press; 2006.
44. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-related posts on Instagram. *Prev Sci* 2016 Dec;17(6):710-720 [FREE Full text] [doi: [10.1007/s11121-016-0669-9](https://doi.org/10.1007/s11121-016-0669-9)] [Medline: [27262456](https://pubmed.ncbi.nlm.nih.gov/27262456/)]
45. Riby DM, Hancock PJB. Viewing it differently: social scene perception in Williams syndrome and autism. *Neuropsychologia* 2008 Sep;46(11):2855-2860. [doi: [10.1016/j.neuropsychologia.2008.05.003](https://doi.org/10.1016/j.neuropsychologia.2008.05.003)] [Medline: [18561959](https://pubmed.ncbi.nlm.nih.gov/18561959/)]
46. Sasson N, Tsuchiya N, Hurley R, Couture SM, Penn DL, Adolphs R, et al. Orienting to social stimuli differentiates social cognitive impairment in autism and schizophrenia. *Neuropsychologia* 2007 Jun 18;45(11):2580-2588 [FREE Full text] [doi: [10.1016/j.neuropsychologia.2007.03.009](https://doi.org/10.1016/j.neuropsychologia.2007.03.009)] [Medline: [17459428](https://pubmed.ncbi.nlm.nih.gov/17459428/)]
47. Moreno MA, van der Stoep A, Parks MR, Zimmerman FJ, Kurth A, Christakis DA. Reducing at-risk adolescents' display of risk behavior on a social networking web site: a randomized controlled pilot intervention trial. *Arch Pediatr Adolesc Med* 2009;163(1):35-41. [doi: [10.1001/archpediatrics.2008.502](https://doi.org/10.1001/archpediatrics.2008.502)]
48. Carotte E, Vella A, Lim M. Predictors of "Liking" Three Types of Health and Fitness-Related Content on Social Media: A Cross-Sectional Study. *J Med Internet Res* 2015 Aug 21;17(8):e205 [FREE Full text] [doi: [10.2196/jmir.4803](https://doi.org/10.2196/jmir.4803)] [Medline: [26297689](https://pubmed.ncbi.nlm.nih.gov/26297689/)]
49. Chu KH, Allem JP, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. *Tob Control* 2016 Sep;26(5):575-578 [FREE Full text] [doi: [10.1136/tobaccocontrol-2016-053052](https://doi.org/10.1136/tobaccocontrol-2016-053052)] [Medline: [27660111](https://pubmed.ncbi.nlm.nih.gov/27660111/)]
50. Allem JP, Chu KH, Cruz TB, Unger JB. Waterpipe promotion and use on Instagram: #hookah. *Nicotine Tob Res* 2017 Oct 01;19(10):1248-1252. [doi: [10.1093/ntr/ntw329](https://doi.org/10.1093/ntr/ntw329)] [Medline: [28077449](https://pubmed.ncbi.nlm.nih.gov/28077449/)]
51. Allem JP, Escobedo P, Chu KH, Cruz TB, Unger JB. Images of little cigars and cigarillos on Instagram identified by the hashtag #swisher: thematic analysis. *J Med Internet Res* 2017 Jul 14;19(7):e255 [FREE Full text] [doi: [10.2196/jmir.7634](https://doi.org/10.2196/jmir.7634)] [Medline: [28710057](https://pubmed.ncbi.nlm.nih.gov/28710057/)]
52. Holland G, Tiggemann M. "Strong beats skinny every time": disordered eating and compulsive exercise in women who post fitspiration on Instagram. *Int J Eat Disord* 2017 Jan;50(1):76-79. [doi: [10.1002/eat.22559](https://doi.org/10.1002/eat.22559)] [Medline: [27302867](https://pubmed.ncbi.nlm.nih.gov/27302867/)]

53. Farace S, van Laer T, de Ruyter K, Wetzels M. Assessing the effect of narrative transportation, portrayed action, and photographic style on the likelihood to comment on posted selfies. *Eur J Market* 2017 Nov 14;51(11/12):1961-1979. [doi: [10.1108/EJM-03-2016-0158](https://doi.org/10.1108/EJM-03-2016-0158)]
54. Litt D. Social networking sites and adolescent alcohol use: the role of social images, social norms, and social comparison (doctoral dissertation). Washington: The George Washington University; 2010.
55. Flaudias V, de Chazeron I, Zerhouni O, Boudesseul J, Begue L, Bouthier R, et al. Preventing alcohol abuse through social networking sites: a first assessment of a two-year ecological approach. *J Med Internet Res* 2015;17(12):e278 [FREE Full text] [doi: [10.2196/jmir.4233](https://doi.org/10.2196/jmir.4233)] [Medline: [26681577](https://pubmed.ncbi.nlm.nih.gov/26681577/)]
56. Katz D, Allport F. Student Attitudes: A Report of the Syracuse University Research Study. Burlington: Craftsman Press; 1931.
57. Hastings G, Anderson S, Cooke E, Gordon R. Alcohol marketing and young people's drinking: a review of the research. *J Pub Health Pol* 2005 Sep;26(3):296-311. [doi: [10.1057/palgrave.jph.3200039](https://doi.org/10.1057/palgrave.jph.3200039)] [Medline: [16167558](https://pubmed.ncbi.nlm.nih.gov/16167558/)]
58. Wills TA, Sargent JD, Gibbons FX, Gerrard M, Stoolmiller M. Movie exposure to alcohol cues and adolescent alcohol problems: a longitudinal analysis in a national sample. *Psychol Addict Behav* 2009 Mar;23(1):23-35 [FREE Full text] [doi: [10.1037/a0014137](https://doi.org/10.1037/a0014137)] [Medline: [19290687](https://pubmed.ncbi.nlm.nih.gov/19290687/)]
59. Waylen A, Leary S, Ness A, Sargent J. Alcohol use in films and adolescent alcohol use. *Pediatrics* 2015 May;135(5):851-858 [FREE Full text] [doi: [10.1542/peds.2014-2978](https://doi.org/10.1542/peds.2014-2978)] [Medline: [25869367](https://pubmed.ncbi.nlm.nih.gov/25869367/)]
60. Yang B, Zhao X. TV, social media, and college students' binge drinking intentions: moderated mediation models. *J Health Commun* 2018 Dec;23(1):61-71. [doi: [10.1080/10810730.2017.1411995](https://doi.org/10.1080/10810730.2017.1411995)] [Medline: [29265924](https://pubmed.ncbi.nlm.nih.gov/29265924/)]
61. Hebden R, Lyons AC, Goodwin I, McCreanor T. "When you add alcohol, it gets that much better". *J Drug Issues* 2015 Mar 22;45(2):214-226. [doi: [10.1177/0022042615575375](https://doi.org/10.1177/0022042615575375)]

Abbreviations

RQ: research question

Edited by G Eysenbach; submitted 07.11.17; peer-reviewed by E Seabrook, JP Allem, L Thompson; comments to author 03.12.17; revised version received 30.01.18; accepted 08.05.18; published 22.06.18.

Please cite as:

Hendriks H, Van den Putte B, Gebhardt WA, Moreno MA

Social Drinking on Social Media: Content Analysis of the Social Aspects of Alcohol-Related Posts on Facebook and Instagram
J Med Internet Res 2018;20(6):e226

URL: <http://www.jmir.org/2018/6/e226/>

doi: [10.2196/jmir.9355](https://doi.org/10.2196/jmir.9355)

PMID: [29934290](https://pubmed.ncbi.nlm.nih.gov/29934290/)

©Hanneke Hendriks, Bas Van den Putte, Winifred A Gebhardt, Megan A Moreno. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.06.2018. 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.

Original Paper

Young People's Experiences of Viewing the Fitspiration Social Media Trend: Qualitative Study

Stephanie Easton¹, BSc (Hons); Katherine Morton¹, BSc (Hons), MSc; Zara Tappy¹, BSc (Hons); Daniella Francis¹, BSc (Hons); Laura Dennison¹, BSc (Hons), MSc, PhD

Academic Unit of Psychology, University of Southampton, Southampton, United Kingdom

Corresponding Author:

Laura Dennison, BSc (Hons), MSc, PhD

Academic Unit of Psychology

University of Southampton

Shackleton Building

University Road

Southampton,

United Kingdom

Phone: 44 2380595000 ext 7657

Email: L.K.Dennison@soton.ac.uk

Abstract

Background: Social media use has become ubiquitous in the lives of many people, especially young adults. A popular recent trend emerging on social media is that of posting and following 'Fitspirational' content - material that purports to motivate and showcase healthy lifestyle habits, particularly relating to exercise and diet. There is very limited existing literature on how engaging with this type of content influences people's psychological and physical health. Initial studies have focused on concerns over potential negative effects on psychological wellbeing including body image, self-esteem and eating disorders.

Objective: We aimed to address a gap in the literature for exploratory research on this topic from the perspective of users. We used a qualitative approach to explore how people experience viewing Fitspiration on social media including why and how they engage with this material and how they perceive that it affects their thoughts, emotions, behaviour and health.

Methods: We recruited 20 young adults (14 females, 6 males, aged 18-25) who self-declared themselves to be Fitspiration followers to participate in either focus groups or individual interviews. We asked detailed, open-ended questions about their motivations for following Fitspiration, experiences of viewing this content and its perceived impact. We used inductive thematic analysis to derive themes that represented common and salient features of the data set.

Results: Four main themes were developed: 1) A tool with the potential to support healthy living, 2) Unrealistic, untrustworthy content, 3) Negative effects on emotional wellbeing, and 4) Vulnerability and protective factors. Following Fitspirational posts on social media can provide young people with knowledge and motivation that may support healthy lifestyle behaviours. However, a range of harms also appeared to arise from Fitspiration viewing ranging from minor annoyances and frustrations to more meaningful negative effects on psychological & physical health. These negative effects seemed to persist despite individuals acknowledging that the material can be unrealistic, and believing that they are personally equipped to minimise harms to themselves.

Conclusions: This study suggests that Fitspiration on social media can be attractive and compelling for young people but appears to bring about negative as well as positive effects. Future research should aim to confirm the scale and intensity of positive and negative effects and investigate ways of harnessing desirable outcomes and minimising undesirable outcomes.

(*J Med Internet Res* 2018;20(6):e219) doi:[10.2196/jmir.9156](https://doi.org/10.2196/jmir.9156)

KEYWORDS

Social media; Fitspiration; behaviour; health; body image; self-esteem; eating disorders; qualitative research; focus groups; interviews

Introduction

In 2017, 66% of UK adults (aged ≥ 16 years) used the internet for social networking, with 96% of young adults (16-24 years) reported to be active users of social networking sites [1]. Social media platforms enable users to create personal profiles and share content (typically photos and text) and view and comment on the posts of users in their network of friends or “followers” who opt to receive updates on the individual’s social media postings.

In recent years, a trend that has emerged on social media is that of posting and following “Fitspirational” content—material that purports to motivate and promote healthy lifestyle habits, especially associated with exercise and diet. “Fitspiration” is a portmanteau, blending “fitness” and “inspiration” and is particularly prevalent on image-based social media platforms such as Instagram, as well as other platforms, including Facebook, Twitter, Tumblr, and Pinterest. It is essential to be aware that anyone can create a Fitspiration post, just by using the #Fitspiration hashtag. Some posts are created by celebrities and fitness and nutrition professionals, whereas others are created by members of the public who may or may not have relevant knowledge or experience. As of February 2018, the authors identified 14.3 million public posts under the metadata #Fitspiration hashtag on Instagram compared with 1.8 million noted in January 2014 [2], signifying a rapid growth in the popularity of this trend.

Fitspiration emerged within the fitness community as an allegedly healthy alternative in response to previous trends such as “Thinspiration” and “Bonespiration” (images that idealize thin bodies and protruding bones) [3]. Tiggemann and Zaccardo [4] reported that social media content tagged with the “#Fitspiration” hashtag tends to feature people (often females) partaking in exercise or dressed in sportswear, or healthy food; these are occasionally overlain with (ostensibly) inspirational quotes or slogans advocating strength, fitness, personal effort, challenge, and empowerment such as “healthy is sexy,” “eating well is a form of self-respect,” or “be stronger than your excuse.” Some social media accounts attract thousands or millions of followers with some of the most famous posters making money through merchandise and product placement [5]. Social media users encounter Fitspirational content either deliberately, by linking with and following users known for posting such content or inadvertently through the posts and shares of other people already in their network [4].

On the surface, the Fitspiration trend could appear benign or even positive. Efforts to motivate and support young people to develop and maintain healthy lifestyle habits are welcomed and might appear refreshing compared with some more apparently worrying online trends and issues (eg, “Thinspiration”). If people are, indeed, motivated and supported in making healthy lifestyle changes through Fitspiration, the scope for benefits to health outcomes could be massive given the extensive reach of these social media platforms. However, no clear evidence demonstrates these positive effects. While Talbot (2017) determined that Fitspiration does appear to be less objectifying than Thinspiration and Bonespiration, it was still concluded that

sufficient similarities existed between these trends to cause concern [3]. Given the newness of the trend, the existing literature remains limited. Several studies using content analysis have concurred that Fitspiration posts perpetuate pervading body image ideals (very lean females and very muscular males), are sexually objectifying, and tend to emphasize appearance over health [6-11]. However, Deighton-Smith and Bell [11] also identified some potentially positive features, including the emphasis on personal control and commitment and building a sense of community and social support. In addition, a concern has been raised about the posters of Fitspirational content, with one study reporting that females who post Fitspirational content have higher scores on measures of disordered eating and compulsive exercise than control females [12]. This study also demonstrated that these females tend to be at higher risk of eating disorders based on scores from the clinical “Drive for Thinness” measure. Furthermore, a content analysis of Fitspirational blogs (in which people write about their experiences of living a healthier lifestyle) provided evidence of problematic eating and negative attitude toward food and body image [13].

Mixed evidence exists regarding how people react to Fitspirational posts. An experimental study reported that exposing female students to Fitspirational images provoked more short-term negative mood and body dissatisfaction and declined appearance-related self-esteem compared with control images [3]. During the study, participants exposed to Fitspiration engaged in more comparison based on their appearance than control participants, and this comparison activity seemingly mediated the effects of the Fitspiration exposure on mood and body dissatisfaction. In addition, participants viewing Fitspiration did exhibit an increase in the measure of inspiration about healthy eating and exercise [4]. A qualitative study interviewed young male Fitspiration followers and revealed that they used the content purposefully (to educate themselves on workout techniques). Although Fitspiration could make them feel inferior about their physical appearance, they engaged in downwards comparison on traits other than appearance and fitness, attributing negative personality traits and neediness to the posters that seemed to help maintain their self-esteem [14].

In summary, although limited research has been conducted to date, the emerging evidence is mixed, suggesting that Fitspiration might exert a negative impact, whereas certain aspects could be inspiring and support behavior change. Thus, further research is warranted on users’ experiences of following Fitspiration. This study used an inductive qualitative approach to address this gap in the literature. Qualitative studies are helpful in emerging fields of research, allowing the emergence of rich data related to a range of perspectives and experiences, rather than constraining data collection to specific anticipated issues [15]. Thus, this study aims to investigate how people experience viewing Fitspiration on social media, including why and how they engage with this material and how they perceive that it affects their thoughts, emotions, behavior, and health.

Methods

Design

This study used a qualitative design and collected data using focus groups and individual interviews. We obtained ethical approval from the Ethics Committee of the Psychology Department at the University of Southampton (Southampton, UK; Study ID: 24273).

Sampling and Recruitment

Participants were eligible for enrollment if they were young adults (aged 18-25 years; as this age group constitutes key consumers of social media, in general [1], and Fitspiration, specifically [16]) who self-defined themselves as followers of Fitspirational content on social media.

Participants were opportunistically enrolled using posters and social media, and course credits were offered to undergraduate students in exchange for participation. In addition, all participants were entered into a prize draw for a chance to win a gift voucher (£15). After viewing adverts, participants contacted the researchers to express interest, read an information sheet, had an opportunity to raise queries, and then signed a consent form before a focus group or interview was arranged.

Data Collection Procedure

We used focus groups to promote the opportunity for participants to share anecdotes and interact with each other to share experiences and perspectives. Notably, four focus groups were held, which comprised of 4, 2, 5, and 3 participants, respectively. In addition, six individual face-to-face interviews were arranged to accommodate participants who could not attend any of the focus groups.

Each focus group was facilitated by two researchers (SE, ZT, or DF, all psychology students) and began with a welcome statement. Then, each participant completed a short questionnaire covering demographics and brief questions about their social media use, and health-related lifestyle behavior to enable us to define our sample. Participants were shown handouts with a selection of nine examples of Fitspiration to elucidate what is implied by Fitspiration on social media and act as a starter activity to prompt their thoughts on the topic. One example included a picture of a female in activewear with text overlay “how bad do you want it.” A question schedule (Appendix 1) was used to elicit in-depth accounts, stories, or opinions about motivations for viewing/following Fitspirational posts on social media, how Fitspiration is used, aspects that are liked and disliked, and perceptions of the ways in which it might affect their behavior, health, thoughts, and feelings. In addition, we used neutral prompts to probe further and encourage participation from all focus groups’ members. Furthermore, individual interviews followed the same procedure and used the same question schedule as the focus groups.

All interviews and focus groups were audiotaped. The recordings were transcribed verbatim, replacing participant names with pseudonyms. All transcripts were carefully checked against recordings.

Data Analysis

Using the inductive thematic analysis, we analyzed the transcripts with techniques from grounded theory, such as constant comparison, to ensure the themes being developed remained close to the original data [17,18]. The analysis was led by SE, an undergraduate psychology student and Fitspiration poster and follower, with regular supervision and analytical input from LD, an experienced qualitative researcher and health psychologist, and KM (a trainee health psychologist with qualitative research experience). Of note, the analysis was inductive, that is themes were developed from the participants’ raw data “upwards,” rather than searching for material that fit with a preexisting theory, model, or structure. The analysis started with reading transcripts and listening to audio-recordings to extensively familiarize with the data. Initial coding involved attaching descriptive labels by hand to parts of the transcripts associated with the research question. The analysis proceeded to develop themes from these codes that captured key patterns and features in the data. Then, the theme development was attained by an iterative process of clustering together similar codes into themes while engaging in the process of constant comparison with the original transcripts to check that themes were grounded in the data and were not being affected by the researchers’ preconceptions or theoretical assumptions [18]. Finally, themes were iteratively reviewed, refined, organized, and relabeled until a set of rich, coherent themes, and subthemes was created in a coding manual (Appendix 2).

In line with the grounded theory approach, analysis and data collection were performed concurrently, enabling us to (1) adjust our sampling strategy to deliberately sample participants with characteristics that were underrepresented in our sample; and (2) adjust the data collection to follow up on analytic insights and emerging ideas [19]. For example, an early analysis of transcripts suggested some possible differences in how males perceive Fitspiration; however, we only had limited male participants. Thus, we sought out more male participants to collect more data to explore this insight further. We ceased recruitment after data had been collected from 20 participants, as it became apparent that significant repetition was occurring, adding little new insight to the ongoing analysis.

Results

Participants

In this study, we enrolled 20 people (14/20, 70%, females and 6/20, 30%, males; age range: 18-25 years; mean age: 20.7 years). Table 1 summarizes the characteristics of the study cohort. The majority of participants (14/20, 70%) were white-British, and 16/20 (80%) were students, including 14 undergraduates and 2 postgraduates.

Of all, 14/20 (70%) reported spending a minimum of 2 hours on social media each day, with 9/20 (45%) spending >1 hour per day on health and fitness-related content, and a minority 1/20 (5%) spending >4 hours per day. The most popular social media platforms for health- and fitness-related content were Instagram, Facebook, and YouTube.

Table 1. Participant characteristics.

Characteristic	Value
Demographics	
Age (years), mean (SD); range	20.7 (1.79); 18–25
Gender, n (%)	
Female	14 (70)
Male	6 (30)
Occupation n (%)	
Undergraduate student	14 (70)
Postgraduate	2 (10)
Non-student	4 (20)
Ethnicity	
White-British	14 (70)
Black-African	1 (5)
Black-Caribbean	1 (5)
Asian-Indian	1 (5)
Other	3 (15)
Social media use, n (%)	
Daily hours spent on social media^a	
<1	2 (10)
1-2	4 (20)
2-3	10 (50)
4+	4 (20)
Daily hours spent on health- and fitness-related social media^a	
<1	11 (55)
1-2	7 (35)
2-3	1 (5)
4+	1 (5)
Social media sites frequently used to view health and fitness related content^a	
Facebook	11 (55)
Twitter	1 (5)
Instagram	18 (90)
Snapchat	5 (25)
Pinterest	2 (10)
YouTube	11 (55)
WordPress	1 (5)
Other	2 (10)
Perceptions of health and lifestyle, n (%)	
Which would you consider yourself to be?^a	
“Underweight”	1 (5)
“About right”	18 (90)
“Overweight”	1 (5)
Would you say you lead an active lifestyle?^a	

Characteristic	Value
Yes	17 (85)
No	3 (15)
Would you say you eat a healthy diet?^a	
Yes	17 (85)
No	3 (15)

^aSelf-reported estimates.

Table 2. Themes and subthemes.

Theme	Subtheme
A tool with some potential to support behavior change	<ul style="list-style-type: none"> • Information and ideas • Being inspired and motivated
Unrealistic, untrustworthy content	<ul style="list-style-type: none"> • Trust and deception • Unrealistic unattainable lifestyles • Inappropriate or abandoned goals
Negative effects on emotional well-being	<ul style="list-style-type: none"> • Feeling guilty about choices and behavior • Feeling low about my body • Concerns about eating • Feeling compelled to keep using Fitspiration
Vulnerability and protective factors	<ul style="list-style-type: none"> • Gender • Age • Mood • Engaging in a critical way • Filtering and choosing relevant content to follow

Furthermore, 18/20 (90%) participants classified themselves as being of healthy weight, 17/20 (85%) felt they led an active lifestyle and 17/20 (85%) ate a healthy diet.

Themes

We developed the following four key themes in this study: (1) *a tool with some potential to support behavior change*; (2) *unrealistic, untrustworthy content*; (3) *negative effects on emotional well-being*; and (4) *vulnerability and protective factors*. Table 2 presents these themes and their subthemes; also, each theme and subtheme is discussed alongside illustrative quotations in the following section. We have replaced participant names with participant numbers, and gender and age have also been indicated (eg, P1, F, 20).

A Tool With Some Potential to Support Behavior Change

Participants revealed benefits that could be gained from Fitspiration content that facilitated making changes to their behavior.

Information and Ideas

Participants discussed following Fitspiration accounts on social media to gain practical ideas and tips about healthy lifestyles. They discussed how Fitspiration content successfully provided them with ideas for healthy recipes, workouts, exercise techniques, and gym merchandise.

It is nice to see to get some ideas [...] when they post exercises and I think “oh that might be something I haven’t tried before.” [P1, F, 20]

Being Inspired and Motivated

Participants described how Fitspiration content boosted their motivation to attend a gym, follow a nutritious diet, and helped them to adopt a positive mind-set. They described how motivation could be explicitly triggered by written, inspirational quotes.

It helps me to set targets [...] see what I need to be doing and then kind of get me the road to doing it. [P3, F, 19]

In addition, observing posters helped them attain their goals and boosted motivation for working toward their health targets, and individuals posting Fitspiration content acted as aspirational figures and role models.

I think they can be good for getting you motivated like definitely, [...], if I try hard I could look like this. [P5, F, 20]

Unrealistic, Untrustworthy Content

All participants discussed how they often find Fitspirational content to be unrealistic and difficult to associate with. Besides creating frustration and negative feelings toward Fitspiration posters, the unrealistic content seemed to adversely affect their goal setting and perseverance.

Trust and Deception

Some participants explained the difficulty in determining which information could be trusted. With no evidence of qualifications, many participants were uncertain which Fitspiration posters possessed adequate expertise to offer valid advice. In addition, participants were conscious that Fitspiration posters mislead and deceive users with filtered content, good lighting and specific poses, and cherry-picked only the best parts of their lives to share.

People are putting up their best photos for a reason, and it's not like real life [P2, F, 21]

Participants were concerned about Fitspiration posters having ulterior motives for posting material, especially an awareness that some posters might gain financially by supporting and endorsing brands.

They might have some like agenda, and maybe they're not being so honest about that [P14, F, 25]

Furthermore, participants discussed various types of products they had seen endorsed or advertised and conveyed frustration that they were being sold products rather than being offered valuable advice. Seemingly, they found it difficult to distinguish whether some Fitspiration posts had hidden intentions to promote products shown in the post, making them wary of trusting the content of the post.

They are just getting money out of it [...] that is all they are doing it for. [P1, F, 20]

Unrealistic Unattainable Lifestyles

Participants perceived that the lifestyles depicted by many posters were difficult or impossible to associate with and emulate. More famous posters were perceived to possess luxuries that suggested an advantage over their followers, making the lifestyle less obtainable.

You have a gym in your house! How is that like real life? [P6, F, 20]

In particular, participants were especially distrustful of celebrities, as they felt that cosmetic surgery and body-alteration made them unrealistic models for followers.

They've all had surgery so people will be working towards a goal that's not achievable [P19, F, 22]

Several participants (whether students or employed) considered cost as a barrier to living a healthy lifestyle, especially because of the perceived cost of the food, gym memberships, and clothing.

I think something that's not taken into account is that being fit is like money as well, like gym membership and gym clothes and healthy food, that's so much money so if you can't afford it you feel like, well I feel like there's no point. [P9, F, 18]

In addition, participants believed they lacked time to adhere to the lifestyle presented, with one participant even describing poster's habits to be a full-time job.

A lot of people I follow are quite into it and sort of do it every day and most of the time that's all they do

[...] you don't really know whether, if they have another job or not. [P16, M, 22]

Inappropriate or Abandoned Goals

Several participants discussed how Fitspiration, though recognized as untrustworthy and unrealistic, could still affect the types of goals they were aiming for and made them less attainable.

It's probably made my personal goals quite different because they [the Fitspiration figures] obviously look amazing... I wouldn't have set goals that unrealistic if I didn't follow them on social media. [P15, F, 21]

It was recognized that this perspective was associated with feelings of disappointment and pressure.

They [the posters] put... 'you can achieve this in 6 weeks'... and it's physically impossible to achieve that kind of physique in that amount of time and I feel it puts an unfair pressure on. [P18, M, 24]

Some participants reflected that unrealistic content resulted in disengagement with goals that were overly ambitious.

It can make you give up quicker I think [P11, F, 19]

In addition, participants commented that following Fitspiration made their goals more focused on appearance and gaining approval from peers, rather than health.

You kind of lose sight of the goal of actually trying to become healthy rather than just looking good for pictures on social media. [P16, M, 22]

In fact, one of the participants perceived this focus on image over health as an issue with Fitspiration posters, implying that they have the wrong priorities.

They don't go to the gym and things because they want to be healthy and lead a healthy lifestyle, they want to have a body from which they can take pictures and post it to Instagram. [P16, M, 22]

Negative Effects on Emotional Well-Being

Participants discussed various negative emotional experiences stemming from viewing Fitspiration content. Mostly, this discussion was about personal experiences and feelings, whereas other comments seemed more speculative and hypothetical.

Feeling Guilty About Choices and Behavior

One of the most often discussed feelings was guilt about not following a similar lifestyle to those advocated in Fitspiration posts. Viewing Fitspiration posts seemingly provoked participants to compare these with their health and fitness-related habits and feel guilty when they did not match up.

It makes me feel quite guilty sometimes, if you're just not really in the mood to um, like be productive or proactive [...] and then you see all these posts, and it's telling you that you should. [P13, F, 20]

In fact, some participants displayed ambivalence around this guilt response as they knew that guilt was not appropriate or logical because of their awareness of the unrealistic nature of the posts.

I feel so guilty if I see all this, but then I'm like why am I feeling guilty? Because what I have just done is normal. [P1, F, 20]

Feeling Low About My Body

Participants reported being left with negative feelings about their body when comparing themselves with Fitspiration images.

It makes me not enjoy things like going to the beach and like taking photos on holiday because you don't look like the photos on Instagram. [P15, F, 21]

When I see fitness accounts where all the girls are like svelte and toned, I'm like oh, it's hard to love me when I look like this. [P6, F, 20]

In fact, one participant highlighted that those who differ from the typical body type within Fitspiration content could be at a heightened risk of experiencing these negative feelings.

If you're of a bigger size, it can make you feel horrendous, it can make you feel completely alien and that you shouldn't look like that. [P15, F, 21]

Moreover, failure to make rapid progress toward the ambitious appearance-related goals that they had set for themselves could trigger negative feelings about themselves.

You can't have this tiny waist and massive bum, [...] you may if you did it [exercise] for a few years, a long time, [...] but it can make you feel kinda down about yourself. [P8, F, 20]

Concerns About Eating

Participants indicated that Fitspiration exerted both positive and negative effects on their eating habits. Although increased awareness about food choices was described by some participants, others found following the eating plans advocated by posters impossible to sustain and were aware of rebounding to extremes of unhealthy eating, or even binge eating.

I'm a lot more aware of food groups, the whole ideal food groups plate arrangement, it's like half vegetables, a quarter of protein, a quarter of carbs, I'm very aware of doing that when I have my dinners. [P6, F, 20]

I think it has made me a lot more wary of what I put into my body but then I will have blow out days and just like literally shove food down. [P5, F, 20]

Some other participants discussed that viewing Fitspiration posts encouraged their obsession with calorie counting. In fact, a few also believed that some of the diet-related material could even instigate an eating disorder, especially if they were unable to recognize that habits were becoming unhealthy.

If I followed their food account where they tell me to eat healthily and I couldn't, I'd probably end up with an eating disorder. [P6, F, 20]

Feeling Compelled to Keep Using Fitspiration

Several participants described experiencing conflict as they knew that Fitspiration posts could elicit various negative

thoughts, behaviors, and moods; yet, they found themselves viewing it regularly.

In one way you're like really attracted to it but in some ways you find it really annoying and it puts you down [P14, F, 25]

In addition, some participants described how they had initially followed Fitspiration content for a specific purpose and believed that it had not successfully fulfilled that purpose, but having got involved in the social and community aspect of it they felt compelled to continue engaging with it. Furthermore, many seemed to find this type of social media usage compelling or even addictive.

Vulnerability and Protective Factors

Participants perceived that various contextual factors affected the degree to which they and others experienced negative impacts from Fitspiration content.

Gender

Several participants (both males and females) believed that females tended to be more vulnerable to the negative effects of the exposure of Fitspiration and, indeed, mass media more generally than males.

Females tend to be more sensitive [...] it can have a bit more of a deeper effect on them, whereas men tend to be a bit more hard-headed. [P16, M, 22]

In line with this, most of the talk about guilt, body image, and concerns about eating and compulsive viewing came from females (refer "Negative Effects on Emotional Well-being" theme above). In addition, participants perceived that Fitspiration perpetuated a long-standing pressure to conform to the existing female body ideals.

For years there's been this problem with media, especially girls like feeling they need to look a certain way. [P8, F, 20]

Yet, one male participant described negative emotions associated with failing to fulfill appearance-based expectations that had been generated by following Fitspiration posts.

I think mentally, it's quite stressful sometimes if you put yourself up to a task that you can't achieve [...] looking at yourself in the mirror [...] you're just not really seeing results, it can definitely have a negative impact [P16, M, 22]

Furthermore, the participants speculated that males may well be negatively affected but might not express their feelings because of gender norms.

I reckon it probably negatively affects boys, but they don't express it [...] If a boy did he'd probably be called a wimp. [P6, F, 20]

Age

Participants anticipated that younger users than themselves tended to be deceived by unrealistic content, consumed by the lifestyle and, thus, most likely to experience negative effects.

In line with this, one participant suggested that her maturity enabled her to control how much the content affected her behavior.

I am old enough, wise enough to know that it is cool too if I have had Uni all day [and therefore not had time to exercise] then that's fine. [P2, F, 21]

Mood

Participants discussed their affective state influencing how they responded to Fitspiration posts and elucidated how Fitspiration could intensify their emotional state if they were in a bad mood. Conversely, if they were viewing Fitspiration while already in a good mood, it seemed to buffer against negative effects and the material could enhance their motivation to emulate healthy behaviors they have witnessed online.

Let's say I'm already feeling up for some workout [...] then I see some Fitspiration post, I might be like more inclined to go and do it and then feel more like, positive about it but if I'm in a bad mood then I don't want to see that, I just get grumpy [P14, F, 25]

More worryingly, participants discussed how distress associated with life events, such as the end of a relationship, could render them highly sensitive to negative effects of Fitspiration content.

Combined with just being broken up with, it just like destroyed my self-esteem seeing all these really fit people. [P5, F, 20]

Engaging in a Critical Way

Seemingly, participants adopted certain approaches to Fitspiration viewing that they believed enabled them to follow this content with a decreased risk of psychological harm. One technique was to use the content purposefully but step back from getting too immersed.

It's kind of best to keep them at a distance [...] use them for inspiration now and then but I don't think it's healthy to be completely immersed [P1, F, 20]

Others, however, felt they had adequate knowledge or education that enabled them to assess the messages and images being presented critically.

I kind of have got the knowledge to know that your abs just aren't going to just appear [P1, F, 20]

Moreover, some were especially aware and well-informed about tricks and effects of the media and believed that their less naïve and more critical approach provided them with some protection against negative effects of Fitspiration.

I study marketing [...] I'm a lot more knowledgeable and less naïve to the content being advertised to me. [P19, F, 22]

However, participants who felt their knowledge protected them still provided accounts of various ways that Fitspiration viewing had adversely affected them personally.

Filtering and Choosing Relevant Content to Follow

Some participants carefully selected and filtered the content that they followed and engaged with to get the most from it.

For example, a few participants described selecting content to view that was consistent with their goals and minimizing exposure to material perceived as irrelevant.

When they put up their personal life and things...I'm not interested in them as a person which is probably quite bad but I just want to see the videos of what they do in the gym. [P15, F, 21]

Furthermore, several participants discussed being selective in following Fitspiration posters who they felt they could associate with to ensure their goals were more realistic to attain.

I always like to follow normal people as well [...] these things are actually achievable. [P19, F, 22]

Discussion

Key Findings

This in-depth qualitative study exploring the experiences and perspectives of Fitspiration followers revealed several crucial insights. Consistent with Palmer [14], this study reported that participants described a desire to gain information as a critical driver of consumption of Fitspirational material; specifically, our participants were interested in gaining information related to exercise techniques, healthy recipes, and workouts. Like Tiggeman and Zaccardo [7], this study demonstrated that participants felt inspired and motivated by Fitspirational content; however, this does not seem to translate into positive dietary change and physical activity routinely, a finding also reported by Palmer [14]. Nevertheless, the majority of our findings highlighted concerning aspects of the Fitspiration use. Our participants discussed several negative effects ranging from minor (eg, frustration about the deceptive nature of posts, jealousy regarding unattainable body appearance or lifestyles, feeling that their usage had become out of control, guilt about not following the lifestyles advocated, and frustration in being encouraged toward inappropriate goal-setting) to more disconcerting (eg, negative feelings toward their own bodies and indications of some concerning eating habits). The issues of frustration, guilt, and feeling addicted to viewing Fitspirational posts are novel insights emerging from this study. Previously, negative effects of the exposure to Fitspiration on self-esteem, body image, and disordered eating have been reported [4] and speculated [6-13]; this study is in line with these studies and also the proposition that social comparison based on appearance is one of the routes through which Fitspiration exerts negative psychological effects [4].

We determined that our participants were critical, cautious, and questioning of Fitspirational content, highly aware of authenticity and credibility issues, and some made mindful decisions about who and what to follow and what aspects of posts to focus on, which corroborated Palmer's [14] study. In our study, some participants felt that their age, gender, education, or approach to using Fitspiration protected them from a negative psychological impact. However, negative psychological effects seemed to persist despite participants possessing characteristics and capabilities that they believed could buffer them against harm. In addition, our participants exhibited a considerable conflict and ambivalence around

Fitspiration. They persisted in viewing content despite feeling it could be frustrating or even harmful. They also reported feeling guilty and comparing themselves unfavorably to posters despite articulating how guilt is unwarranted as posters are showing edited versions of themselves and sending invitations to lead unviable lifestyles and fulfill impossible appearance and health-related goals. Finally, participants discussed female gender as a factor associated with vulnerability to negative effects from using Fitspiration; however, this finding might have been driven by stereotypes and gender norms. When reviewing males' descriptions of impact, this study provided evidence that males, as well as females, could be adversely affected by content and were similar in the ways they thought and felt about the material they were seeing.

Study Strengths and Limitations

This study has several strengths and limitations that merit consideration while interpreting the study findings. First, sample composition. Our sample size was relatively small ($n=20$). However, we attained saturation with this number of participants. In addition, we successfully sampled participants with various views and experiences. We attained a variability in age, gender, and ethnicity and enrolled participants that varied in their intensity of engagement with social media, generally, and Fitspiration, specifically. Furthermore, the sample comprised participants who reported being both committed and uncommitted to healthy eating and exercise, although the majority reported being committed. Most participants perceived their weight as "about right," although it is essential to note that the questionnaire only captured self-reported perceptions of weight and not actual body mass index (BMI). Being an exploratory qualitative study, we did not assess participants' objective weight, although this could be an exciting avenue for further research on the impact of Fitspiration.

University students accounted for the majority of our sample. Thus, the findings of this study represent views and experiences of a highly educated group of young people. In addition, since our recruitment strategy resulted in the inclusion of a high number of psychology students, our participants may, because of their education, have been especially attuned to the negative effects of the media and possibly more likely to consider and discuss issues associated with eating disorders and body image. Nonetheless, while some of the discussion was speculative and hypothetical, there was also a considerable discussion of personal experiences and feelings. Overall, because of our sampling, we suspect that our participants might have been more critical and careful with their social media use and plausibly better protected against the negative impact than other young people who follow Fitspiration.

Second, another methodological consideration associated with data collection. The use of focus groups maximized opportunities to create discussion and enable participants to draw on, compare, and contrast experiences and perspectives. However, perhaps social desirability, gender norms, or the effect of dominant individuals could have silenced dissenting opinions and rendered it more difficult to talk about specific personal experiences. In addition, participants might have found that having young female students as interviewers and facilitators

made them reluctant to mention specific issues. We did, however, determine that several participants discussed emotive and personal experiences, and we were reassured that individual interviews elicited similar issues to those raised in focus groups.

Finally, it is imperative to consider that participants might have had concerns about their body image and eating habits before their exposure to Fitspiration. From this study, it is not possible to understand to what extent Fitspiration causes these negative outcomes. However, it remains interesting that participants perceive that Fitspiration is accountable for these feelings.

Future Research

This study suggests several useful directions for future Fitspiration research. Experimental, quasi-experimental, or observational studies could be conducted to attain a quantitative, more objective assessment of the factors emerging as important from this exploratory qualitative study [20,21], including short- and long-term emotional, cognitive, and behavioral responses to Fitspirational content. This study suggests that frustration, guilt, compulsive social media use, self-esteem, body image, and concerns with eating are relevant negative outcomes to measure and that health- and fitness-related knowledge, motivation, and behavior change are relevant positive outcomes to measure. In addition, this study points toward the utility of examining possible moderators of the impact of Fitspiration, including the realism and relatability of posters and the posted material and the characteristics of the follower, including age, gender, preexposure mood, and media-related literacy and critical appraisal skills. This study supports previous suggestions that social comparison could be investigated as one potential mediator of negative effects on the body image and related outcomes [3]. Furthermore, future studies could consider focusing on different groups of Fitspiration followers. A slightly younger (ie, teenage) group might have different motivations for following Fitspiration. However, this age group has been neglected in Fitspiration research thus far.

Given that the emerging research literature on Fitspiration has indicated potential harm yet statistics demonstrate that Fitspiration is a growing trend, it seems useful to consider interventions to decrease or prevent harm from body and appearance-related images and content on social media. Previously, some studies have suggested that psycho-educational interventions could focus on raising awareness of the harms of this sort of content and increasing followers' media literacy and critical appraisal skills [22,23]. This study suggests young people believe that being armed with information and critical appraisal skills will protect them from possible negative effects. However, this study also suggests that critical and careful consumption of Fitspiration might not actually be successful in averting the negative psychological impact. Thus, further exploration and evaluation of potential intervention approaches is warranted.

An alternative direction for research is capitalizing on the positive aspects of Fitspiration in social media. Researchers could develop or adapt theory-based behavior change interventions to enhance the diet and physical activity levels of young adults, which make use of social media platforms to deliver appropriate health-related messages [24]; these platforms

are appealing, influential, and draw in young females in particular [16]. A key challenge could be mimicking what is appealing, engaging, and positive about Fitspiration while omitting what seems harmful.

Conclusions

This study suggests that following Fitspirational posts on social media can provide young people with knowledge and motivation to support healthy lifestyle behaviors; however, following such content also seems to exert some undesirable effects. In addition,

this study suggests the possibility of various harms ranging from minor annoyances and frustrations to more meaningful effects on the mental and physical health. These effects might persist despite users being aware of how unrealistic the material portrayed can be and despite users believing that they are personally well-equipped to minimize harms to themselves. Furthermore, studies to confirm benefits and harms and investigate ways of harnessing positives and minimizing negatives would be helpful additions to this field of research.

Acknowledgments

The study concept and design were led by SE and LD. The data collection was performed by SE, ZT, and DF. The data analysis was performed by SE, LD, KM, ZT, and DF. SE, LD, and KM wrote the paper. All authors revised and approved the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview schedule.

[PDF File (Adobe PDF File), 22 KB - [jmir_v20i6e219_app1.pdf](#)]

Multimedia Appendix 2

Coding manual.

[PDF File (Adobe PDF File), 54 KB - [jmir_v20i6e219_app2.pdf](#)]

References

- Office of National Statistics. Internet access – households and individuals: 2017 URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017> [WebCite Cache ID 6zpKwe22K]
- Marklew D. NewsVineWA. 2014. #Fitspiration or #fitsploitation? How health hashtags can harm URL: <https://newsvinewa.com.au/14170/news/fitspiration-or-fitsploitation/> [WebCite Cache ID 6u6l3gz2H]
- Talbot CV, Gavin J, van ST, Morey Y. A content analysis of thinspiration, fitspiration, and bonespiration imagery on social media. *J Eat Disord* 2017;5:40 [FREE Full text] [doi: [10.1186/s40337-017-0170-2](https://doi.org/10.1186/s40337-017-0170-2)] [Medline: [29021900](https://pubmed.ncbi.nlm.nih.gov/29021900/)]
- Tiggemann M, Zaccardo M. "Exercise to be fit, not skinny": The effect of fitspiration imagery on women's body image. *Body Image* 2015 Sep;15:61-67. [doi: [10.1016/j.bodyim.2015.06.003](https://doi.org/10.1016/j.bodyim.2015.06.003)] [Medline: [26176993](https://pubmed.ncbi.nlm.nih.gov/26176993/)]
- Norton M. Capstone Projects and Theses. Fitspiration: Social Media's Fitness Culture and its Effect on Body Image (2017) URL: http://digitalcommons.csumb.edu/caps_thes_all/138 [accessed 2017-10-10] [WebCite Cache ID 6u6lRh5ow]
- Borzekowski DLG, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: A content analysis of pro-eating disorder Web sites. *Am J Public Health* 2010 Aug;100(8):1526-1534. [doi: [10.2105/AJPH.2009.172700](https://doi.org/10.2105/AJPH.2009.172700)] [Medline: [20558807](https://pubmed.ncbi.nlm.nih.gov/20558807/)]
- Tiggemann M, Zaccardo M. 'Strong is the new skinny': A content analysis of #fitspiration images on Instagram. *J Health Psychol* 2016 Mar 01;1359105316639436. [doi: [10.1177/1359105316639436](https://doi.org/10.1177/1359105316639436)] [Medline: [27611630](https://pubmed.ncbi.nlm.nih.gov/27611630/)]
- Carrotte ER, Prichard I, Lim MSC. "Fitspiration" on Social Media: A Content Analysis of Gendered Images. *J Med Internet Res* 2017 Mar 29;19(3):e95 [FREE Full text] [doi: [10.2196/jmir.6368](https://doi.org/10.2196/jmir.6368)] [Medline: [28356239](https://pubmed.ncbi.nlm.nih.gov/28356239/)]
- Simpson CC, Mazzeo SE. Skinny Is Not Enough: A Content Analysis of Fitspiration on Pinterest. *Health Commun* 2017 May;32(5):560-567. [doi: [10.1080/10410236.2016.1140273](https://doi.org/10.1080/10410236.2016.1140273)] [Medline: [27326747](https://pubmed.ncbi.nlm.nih.gov/27326747/)]
- Ghaznavi J, Taylor LD. Bones, body parts, and sex appeal: An analysis of #thinspiration images on popular social media. *Body Image* 2015 Jun;14:54-61. [doi: [10.1016/j.bodyim.2015.03.006](https://doi.org/10.1016/j.bodyim.2015.03.006)] [Medline: [25880783](https://pubmed.ncbi.nlm.nih.gov/25880783/)]
- Deighton-Smith N, Bell BT. Objectifying Fitness: A Content and Thematic Analysis of #Fitspiration Images on Social Media. *Psychology of Popular Media Culture* 2017;- [doi: [10.1037/ppm0000143](https://doi.org/10.1037/ppm0000143)]
- Holland G, Tiggemann M. "Strong beats skinny every time": Disordered eating and compulsive exercise in women who post fitspiration on Instagram. *Int J Eat Disord* 2017 Jan;50(1):76-79. [doi: [10.1002/eat.22559](https://doi.org/10.1002/eat.22559)] [Medline: [27302867](https://pubmed.ncbi.nlm.nih.gov/27302867/)]
- Boepple L, Thompson JK. A content analysis of healthy living blogs: evidence of content thematically consistent with dysfunctional eating attitudes and behaviors. *Int J Eat Disord* 2014 May;47(4):362-367. [doi: [10.1002/eat.22244](https://doi.org/10.1002/eat.22244)] [Medline: [24420676](https://pubmed.ncbi.nlm.nih.gov/24420676/)]

14. Palmer L. Poppin bottles, getting wheysted. Exploring young mens engagement with fitspiration content and its consequential influences on attitudes and behaviour. *Journal of Promotional Communications* 2015;425-445 [[FREE Full text](#)]
15. Bishop F, Yardley L. Qualitative Assessment. In: In: Ayers S, Baum A, McManus C, Newman S, Wallston K, Weinman J, West R. *Cambridge Handbook of Psychology, Health and Medicine*. 2nd ed. Cambridge: Cambridge University Press; 2007:314-319.
16. Carrotte ER, Vella AM, Lim MSC. Predictors of “Liking” Three Types of Health and Fitness-Related Content on Social Media: A Cross-Sectional Study. *J Med Internet Res* 2015 Aug 21;17(8):e205 [[FREE Full text](#)] [doi: [10.2196/jmir.4803](#)] [Medline: [26297689](#)]
17. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](#)]
18. Glaser B, Strauss A, Strutzel E. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine; 1967.
19. Charmaz K. *Constructing Grounded Theory*. California: Sage; 2014.
20. Yardley L, Bishop F. Mixing qualitative and quantitative methods: A pragmatic approach. In: In: Willig, Carla and Stainton-Rogers, Wendy (eds.) *The SAGE Handbook of Qualitative Research in Psychology*. California: Sage; 2008:352-370.
21. Willig C, Rogers W. *The SAGE handbook of qualitative research in psychology*. California: Sage; 2008.
22. McLean SA, Wertheim EH, Masters J, Paxton SJ. A pilot evaluation of a social media literacy intervention to reduce risk factors for eating disorders. *Int J Eat Disord* 2017 Jul;50(7):847-851. [doi: [10.1002/eat.22708](#)] [Medline: [28370321](#)]
23. McLean SA, Paxton SJ, Wertheim EH. The role of media literacy in body dissatisfaction and disordered eating: A systematic review. *Body Image* 2016 Dec;19:9-23. [doi: [10.1016/j.bodyim.2016.08.002](#)] [Medline: [27572000](#)]
24. 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](#)]

Edited by G Eysenbach; submitted 10.10.17; peer-reviewed by N Deighton-Smith, S Aggarwal, N Bragazzi, M Acosta, H Miller; comments to author 07.01.18; revised version received 02.03.18; accepted 19.03.18; published 18.06.18.

Please cite as:

Easton S, Morton K, Tappy Z, Francis D, Dennison L

Young People's Experiences of Viewing the Fitspiration Social Media Trend: Qualitative Study
J Med Internet Res 2018;20(6):e219

URL: <http://www.jmir.org/2018/6/e219/>

doi: [10.2196/jmir.9156](#)

PMID: [29914859](#)

©Stephanie Easton, Katherine Morton, Zara Tappy, Daniella Francis, Laura Dennison. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 18.06.2018. 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.

Original Paper

Recruiting Hard-to-Reach Populations for Survey Research: Using Facebook and Instagram Advertisements and In-Person Intercept in LGBT Bars and Nightclubs to Recruit LGBT Young Adults

Jamie Guillory¹, MSc, PhD; Kristine F Wiant¹, MSc, PhD; Matthew Farrelly¹, PhD; Leah Fiocco¹, BS; Ishrat Alam¹, BA; Leah Hoffman², MPH; Erik Crankshaw¹, MPH, PhD; Janine Delahanty², MA, PhD; Tesfa N Alexander³, MA, PhD

¹RTI International, Research Triangle Park, NC, United States

²US Food and Drug Administration, Center for Tobacco Products, Silver Spring, MD, United States

³Battelle Memorial Institute, Aberdeen, MD, United States

Corresponding Author:

Jamie Guillory, MSc, PhD

RTI International

3040 E Cornwallis rd

P.O. Box 12194

Research Triangle Park, NC, 27709

United States

Phone: 1 9193163725

Fax: 1 9195416683

Email: jguillory@rti.org

Abstract

Background: Tobacco public education campaigns focus increasingly on hard-to-reach populations at higher risk for smoking, prompting campaign creators and evaluators to develop strategies to reach hard-to-reach populations in virtual and physical spaces where they spend time.

Objective: The aim of this study was to describe two novel recruitment strategies (in-person intercept interviews in lesbian, gay, bisexual, and transgender [LGBT] social venues and targeted social media ads) and compares characteristics of participants recruited via these strategies for the US Food and Drug Administration's *This Free Life* campaign evaluation targeting LGBT young adults who smoke cigarettes occasionally.

Methods: We recruited LGBT adults aged 18-24 years in the United States via Facebook and Instagram ads (N=1709, mean age 20.94, SD 1.94) or intercept in LGBT social venues (N=2348, mean age 21.98, SD 1.69) for the baseline evaluation survey. Covariates related to recruitment strategy were age; race or ethnicity; LGBT identity; education; pride event attendance; and alcohol, cigarette, and social media use.

Results: Lesbian or gay women (adjusted odds ratio, AOR 1.88, 95% CI 1.54-2.29, $P<.001$), bisexual men and women (AOR 1.46, 95% CI 1.17-1.82, $P=.001$), gender minorities (AOR 1.68, 95% CI 1.26-2.25, $P<.001$), and other sexual minorities (AOR 2.48, 95% CI 1.62-3.80, $P<.001$) were more likely than gay men to be recruited via social media (than intercept). Hispanic (AOR 0.73, 95% CI 0.61-0.89, $P=.001$) and other or multiracial, non-Hispanic participants (AOR 0.70, 95% CI 0.54-0.90, $P=.006$) were less likely than white, non-Hispanic participants to be recruited via social media. As age increased, odds of recruitment via social media decreased (AOR 0.76, 95% CI 0.72-0.80, $P<.001$). Participants with some college education (AOR 1.27, 95% CI 1.03-1.56, $P=.03$) were more likely than those with a college degree to be recruited via social media. Participants reporting past 30-day alcohol use were less likely to be recruited via social media (AOR 0.33, 95% CI 0.24-0.44, $P<.001$). Participants who reported past-year pride event attendance were more likely to be recruited via social media (AOR 1.31, 95% CI 1.06-1.64, $P=.02$), as well as those who used Facebook at least once daily (AOR 1.43, 95% CI 1.14-1.80, $P=.002$). Participants who reported using Instagram at least once daily were less likely to be recruited via social media (AOR 0.73, 95% CI 0.62-0.86, $P<.001$). Social media recruitment was faster (incidence rate ratio, IRR=3.31, 95% CI 3.11-3.52, $P<.001$) and less expensive (2.2% of combined social media and intercept recruitment cost) but had greater data quality issues—a larger percentage of social media respondents were lost because of duplicate and low-quality responses (374/4446, 8.41%) compared with intercept respondents lost to interviewer misrepresentation (15/4446, 0.34%; $P<.001$).

Conclusions: Social media combined with intercept provided access to important LGBT subpopulations (eg, gender and other sexual minorities) and a more diverse sample. Social media methods have more data quality issues but are faster and less expensive than intercept. Recruiting hard-to-reach populations via audience-tailored strategies enabled recruitment of one of the largest LGBT young adult samples, suggesting these methods' promise for accessing hard-to-reach populations.

(*J Med Internet Res* 2018;20(6):e197) doi:[10.2196/jmir.9461](https://doi.org/10.2196/jmir.9461)

KEYWORDS

sexual minorities; social media; public health; tobacco; survey methods; transgender persons

Introduction

Background

Public education campaigns aimed at educating the public on the risks of tobacco use are increasingly targeting specific segments of the population who are at risk for tobacco use and hard to reach via traditional methods [1,2]. In the United States, the Center for Tobacco Products (CTP) at the US Food and Drug Administration (FDA) implements a number of these campaigns as part of its mission to educate the public on the harms of tobacco use. Each campaign addresses at-risk, hard-to-reach populations (ie, populations that are difficult to reach for inclusion in surveys via traditional survey research recruitment methods) and delivers compelling content relevant to that specific target population. These campaigns include FDA's flagship campaign, *The Real Cost*, that targets young people aged 12-17 years who are at risk for initiating cigarette smoking or are experimenting with smoking (ie, have smoked fewer than 100 cigarettes in their lifetime) and includes a component that focuses on educating hard-to-reach rural male youth at risk for using smokeless tobacco [1]. FDA's *Fresh Empire* campaign targets African American, Hispanic, Asian or Pacific Islander, and multiracial youth who are influenced by the hip hop peer crowd and at risk for cigarette smoking [2]. *This Free Life* is FDA's public education campaign focusing on lesbian, gay, bisexual, and transgender (LGBT) young adults aged 18 to 24 years who smoke cigarettes occasionally [3].

LGBT young adults are hard to reach [4] and have significant tobacco use disparities compared with non-LGBT young adults [5-7], being almost twice as likely to use tobacco as their non-LGBT peers [8,9]. Elevated risk of tobacco use among LGBT individuals has been attributed to LGBT-targeted tobacco product marketing [5,7,10-12] and minority stress (ie, strain resulting from stigma and discrimination associated with having a minority identity) [13-17]. LGBT minority stress increases risk of depression, alcohol and other substance abuse, homelessness, and poorer health, which are all factors associated with tobacco use [6,15,16,18-22].

Survey Recruitment Methods for Hard-To-Reach Populations

Developing tobacco public education campaigns for hard-to-reach populations such as LGBT young adults also comes with the challenge of reaching these populations to evaluate whether the campaign is effectively educating them about the harms of tobacco use. Researchers are increasingly turning to innovative strategies for recruiting hard-to-reach populations vs traditional methods [23-34]. One strategy that

researchers use to recruit young adults involves conducting intercept interviews in social venues (eg, bars and nightclubs) where the target population spends time. A number of researchers have used this strategy to recruit young adults who are at a higher risk for smoking and alcohol use [24-28]. Furthermore, Fallin and colleagues used this strategy to successfully recruit LGBT young adults in bars and nightclubs [23].

A second strategy that has become popular for recruiting hard-to-reach populations for survey research is the use of targeted ads on social media platforms such as Facebook, Instagram, and Twitter [29,31-37]. Social media platforms possess massive quantities of user data that allow for highly specific targeting of ads to hard-to-reach populations on multiple features such as age, gender, location, interests, and relationship preferences (women interested in women, women interested in women and men, men interested in men, and men interested in women and men). A growing number of studies have successfully recruited hard-to-reach populations via social media, including young adult and adolescent smokers in the United States [29,31-32], adult electronic cigarette users [35], adult gay men [33,38,39], gay and bisexual youth [30], adolescent and young adult women in Australia [37,40], and HIV-positive adults in the United States [34].

This Study

In the present research, we focus on data collected for the evaluation of FDA's *This Free Life* tobacco public education campaign targeting LGBT young adults. *This Free Life* engages with the target group in 12 designated market areas (DMAs) in the United States through multiple channels including social media and LGBT-specific digital sites, streaming radio, LGBT print media, branded promotions at LGBT events and social venues, and out-of-home media such as signage at bus stops in areas where LGBT young adults are likely to socialize. From a campaign evaluation perspective, we consider these 12 DMAs to be treatment DMAs and compare them against data we collected for the evaluation in 12 control, or comparison, DMAs where no events occur, and minimal campaign advertising occurs. The data we present in this paper come from the baseline wave of data collection that occurred immediately before the launch of the *This Free Life* media campaign in the 12 treatment DMAs in the United States. We employed two unique strategies to recruit LGBT young adults for this study. First, we conducted in-person intercept screening interviews using tablet devices in social venues such as bars and nightclubs that we identified as popular among LGBT young adults in each of the 24 DMAs. Second, we used social media ads on Facebook and Instagram that linked to a Web screening instrument. For social media

ads, we used targeting tools and targeted ad content to recruit LGBT young adults in the 24 DMAs. In this study, we compare the cost and time efficiency of recruitment and quality of data gleaned between the two novel methods. Considering that findings from previous research indicate that data collection via social recruitment is more time-efficient than traditional methods (eg, email invitations and print ads) [36], in combination with the fact that a large amount of resources and time are required for intercept recruitment, we hypothesize that the sample will be recruited more quickly and at a lower cost via social media than intercept. Furthermore, we explore how the LGBT young adults we recruited via these methods differ by LGBT identity, demographics, cigarette use, alcohol use, social media use, and participation in LGBT culture.

Methods

Participants

Eligible participants were young adults, aged 18 to 24 years, who self-identified as LGBT, and lived in one of the 24 DMAs in the United States (N=4057, mean age 21.54, SD 1.87). We recruited participants from February 2016 to May 2016 before the launch of FDA’s *This Free Life* media campaign. The study was approved by RTI International’s institutional review board.

Recruitment Method

We recruited participants using either in-person intercept interviews in LGBT social venues (N=2348) or via social media ads (N=1709).

Intercept Recruitment

We intercepted and asked participants to complete a screener (ie, screening instrument) in LGBT social venues that we identified through Web searches and recommendations from local field staff. We discuss full details of intercept recruitment in the Procedure section.

Social Media Recruitment

We used Facebook and Instagram ads for social media recruitment. Facebook and Instagram ads run through the same platform and ad sets; thus, content and other targeting features are identical. To recruit a broad sample that represented a range of LGBT subgroups, we used three ad sets with different images

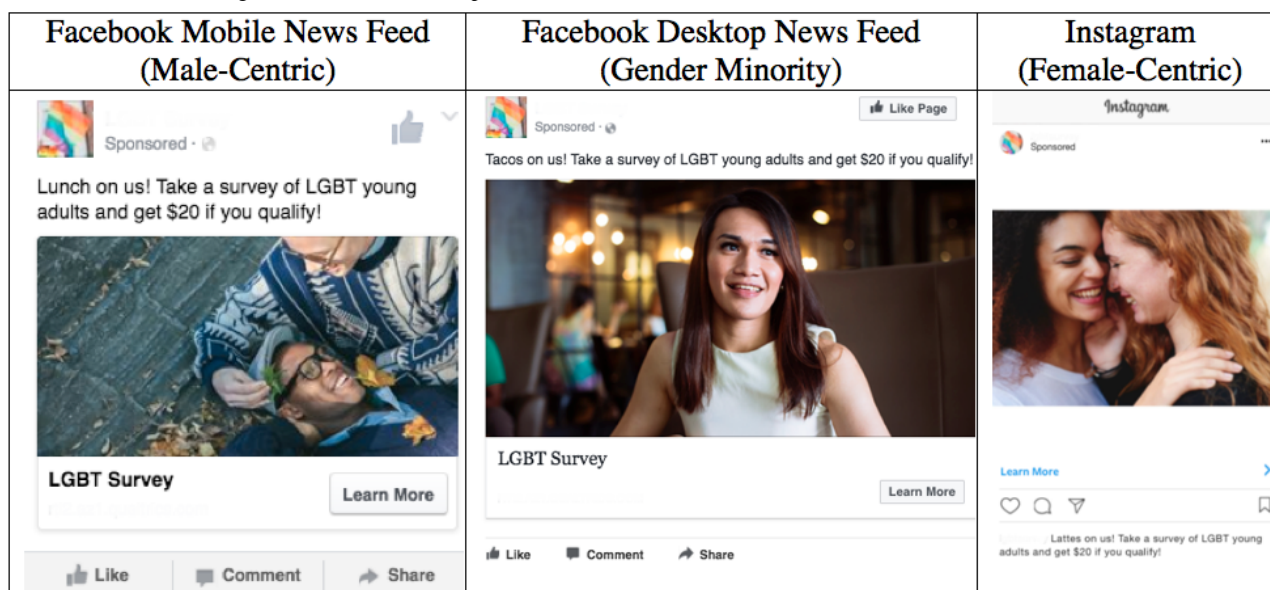
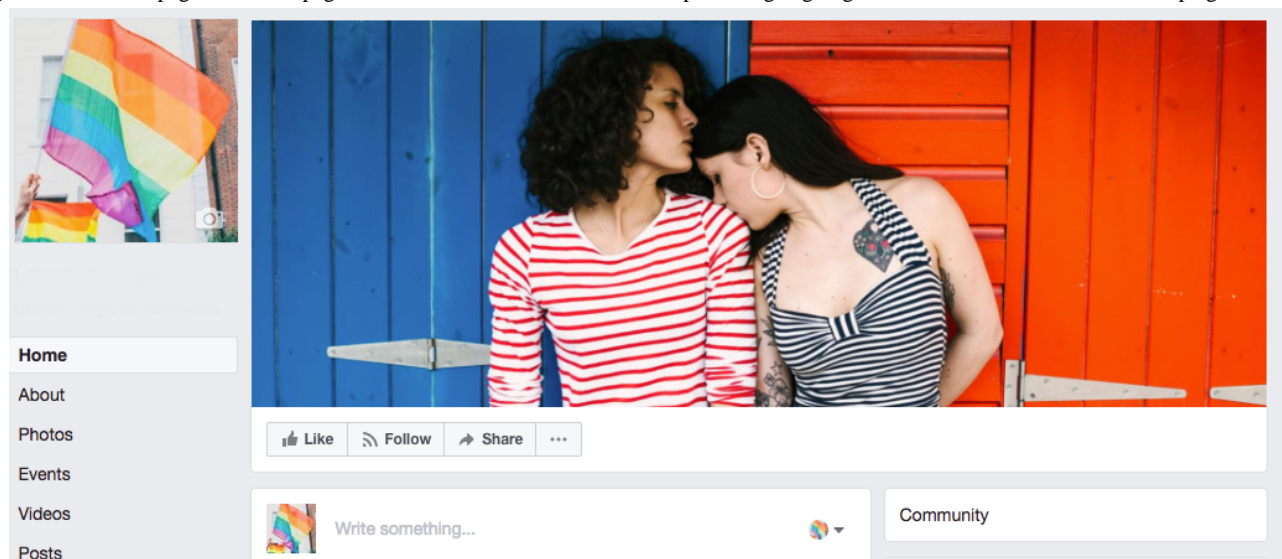
and targeting criteria. The first was a male-centric ad set targeted to men whose relationship interest was in men or men and women. The second was a female-centric ad set targeted to women whose relationship interest was in women or women and men. The third ad set targeted gender minorities (eg, transgender and genderqueer) and used keywords representing transgender and other gender minority celebrities, historic figures, causes (eg, Transgender Student Rights), and outreach groups. Several of the study’s authors identified keywords for the gender minority ad set through formative research conducted during the development stages of the media campaign content with focus groups of LGBT young adults. We outline targeting strategies unique to each of the ad sets in Table 1. All ad sets targeted English-speaking young adults, aged 18 to 24 years, and who live in one of the 24 evaluation DMAs.

Facebook and Instagram ads consisted of a brief text description of the survey and incentive amount for qualifying participants (eg, “Breakfast on us! Take a survey of LGBT young adults and get \$20 if you qualify!”), a reference to the Facebook page or Instagram account associated with the ads, and an image. Figure 1 shows sample Facebook and Instagram ads for male-centric, gender minority, and female-centric campaigns. We crafted text descriptions for all ad sets to appeal to the target audience of young adults, referencing goods and services that young adults would likely want (and realistically be able) to purchase with the US \$20 gift card that they would receive if they qualified and completed the survey (eg, “Tacos on us! ...”; “Cupcakes on us! ...”; “Treat yo self! ...”; “Burgers on us! ...”; and “Lattes on us! ...”). We chose images to represent the specific target audience for each of the three ad sets, with the male-centric ads including images of male young adult couples and an LGBT pride flag, female-centric ads including images of female young adult couples and an LGBT pride flag, and gender minority ads including images of transgender young adults. We also created an Instagram account that was associated with Instagram recruitment ads. Images from all ad sets were included in the Facebook page and Instagram account for the study.

To distribute Facebook ads, Facebook requires that a Facebook page be associated with ads (see Figure 2). We created a Facebook page to use for the study and associated it with Facebook recruitment ads.

Table 1. Facebook and Instagram targeting criteria.

Facebook or Instagram targeting strategy	Targeting criteria
Male-centric	Interested in <ul style="list-style-type: none">MenMen and women
Female-centric	Interested in <ul style="list-style-type: none">WomenWomen and men
Gender minority	Keywords: Against Me!, Caitlyn Jenner, Chaz Bono, Fallon Fox, Janet Mock, Jenna Talackova, Laura Jane Grace, Laverne Cox, Lea T, Lili Elbe, National Center for Transgender Equality, transgender, Transgender Law Center, Transgender Student Rights, Transgender youth, Wendy Carlos

Figure 1. Facebook and Instagram advertisement examples.**Figure 2.** Facebook page. Facebook page name has been removed to avoid compromising ongoing waves of data collection for the campaign evaluation.

Ads were displayed on desktop and laptop computers (as sidebar or Facebook Desktop News Feed ads), smartphones (as mobile Facebook News Feed ads), on Instagram (via the mobile Instagram app), and third-party apps in which Facebook places ads.

Dependent Variable

The dependent variable in this study was recruitment method (in-person intercept vs social media) for number of completed surveys.

Procedure

Intercept Recruitment

We fielded the baseline intercept recruitment from February 2016 to May 2016. Trained field staff visited the potential venues to confirm the venue's popularity with the target audience and, as appropriate, gain permission to conduct the intercept study there. Field supervisors followed up with recruited venues via email to confirm arrangements and

schedules. Within each DMA, two to five local data collectors attended a 5-hour in-person training before collecting data. Once trained, groups of two to five interviewers visited the recruited venues at the agreed-upon date and time. We designated one of the interviewers as the recruiter. Interviewers intercepted potential participants at venues where the recruiter approached patrons who appeared to be in the target age range and used a standardized script to introduce themselves, describe the screening process and incentive amount, prescreen patrons for age eligibility (ie, verify that participants were younger than 25 years), and ask patrons who stated they were in the age range of 18 to 24 years to complete the 5-min, self-administered screener on a tablet. To promote data quality, the trainer accompanied the data collectors on the first 2 to 5 nights of data collection to monitor compliance with data collection protocols and provide additional on-the-job training when needed. We also used global positioning system (GPS) coordinates and timing data captured within the screeners to identify screeners of questionable authenticity (ie, screeners taken at times and in

locations that did not align with the times and locations of data collection, suggesting interviewer misrepresentation). The recruiter was responsible for maintaining tallies of the outcome of this contact (ie, refused, not aged 18-24 years, and agreed to participate).

As patrons agreed, a data collector helped the patron launch the screener to determine eligibility for participating in the main survey that included questions about age, home zip code, LGBT identity, and cigarette use. Data collectors provided each intercept respondent US \$10 in cash for completing the screener and provided respondents who screened as eligible with a study information card with details about next steps. For those who screened as eligible, within 2 days an invitation was sent via SMS text message (short message service, SMS) or email (based on the participant's stated preference) to complete the full 30-min Web survey. Participants who clicked on the Web link for this survey were first directed to a brief consent form. Those who consented completed the main survey that included questions about demographics; tobacco, alcohol, and social media use; participation in and identification with LGBT culture; and tobacco-related knowledge, attitudes, and beliefs. Those who completed the survey received a US \$20 digital gift card with a US \$5 bonus (total of US \$25) for completing the full survey within 2 days of receiving the first invitation. Respondents who did not respond to the first invitation received up to three additional prompts, spaced every other day.

Social Media Recruitment

We conducted the male- and female-targeted social media recruitment during 1 week in March 2016 and the gender minority recruitment over 5 days from April 2016 to May 2016. We made first contact with potential participants via Facebook and Instagram ads. Participants clicked on ads that sent them to a link for the same screening instrument that intercept respondents completed. Social media participants did not receive an incentive for completing the screener. Eligible participants, identified via responses to the screener, proceeded directly to the same consent form provided to intercept participants. Consenting participants completed the same survey as intercept participants. Participants recruited via social media received a US \$20 digital gift card for their participation. We did not provide the US \$5 bonus to social media respondents because eligible participants proceeded directly to the main survey from the screener.

Predictor Variables

Independent variables were age; LGBT identity—lesbian (ie, cisgender lesbian or gay women), gay (ie, cisgender gay men), bisexual (ie, cisgender bisexual men and women), gender minority (ie, transgender, genderqueer, and gender-variant men and women), other sexual minority (eg, pansexual, omnisexual, and trisexual)—education; race or ethnicity; past 30-day cigarette use; past 30-day alcohol use; past-year pride event attendance; and social media use (ie, Facebook and Instagram use frequency). Only participants who are not gender minorities or other sexual minorities were grouped by their sexual identity.

Statistical Analysis

First, we compared the two recruitment strategies on the total cost and time efficiency of data collection. Data on cost of recruitment were available only as a proportion of total recruitment costs for each recruitment method; thus, our cost comparisons are limited to simple descriptive comparisons of the proportion of total cost of recruitment between intercept and social media.

To ascertain which of the two recruitment methods was more time-efficient in recruiting participants, we conducted a Poisson regression on the count of people who completed the full survey, with weeks required to recruit the sample of completed surveys for each recruitment method included as an offset variable and method of recruitment as the predictor variable [35,36]. The inclusion of weeks to survey completion as an offset variable allowed us to calculate recruitment efficiency as an incidence rate ratio (IRR) for weeks required for each recruitment method to recruit a single participant.

We then compared intercept and social media data collection on the quality of data collected and their effectiveness in identifying eligible participants. We cleaned the raw dataset to remove low-quality (ie, non-US-based internet protocol (IP) addresses and IP addresses known to be associated with malicious software or services) and duplicate responses from the social media data (ie, multiple responses from the same IP address in a small window of time and responses associated with an email address that is at least 80% similar to an email address already associated with a completed Web survey) and interviewer misrepresentation from the intercept data (ie, generation of fake responses as detected by GPS coordinates and timing data associated with completed screeners). We used *t* tests to compare the percentage of data lost because of low-quality or duplicate responses from social media with the percentage of data lost because of interviewer misrepresentation from intercept. We used a series of unpaired sample means tests to compare the percentage of people from social media vs intercept who completed screeners, were eligible, and completed the baseline survey.

Third, we compared the characteristics of participants (see the predictor variables described previously) recruited via the two methods. We used descriptive statistics (means, frequencies, and percentages) to describe the sample characteristics for each recruitment method. We then conducted bivariate analyses to determine differences between participants recruited via intercept vs social media for each of the predictor variables. We created the final multivariate model using predictor variables that we found to be related to recruitment method in bivariate analyses (at the $P < .25$ level following methods from Hosmer and Lemeshow [41]). The final model was a multivariate logistic regression with recruitment method as the outcome variable and the following predictor variables: age, education, race or ethnicity, LGBT identity, past 30-day cigarette and alcohol use, past-year pride event attendance, and Facebook and Instagram use frequency. Analyses were run in Stata 14 (StataCorp, LLC).

Results

Intercept Recruitment Screener Completion Rate

For intercept recruitment, we approached 9552 individuals who appeared to be within the eligible age range in venues and asked them to complete the screener. Of those asked, 7375 completed the screener, resulting in a 77.21% screener completion rate among those approached.

Social Media Recruitment Advertisement Performance and Screener Completion Rate

The Facebook and Instagram ads used to recruit participants reached a total of 324,959 individual users (exposed to ads at least once): 81,312 with female-centric ads, 44,802 with male-centric ads, and 204,614 with gender minority ads. Ads resulted in 7249 total clicks, with 2225 clicks on female-centric ads (2225/81,312, 2.74% of people exposed to ads, clicked on links), 1558 clicks on male-centric ads (1558/44,802, 3.48% of people exposed to ads, clicked on links), and 3466 clicks on gender minority ads (3466/204,614, 1.69% of people exposed to ads, clicked on links). It is important to note that because gender minority participants are a particularly hard-to-reach subpopulation within the LGBT young adult population, we devoted a larger budget and more run time to gender minority ads to generate a more diverse sample.

Of social media respondents who clicked on links, 6611 completed the screener, resulting in a 91.20% (6611/7249) screener completion rate for people who clicked on ads. Due to privacy features on the ad platforms, we cannot tie link clicks from specific ad sets to completed screeners; thus, the screener completion rate could only be generated for ad sets in combination.

Cost and Time Efficiency Comparisons Between Intercept and Social Media Recruitment

Descriptive comparisons of recruitment costs between recruitment methods show that social media recruitment is less expensive than intercept recruitment, with social recruitment making up just 2.2% of total recruitment costs and intercept recruitment making up 97.8% of total recruitment costs. This substantial difference is largely because of the large number of resources required to conduct intercept recruitment, as costs include labor (for staff training and time in the field recruiting participants), mileage for traveling to and from recruitment venues, screener incentives, and miscellaneous expenses (ie, parking and cover costs for venue entry), whereas the only cost for social media recruitment is the cost of placing ads on Facebook and Instagram.

Social media recruitment is also more time-efficient than intercept recruitment, as we found that the IRR for time to survey completion was 3.31 (95% CI 3.11-3.52) times faster for social media participants than intercept participants for the full data collection period ($P<.001$). [Figure 3](#) illustrates the number of surveys completed for each recruitment method by week of data collection. Time required for intercept recruitment includes training, travel time to and from recruitment venues,

and time spent recruiting participants in venues, whereas social media recruitment time includes the total number of weeks that ads were run to recruit participants. As is the case with cost, these large differences in time efficiency are attributed to additional time required for intercept data collection.

Data Quality Comparisons Between Intercept and Social Media Recruitment

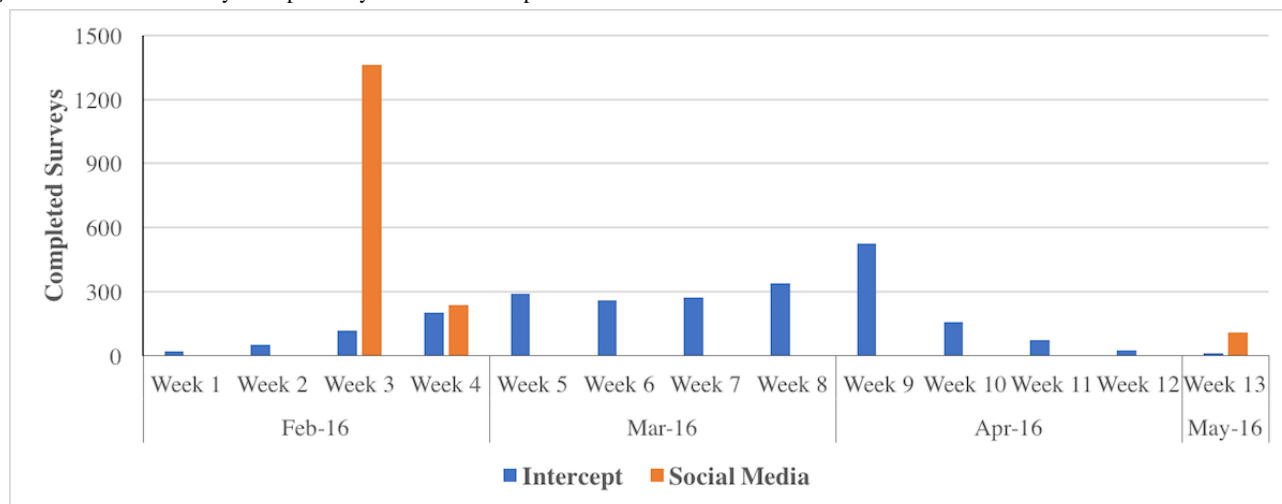
Social media recruitment was more vulnerable to data quality issues than intercept. During data cleaning, we dropped a significantly larger percentage of social media respondents because of low-quality and duplicate responses (374/4446, 8.41%) than intercept respondents dropped because of interviewer misrepresentation (15/4446, 0.34%; $P<.001$). Denominators represent the sample N before we dropped cases from the analytic sample (denominator = analytic N + N dropped for low-quality or duplicate responses + N dropped for interviewer misrepresentation).

Completed Screeners, Eligibility, and Survey Completion

Across the full sample, more than half of the people who completed screeners were eligible to participate in the main survey (7965/13986, 56.95%), and half of eligible participants (ie, young adults aged 18-24 years who self-identified as LGBT and reported living in a zip code in one of the 24 study DMAs) completed the main survey (4057/7965, 50.93%). A significantly larger percentage of participants who completed the screener via intercept (4608/7375, 62.48%) were eligible to complete the main survey than participants recruited via social media (3357/6611, 50.78%; $P<.001$). The proportion of eligible participants who completed the survey did not differ between recruitment methods. Results are presented in [Table 2](#).

Sample Characteristics

We describe the sample in [Table 3](#). We recruited more than half of the sample via intercept (2348/4057, 57.88%). The mean age of the sample was 22 years (SD 1.87), with participants recruited via intercept being significantly older than those recruited via social media ($P<.001$). Participants who self-identified as gay men made up the largest proportion of the overall sample (1822/4057, 44.91%), followed by lesbian or gay women (882/4057, 21.74%), bisexual men (219/4057, 5.40%) and women (639/4057, 15.75%), gender minorities (342/4057, 8.43%), and other sexual minorities (152/4057, 3.75%). We recruited a larger percentage of people who self-identified as gay men via intercept than social media ($P<.001$) and a larger percentage of people who self-identified as lesbian or gay women, gender minorities, and other sexual minorities via social media ($P<.001$). The majority of the sample reported having some college education (2049/3976, 51.53%), followed by a high school education or less (1040/3976, 26.63%), and a college degree or greater (887/3976, 22.31%). Participants who reported attending high school or less ($P<.001$) or some college ($P=.005$) were more likely to be recruited via social media than intercept, and those who reported having a college education or greater were more likely to be recruited via intercept ($P<.001$).

Figure 3. Number of surveys completed by week for intercept and social media recruitment.**Table 2.** Eligibility and survey completion by recruitment method. The denominator for each row percentage is from the row above. Comparisons between counts for *eligible* and *completed survey* rows are horizontal only.

Stage of completion	Intercept	Social media	P value
Completed screener, n	7375	6611	N/A ^a
Eligible, n (%)	4608 (62.48)	3357 (50.78)	<.001
Completed survey, n (%)	2348 (50.95)	1709 (50.90)	.97

^aN/A: not applicable.

The sample was primarily white, non-Hispanic (1856/4057, 45.75%); followed by Hispanic (1211/4057, 29.85%), other or multiracial, non-Hispanic (587/4057, 14.47%); and black, non-Hispanic (403/4057, 9.93%). Black, non-Hispanic ($P=.002$) and Hispanic participants ($P=.01$) were more likely to be recruited via intercept, whereas white, non-Hispanic participants were more likely to be recruited via social media ($P=.001$).

On average, participants reported smoking cigarettes on 14 of the past 30 days (SD 11.73). Past 30-day smoking was higher among participants recruited via intercept ($P=.02$). About one-quarter of participants reported using alcohol on 3 to 5 of the past 30 days (904/3487, 25.92%), followed by 6 to 9 (849/3487, 24.35%), 10 to 19 (764/3487, 21.91%), 1 or 2 (671/3487, 19.24%), 20 to 29 (249/3487, 7.14%), and 30 of the past 30 days (50/3487, 1.43%). Participants who consumed alcohol on 1 or 2 ($P<.001$) or 3 to 5 of the past 30 days ($P=.05$) were more likely to be recruited via social media, whereas those who consumed alcohol on 10 to 19 or 20 to 29 of the past 30 days were more likely to be recruited via intercept ($P<.001$). The majority of participants reported past-year pride event attendance (2591/3092, 83.80%). A larger percentage of participants recruited via social media reported past-year pride event attendance ($P<.001$).

Social media use was high across the overall sample, with more than half of participants reporting Facebook use several times a day (2666/3943, 67.61%) and half of participants reporting Instagram use several times a day (1952/3942, 49.52%). Among those who reported using Facebook about once a day, a larger percentage were recruited via social media ($P=.02$). Among

those who reported using Facebook less than once a day, a larger percentage were recruited via intercept ($P=.006$). Finally, among participants who reported using Instagram less than once a day, a larger percentage were recruited via social media ($P=.02$).

Bivariate Analyses

Bivariate analyses (reported in text only) revealed that, as age increased, odds of recruitment via social media decreased (odds ratio, OR 0.73, 95% CI 0.71-0.76, $P<.001$). Lesbian or gay women (OR 1.80, 95% CI 1.53-2.12, $P<.001$), bisexual men and women (OR 1.43, 95% CI 1.22-1.69, $P<.001$), gender minorities (OR 2.08, 95% CI 1.65-2.62, $P<.001$), and other sexual minorities (OR 3.64, 95% CI 2.57-5.16, $P<.001$) were more likely than gay men to be recruited via social media. Participants with a high school education or less (OR 2.08, 95% CI 1.73-2.51, $P<.001$) and with some college education (OR 1.80, 95% CI 1.52-2.13, $P<.001$) were more likely than those with at least a college degree to be recruited via social media. Hispanic (OR 0.79, 95% CI 0.69-0.92, $P=.002$) and black, non-Hispanic participants (OR 0.66, 95% CI 0.53-0.83, $P<.001$) were less likely than white, non-Hispanic participants to be recruited via social media. Participants reporting past 30-day smoking (OR 0.84, 95% CI 0.74-0.95, $P=.006$) and drinking alcohol (OR 0.25, 95% CI 0.20-0.31, $P<.001$) were less likely to be recruited via social media. Participants who had attended a pride event in the past year were more likely to be recruited via social media (OR 1.48, 95% CI 1.21-1.81, $P<.001$), as were those who used Facebook at least once a day (OR 1.28, 95% CI 1.08-1.53, $P=.005$). In contrast, participants who reported using Instagram at least once a day were less likely to be recruited via social media (OR 0.85, 95% CI 0.74-0.97, $P=.01$).

Table 3. Sample characteristics.

Characteristic	Total sample (N=4057)	Social media (N=1709)	Intercept (N=2348)	P value
Age (years), mean (SD)	21.54 (1.87)	20.94 (1.94)	21.98 (1.69)	<.001
LGBT^a identity, n (%)				
Gay men	1822 (44.91)	630 (36.86)	1191 (50.75)	<.001
Lesbian or gay women	882 (21.74)	430 (25.16)	453 (19.30)	<.001
Bisexual men and women	858 (21.14)	370 (21.65)	488 (20.79)	.51
Gender minorities	342 (8.43)	179 (10.47)	163 (6.94)	<.001
Other sexual minorities	152 (3.75)	100 (5.85)	52 (2.21)	<.001
Education, n (%)				
High school or less	1040 (26.63)	489 (29.76)	551 (23.62)	<.001
Some college	2049 (51.53)	889 (54.11)	1160 (49.72)	.005
College plus	887 (22.31)	265 (16.13)	622 (26.67)	<.001
Race or ethnicity, n (%)				
White, non-Hispanic	1856 (45.75)	832 (48.68)	1024 (43.61)	.001
Black, non-Hispanic	403 (9.93)	141 (8.25)	262 (11.16)	.002
Hispanic	1211 (29.85)	475 (27.79)	736 (31.35)	.01
Other or multiracial, non-Hispanic	587 (14.47)	261 (15.27)	326 (13.88)	.23
Past 30-day cigarette use (N=1833), mean (SD)	13.54 (11.73)	12.73 (11.33)	14.07 (11.96)	.02
Past 30-day alcohol use, n (%)				
1 or 2 days	671 (19.24)	343 (26.61)	328 (14.92)	<.001
3-5 days	904 (25.92)	366 (28.39)	538 (24.48)	.05
6-9 days	849 (24.35)	302 (23.43)	547 (24.88)	.33
10-19 days	764 (21.91)	199 (15.44)	565 (25.71)	<.001
20-29 days	249 (7.14)	64 (4.97)	185 (8.42)	<.001
30 days	50 (1.43)	15 (1.16)	35 (1.59)	.28
Past-year pride event attendance, n (%)	2591 (83.80)	1113 (86.81)	1478 (81.66)	<.001
Facebook use frequency, n (%)				
Several times a day	2666 (67.61)	1095 (67.97)	1571 (67.37)	.78
About once a day	639 (16.20)	287 (17.82)	352 (15.09)	.02
Less than once a day	638 (16.18)	229 (14.21)	409 (17.54)	.006
Instagram use frequency, n (%)				
Several times a day	1952 (49.52)	776 (48.17)	1176 (50.45)	.16
About once a day	563 (14.28)	215 (13.35)	348 (14.93)	.17
Less than once a day	1427 (36.20)	620 (38.49)	807 (34.62)	.02

^aLGBT: lesbian, gay, bisexual, and transgender.

Table 4. Multivariate logistic regressions of lesbian, gay, bisexual, and transgender (LGBT) young adults recruited via social media (vs intercept). Predictors include variables related to the recruitment methods in the bivariate analyses ($P<.25$). Variables for past 30-day cigarette and alcohol use are dichotomized to any past 30-day use vs no past 30-day use (reference category, REF). Social media use frequency variables are dichotomized to at least once a day vs less than once a day (REF). Analytic N=2945 (social media N=1183, intercept N=1762).

Variable	AOR ^a	95% CI	P value
Age	0.76	0.72-0.80	<.001
LGBT^b identity			
Gay men	REF ^c	REF	REF
Lesbian or gay women	1.88	1.54-2.29	<.001
Bisexual men and women	1.46	1.17-1.82	.001
Gender minorities	1.68	1.26-2.25	<.001
Other sexual minorities	2.48	1.62-3.80	<.001
Education			
High school or less	1.07	0.83-1.40	.60
Some college	1.27	1.03-1.56	.03
College plus	REF	REF	REF
Race or ethnicity			
White, non-Hispanic	REF	REF	REF
Black, non-Hispanic	0.76	0.58-1.01	.05
Hispanic	0.73	0.61-0.89	.001
Other or multiracial, non-Hispanic	0.70	0.54-0.90	.006
Past 30-day cigarette use	0.94	0.80-1.10	.42
Past 30-day alcohol use	0.33	0.24-0.44	<.001
Past-year pride event attendance	1.31	1.06-1.64	.02
Facebook use at least once a day	1.43	1.14-1.80	.002
Instagram use at least once a day	0.73	0.62-0.86	<.001

^aAOR: adjusted odds ratio.

^bLGBT: lesbian, gay, bisexual, and transgender.

^cREF: Reference category.

Logistic Regression Analyses

We present results from the final multivariate logistic regression in Table 4. As age increased, odds of recruitment via social media decreased (AOR 0.76, 95% CI 0.72-0.80, $P<.001$). Lesbian or gay women (AOR 1.88, 95% CI 1.54-2.29, $P<.001$), bisexual men and women (AOR 1.46, 95% CI 1.17-1.82, $P=.001$), gender minorities (AOR 1.68, 95% CI 1.26-2.25, $P<.001$), and other sexual minorities (AOR 2.48, 95% CI 1.62-3.80, $P<.001$) were more likely than gay men to be recruited via social media. Hispanic (AOR 0.73, 95% CI 0.61-0.89, $P=.001$); black, non-Hispanic (AOR 0.76, 95% CI 0.58-1.01, $P=.05$); and other or multiracial, non-Hispanic participants (AOR 0.70, 95% CI 0.54-0.90, $P=.006$) were less likely than white, non-Hispanic participants to be recruited via social media. Participants with some college education (AOR 1.27, 95% CI 1.03-1.56, $P=.03$) were more likely than those with at least a college degree to be recruited via social media.

Participants reporting past 30-day alcohol use were less likely to be recruited via social media (AOR 0.33, 95% CI 0.24-0.44,

$P<.001$). Past 30-day smoking was not related to the likelihood of being recruited via social media vs intercept. Participants who reported past-year pride event attendance were more likely to be recruited via social media (AOR 1.31, 95% CI 1.06-1.64, $P<.05$) as were those who used Facebook at least once a day (AOR 1.43, 95% CI 1.14-1.80, $P=.002$). Participants who reported using Instagram at least once a day were less likely to be recruited via social media (AOR 0.73, 95% CI 0.62-0.86, $P<.001$).

Discussion

Principal Findings

Overall, our findings indicate that innovative recruitment methods that reached hard-to-reach populations in the virtual and physical spaces where they spend time were an effective means of recruiting LGBT young adults for FDA's *This Free Life* campaign evaluation. Social media participants were younger and less educated (ie, more likely to report some college education than having a college degree or greater) compared with participants recruited via intercept. Social media

participants were more likely to self-identify as lesbian or gay women, bisexual men or women, gender minorities, or other sexual minorities than as gay men compared with intercept participants. Social media participants were also more likely to be white, non-Hispanic than racial or ethnic minorities compared with intercept participants. Social media participants were more likely to have attended a pride event in the past year than were intercept participants. Finally, participants reporting past 30-day alcohol use were less likely to be recruited via social media than intercept.

Findings specific to age, LGBT identity, education, and alcohol consumption were not unexpected for several reasons. First, LGBT social venues and events tend to disproportionately cater to gay men in comparison with other LGBT subgroups (eg, lesbian or gay women). Many LGBT social venues are also restricted to those aged 21 years and older because they serve alcoholic beverages. Thus, attendance is not only limited to older individuals who are of legal drinking age but may also be limited to those with higher levels of education who are more likely to be able to afford drinking in bars and nightclubs. The highly specific targeting tools provided by social media platforms enabled us to recruit particularly hard-to-reach subgroups of LGBT young adults (eg, gender and other sexual minorities). In combination, our findings suggest that the unique features of the two recruitment methods complemented one another, allowing us to recruit a more balanced population of LGBT young adults. For example, the targeting tools available via social media advertising platforms allowed us to recruit a broader age range of LGBT young adults from hard-to-reach LGBT subgroups, whereas the tendency for LGBT social venues to be 21 years-and-over bars and nightclubs allowed us to recruit higher-risk LGBT young adults (ie, those who consume more alcohol and smoke more cigarettes).

One unexpected finding was that social media participants were more likely to be white, non-Hispanic than racial or ethnic minorities (ie, Hispanic, black, non-Hispanic) compared with intercept participants. It is unclear why this difference emerged given that white and non-white adults' level of social media use is about equal [42]. White, non-Hispanic individuals have higher levels of at-home broadband internet access (82%) than black individuals (74%) [43], suggesting the possibility that a larger number of white, non-Hispanic participants recruited via social media ads completed surveys because they had a greater opportunity to click on social media ads and immediately proceed to completing the 30-min survey instrument while using social media at home. Because fewer racial or ethnic minorities have at-home broadband internet, they may have been more likely to see the social media ads while using smartphones on-the-go (ie, not at home), which may not have been an optimal time for completing the 30-min survey. It is unsurprising that this was not the case for intercept as eligible participants were emailed the survey link to complete at their convenience.

Although social media participants were more likely to use Facebook at least once a day, one counterintuitive finding was that social media participants were less likely to use Instagram at least once a day compared with intercept participants. It is important to note that, at the time of recruitment, Instagram advertising was newly available to researchers and has not yet

been used extensively for participant recruitment. Although the reasons for this finding are unclear, it suggests that there may be important differences in recruitment methods that are most effective for recruiting Facebook and Instagram users. Furthermore, these differences may vary depending on frequency of use of each platform. We will explore this possibility in future research.

Beyond recruiting a diverse sample of LGBT young adults, both recruitment methods resulted in high eligibility rates among participants who completed the screening instrument. Eligibility was higher among intercept than social media participants, suggesting that intercept interviews are more likely to identify eligible members of the target population. Social media ads may reach a wider population beyond those who are LGBT, whereas more LGBT social venue attendees may be LGBT. Survey completion rates among eligible participants were similar between recruitment methods. This finding is promising for social media recruitment as a data collection method—previous research has shown that collecting data entirely via the Web leads to lower levels of participant accountability, which results from higher levels of psychological distance between researcher and participant in Web-based studies (compared with studies involving some level of face-to-face contact between researcher and participant) [44-46]. Rather than affecting completion rates, this psychological distance may instead have played out in terms of data quality. Data collected via social media were more vulnerable to low-quality and duplicate responses from participants trying to complete the survey more than once for additional incentives. In comparison, we needed to throw out only a nominal percentage of intercept surveys because of interviewer misrepresentation (ie, fake responses generated by interviewers).

Although intercept recruitment resulted in higher eligibility and had fewer data quality issues, from a practical perspective, social media recruitment was significantly more time efficient than intercept. From a proportion of recruitment costs perspective, social recruitment was also less expensive than intercept. We completed social media data collection in all markets in less than 2 weeks for the nominal cost of posting ads. In contrast, for manageability considerations, we launched intercept collection in markets on a rolling basis over the course of 7 weeks with two to five markets being launched in any given week. Due to wide variation by market in the volume of LGBT young adults present at local LGBT venues, intercept data collection lasted anywhere from 1 week to 9 weeks within each market. Slow or lengthy data collection in some markets reduced data collector morale and retention, requiring the expense of training additional local data collectors or temporarily relocating data collectors from other areas. From a cost perspective, intercept methods involve wages and other expenses associated with data collector time spent intercepting respondents and the expense of recruiting, training, and managing those data collectors. Furthermore, intercept respondents were provided with an additional US \$10 cash incentive for taking the eligibility screener and a US \$5 bonus for completing the main survey within 2 days of receiving the first invitation. Social media respondents were not offered additional incentives

because they completed the screener after clicking on ads, and eligible participants proceeded directly to the main survey.

Limitations

Although our research provides important insights about how samples can be recruited via innovative methods, we acknowledge several important limitations. First, samples that were recruited both via social media and intercept were convenience, nonprobability samples. Second, it is possible that participants were exposed to both recruitment methods more than once, which may have influenced their decision to participate in the study and was not measured in this study. Third, we were unable to conduct significance tests to compare costs between social media and intercept recruitment because cost data was available only as a proportion of total recruitment cost for each recruitment method. Fourth, the higher monetary incentive offered to participants recruited via intercept (ie, US \$10 incentive for completing the screener and potential US \$5 early survey completion bonus for main survey) was a confounding factor that may have influenced recruitment rates between methods. A final limitation was the difference in the method for approaching potential participants to complete the screening instrument via social media vs intercept. Intercept respondents were approached by data collectors to complete the screening instrument in person and may have experienced more perceived pressure to complete the screener than those recruited via social media who were shown ads and could choose whether or not to click on ads without being physically monitored by a third party.

Comparison With Prior Work

Our research marks the recruitment of one of the largest samples of LGBT young adults to date. This research provides important contributions to the literature on using novel methods to recruit hard-to-reach populations; LGBT young adults in particular. Although previous studies have used social media to recruit members of the LGBT community, such as gay men [33,38-39], gay and bisexual male youth [30], and transgender women [47], our study demonstrates that social media can be used to recruit large numbers of particularly hard-to-reach and underrepresented LGBT subgroup members (eg, lesbian or gay women, bisexual men and women, gender and other sexual minorities).

In a similar vein, our research shows that large numbers of LGBT young adults in a number of different regions of the United States can be recruited in-person in social venues. These findings provide important contributions to the existing literature that has shown that these methods can be used to recruit LGBT

young adults in bars and nightclubs in a single market [23] and young adults who are at a higher risk for smoking and alcohol use in a number of markets in the United States [24-28].

Our research also shows that a large and diverse sample of LGBT young adults can be recruited using Facebook and Instagram ads. To the best of our knowledge, all of the published literature demonstrating the effectiveness of using social media for participant recruitment has focused on using Facebook to recruit LGBT samples [33,38-39,47]. Instagram ads are somewhat newly available to researchers (since September 2015) [48] and show promise for recruiting hard-to-reach populations such as young adults, who have much higher representation on Instagram (59% of adults aged 18-29 years) than adults over the age of 30 years (33% of adults aged 30-49 years, 18% of adults aged 50-64 years, and 8% of adults aged 65+ years) [49]. Indeed, Instagram ads have rapidly evolved and now offer similar capabilities as Facebook ads because they use the same platform. These capabilities make Instagram ads a seamless tool to use alongside Facebook ads for recruitment.

Finally, this study shares explicit methodological details regarding our development of social media ads and strategy. This information may be particularly useful for researchers who seek to implement these tools in their own studies. Few published studies provide this level of detail on their social media recruitment methods [32].

Conclusions

Novel methods that reach hard-to-reach populations such as LGBT young adults where they frequently spend time are effective participant recruitment strategies. We recruited LGBT young adults in LGBT social venues and via Facebook and Instagram ads. Using these methods in combination, we recruited a more diverse sample of LGBT young adults from a broader range of LGBT identities, race or ethnicities, ages, and education levels than we could have using either method in isolation. Importantly, social media ads provided enhanced access to particularly hard-to-reach subpopulations of LGBT young adults (ie, bisexual, gender, and sexual minorities) who were less easily accessed via intercept recruitment. Although social media data collection is a more efficient and inexpensive recruitment method, it is more subject to data quality issues than intercept data collection. Together, these methods enabled recruitment of one of the largest known LGBT young adult samples, suggesting their promise for recruiting hard-to-reach populations.

Acknowledgments

The authors would like to thank Ashley Smith, Shaola Abagat, Gem Benozza, and April Brubach of FDA's CTP for developing the *This Free Life* campaign and overall strategy, which was integral to informing the campaign evaluation design and this study. They would also like to thank Brandon S Tate, Mayo Djakaria, Dana E Wagner, Michelle Bellon, and Jeffrey W Jordan of Rescue Agency for developing the *This Free Life* campaign and overall strategy and for their guidance in developing the social media recruitment strategy for this study.

This publication represents the views of the authors and does not represent FDA CTP's position or policy.

Authors' Contributions

JG, KFW, MF, LH, JD, and TNA conceptualized the study. JG, MF, LH, JD, and TNA developed the survey measures. JG, MF, and KFW led the study implementation. JG developed social media methods and led social media data collection. KFW developed the intercept recruitment methods and led the intercept data collection. JG, LF, and IA led the data analysis. JG, LF, IA, KFW, MF, and LH contributed to the interpretation of findings. JG led the writing of the manuscript. LF, IA, KFW, and LH assisted in writing the manuscript. All co-authors provided critical review of the manuscript.

Conflicts of Interest

None declared.

References

- Center for Tobacco Products (CTP). FDA. 2017. The real cost campaign URL: <https://www.fda.gov/tobaccoproducts/publichealtheducation/publiceducationcampaigns/therealcostcampaign/default.htm> [accessed 2017-10-09] [WebCite Cache ID 6sxt5awpe]
- Center for Tobacco Products (CTP). FDA. 2017. Fresh empire campaign URL: <https://www.fda.gov/TobaccoProducts/PublicHealthEducation/PublicEducationCampaigns/ThisFreeLifeCampaign/default.htm> [accessed 2017-10-09] [WebCite Cache ID 6sxsmCfck]
- Center for Tobacco Products (CTP). FDA. 2016. This free life campaign URL: <https://www.fda.gov/TobaccoProducts/PublicHealthEducation/PublicEducationCampaigns/ThisFreeLifeCampaign/default.htm> [accessed 2017-10-09] [WebCite Cache ID 6sxtC3bCs]
- Gates G, Newport F. Gallup. Washington, DC: Gallup; 2012 Oct 18. Special report 3.4% of US adults identify as LGBT URL: <http://www.gallup.com/poll/158066/special-report-adults-identify-lgbt.aspx> [accessed 2018-04-30] [WebCite Cache ID 6z4C7q9SO]
- US Department of Health and Human Services (USDHHS). 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/sgr50-supplement-tables.pdf> [accessed 2017-09-18] [WebCite Cache ID 6z4CrtdQw]
- Shires DA, Jaffee KD. Structural discrimination is associated With smoking status among a national sample of transgender individuals. *Nicotine Tob Res* 2016 Dec;18(6):1502-1508. [doi: [10.1093/ntr/ntv221](https://doi.org/10.1093/ntr/ntv221)] [Medline: [26438646](https://pubmed.ncbi.nlm.nih.gov/26438646/)]
- Dilley JA, Spigner C, Boysun MJ, Dent CW, Pizacani BA. Does tobacco industry marketing excessively impact lesbian, gay and bisexual communities? *Tob Control* 2008 Dec;17(6):385-390 [FREE Full text] [doi: [10.1136/tc.2007.024216](https://doi.org/10.1136/tc.2007.024216)] [Medline: [18723561](https://pubmed.ncbi.nlm.nih.gov/18723561/)]
- Agaku IT, King BA, Husten CG, Bunnell R, Ambrose BK, Hu SS, Centers for Disease Control and Prevention (CDC). Tobacco product use among adults--United States, 2012-2013. *MMWR Morb Mortal Wkly Rep* 2014 Jun 27;63(25):542-547 [FREE Full text] [Medline: [24964880](https://pubmed.ncbi.nlm.nih.gov/24964880/)]
- Lee J, Griffin G, Melvin C. Tobacco use among sexual minorities in the USA, 1987 to May 2007: a systematic review. *Tob Control* 2009 Aug;18(4):275-282. [doi: [10.1136/tc.2008.028241](https://doi.org/10.1136/tc.2008.028241)] [Medline: [19208668](https://pubmed.ncbi.nlm.nih.gov/19208668/)]
- Smith EA, Malone RE. The outing of Philip Morris: advertising tobacco to gay men. *Am J Public Health* 2003 Jun;93(6):988-993. [Medline: [12773366](https://pubmed.ncbi.nlm.nih.gov/12773366/)]
- Stevens P, Carlson LM, Hinman JM. An analysis of tobacco industry marketing to lesbian, gay, bisexual, and transgender (LGBT) populations: strategies for mainstream tobacco control and prevention. *Health Promot Pract* 2004 Jul;5(3 Suppl):129S-134S. [doi: [10.1177/1524839904264617](https://doi.org/10.1177/1524839904264617)] [Medline: [15231106](https://pubmed.ncbi.nlm.nih.gov/15231106/)]
- Smith EA, Thomson K, Offen N, Malone RE. "If you know you exist, it's just marketing poison": meanings of tobacco industry targeting in the lesbian, gay, bisexual, and transgender community. *Am J Public Health* 2008 Jun;98(6):996-1003 [FREE Full text] [doi: [10.2105/AJPH.2007.118174](https://doi.org/10.2105/AJPH.2007.118174)] [Medline: [18445800](https://pubmed.ncbi.nlm.nih.gov/18445800/)]
- Gruskin EP, Byrne KM, Altschuler A, Dibble SL. Smoking it all away: influences of stress, negative emotions, and stigma on lesbian tobacco use. *J LGBT Health Res* 2008;4(4):167-179. [Medline: [19928044](https://pubmed.ncbi.nlm.nih.gov/19928044/)]
- Hatzenbuehler ML, Slopen N, McLaughlin KA, McLaughlin KA. Stressful life events, sexual orientation, and cardiometabolic risk among young adults in the United States. *Health Psychol* 2014 Oct;33(10):1185-1194 [FREE Full text] [doi: [10.1037/hea0000126](https://doi.org/10.1037/hea0000126)] [Medline: [25133830](https://pubmed.ncbi.nlm.nih.gov/25133830/)]
- Cochran SD, Björkenstam C, Mays VM. Sexual orientation and all-cause mortality among US adults aged 18 to 59 Years, 2001-2011. *Am J Public Health* 2016 May;106(5):918-920. [doi: [10.2105/AJPH.2016.303052](https://doi.org/10.2105/AJPH.2016.303052)] [Medline: [26985610](https://pubmed.ncbi.nlm.nih.gov/26985610/)]
- Grant J, Mottet L, Tanis J. Researchgate. Washington, DC: National Center for Transgender Equality and the National Gay and Lesbian Task Force; 2010. National transgender discrimination survey report on health and health care URL: <https://tinyurl.com/ydgsbgrk> [WebCite Cache ID 6z4EEHIN4]
- Blondeel K, Say L, Chou D, Toskin I, Khosla R, Scolaro E, et al. Evidence and knowledge gaps on the disease burden in sexual and gender minorities: a review of systematic reviews. *Int J Equity Health* 2016 Jan 22;15:16 [FREE Full text] [doi: [10.1186/s12939-016-0304-1](https://doi.org/10.1186/s12939-016-0304-1)] [Medline: [26800682](https://pubmed.ncbi.nlm.nih.gov/26800682/)]

18. Conron KJ, Mimiaga MJ, Landers SJ. A population-based study of sexual orientation identity and gender differences in adult health. *Am J Public Health* 2010 Oct;100(10):1953-1960. [doi: [10.2105/AJPH.2009.174169](https://doi.org/10.2105/AJPH.2009.174169)] [Medline: [20516373](https://pubmed.ncbi.nlm.nih.gov/20516373/)]
19. Blossich J, Lee JG, Horn K. A systematic review of the aetiology of tobacco disparities for sexual minorities. *Tob Control* 2013 Mar;22(2):66-73 [FREE Full text] [doi: [10.1136/tobaccocontrol-2011-050181](https://doi.org/10.1136/tobaccocontrol-2011-050181)] [Medline: [22170335](https://pubmed.ncbi.nlm.nih.gov/22170335/)]
20. McLaughlin KA, Hatzenbuehler ML, Xuan Z, Conron KJ. Disproportionate exposure to early-life adversity and sexual orientation disparities in psychiatric morbidity. *Child Abuse Negl* 2012 Sep;36(9):645-655 [FREE Full text] [doi: [10.1016/j.chiabu.2012.07.004](https://doi.org/10.1016/j.chiabu.2012.07.004)] [Medline: [22964371](https://pubmed.ncbi.nlm.nih.gov/22964371/)]
21. Corliss HL, Rosario M, Wypij D, Wylie SA, Frazier AL, Austin SB. Sexual orientation and drug use in a longitudinal cohort study of U.S. adolescents. *Addict Behav* 2010 May;35(5):517-521 [FREE Full text] [doi: [10.1016/j.addbeh.2009.12.019](https://doi.org/10.1016/j.addbeh.2009.12.019)] [Medline: [20061091](https://pubmed.ncbi.nlm.nih.gov/20061091/)]
22. Caceres BA, Brody A, Luscombe RE, Primiano JE, Marusca P, Sitts EM, et al. A systematic review of cardiovascular disease in sexual minorities. *Am J Public Health* 2017 Apr;107(4):570. [doi: [10.2105/AJPH.2016.303630a](https://doi.org/10.2105/AJPH.2016.303630a)] [Medline: [28272943](https://pubmed.ncbi.nlm.nih.gov/28272943/)]
23. Fallin A, Neilands TB, Jordan JW, Ling PM. Secondhand smoke exposure among young adult sexual minority bar and nightclub patrons. *Am J Public Health* 2014 Feb;104(2):e148-e153 [FREE Full text] [doi: [10.2105/AJPH.2013.301657](https://doi.org/10.2105/AJPH.2013.301657)] [Medline: [24328626](https://pubmed.ncbi.nlm.nih.gov/24328626/)]
24. Guillory J, Johns M, Farley SM, Ling PM. Loose cigarette purchasing and nondaily smoking among young adult bar patrons in New York City. *Am J Public Health* 2015 Aug;105(8):e140-e147 [FREE Full text] [doi: [10.2105/AJPH.2014.302518](https://doi.org/10.2105/AJPH.2014.302518)] [Medline: [25880951](https://pubmed.ncbi.nlm.nih.gov/25880951/)]
25. Guillory J, Lisha N, Lee YO, Ling PM. Phantom smoking among young adult bar patrons. *Tob Control* 2017 Dec;26(2):153-157 [FREE Full text] [doi: [10.1136/tobaccocontrol-2015-052821](https://doi.org/10.1136/tobaccocontrol-2015-052821)] [Medline: [27048205](https://pubmed.ncbi.nlm.nih.gov/27048205/)]
26. Jiang N, Ling PM. Impact of alcohol use and bar attendance on smoking and quit attempts among young adult bar patrons. *Am J Public Health* 2013 May;103(5):e53-e61 [FREE Full text] [doi: [10.2105/AJPH.2012.301014](https://doi.org/10.2105/AJPH.2012.301014)] [Medline: [23488485](https://pubmed.ncbi.nlm.nih.gov/23488485/)]
27. Lee YO, Bahreinifar S, Ling PM. Understanding tobacco-related attitudes among college and noncollege young adult hookah and cigarette users. *J Am Coll Health* 2014;62(1):10-18 [FREE Full text] [doi: [10.1080/07448481.2013.842171](https://doi.org/10.1080/07448481.2013.842171)] [Medline: [24313692](https://pubmed.ncbi.nlm.nih.gov/24313692/)]
28. Ling PM, Lee YO, Hong J, Neilands TB, Jordan JW, Glantz SA. Social branding to decrease smoking among young adults in bars. *Am J Public Health* 2014 Apr;104(4):751-760 [FREE Full text] [doi: [10.2105/AJPH.2013.301666](https://doi.org/10.2105/AJPH.2013.301666)] [Medline: [24524502](https://pubmed.ncbi.nlm.nih.gov/24524502/)]
29. Rait MA, Prochaska JJ, Rubinstein ML. Recruitment of adolescents for a smoking study: use of traditional strategies and social media. *Transl Behav Med* 2015 Sep;5(3):254-259 [FREE Full text] [doi: [10.1007/s13142-015-0312-5](https://doi.org/10.1007/s13142-015-0312-5)] [Medline: [26327930](https://pubmed.ncbi.nlm.nih.gov/26327930/)]
30. Prescott TL, Phillips IG, DuBois LZ, Bull SS, Mustanski B, Ybarra ML. Reaching adolescent gay, bisexual, and queer men online: development and refinement of a national recruitment strategy. *J Med Internet Res* 2016 Aug 04;18(8):e200 [FREE Full text] [doi: [10.2196/jmir.5602](https://doi.org/10.2196/jmir.5602)] [Medline: [27492781](https://pubmed.ncbi.nlm.nih.gov/27492781/)]
31. 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 Feb 23;14(1):e28 [FREE Full text] [doi: [10.2196/jmir.1878](https://doi.org/10.2196/jmir.1878)] [Medline: [22360969](https://pubmed.ncbi.nlm.nih.gov/22360969/)]
32. Ramo DE, Rodriguez TM, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. *Internet Interv* 2014 Apr;1(2):58-64 [FREE Full text] [doi: [10.1016/j.invent.2014.05.001](https://doi.org/10.1016/j.invent.2014.05.001)] [Medline: [25045624](https://pubmed.ncbi.nlm.nih.gov/25045624/)]
33. Thériault N, Bi P, Hiller JE, Nor M. Use of web 2.0 to recruit Australian gay men to an online HIV/AIDS survey. *J Med Internet Res* 2012 Nov 06;14(6):e149 [FREE Full text] [doi: [10.2196/jmir.1819](https://doi.org/10.2196/jmir.1819)] [Medline: [23128646](https://pubmed.ncbi.nlm.nih.gov/23128646/)]
34. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. *J Med Internet Res* 2014 May 01;16(5):e117 [FREE Full text] [doi: [10.2196/jmir.3229](https://doi.org/10.2196/jmir.3229)] [Medline: [24784982](https://pubmed.ncbi.nlm.nih.gov/24784982/)]
35. Guillory J, Kim A, Murphy J, Bradfield B, Nonnemaker J, Hsieh Y. Comparing twitter and online panels for survey recruitment of e-cigarette users and smokers. *J Med Internet Res* 2016 Nov 15;18(11):e288 [FREE Full text] [doi: [10.2196/jmir.6326](https://doi.org/10.2196/jmir.6326)] [Medline: [27847353](https://pubmed.ncbi.nlm.nih.gov/27847353/)]
36. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. *Internet Interv* 2016 May;4:1-10. [doi: [10.1016/j.invent.2016.01.001](https://doi.org/10.1016/j.invent.2016.01.001)]
37. Loxton D, Powers J, Anderson AE, Townsend N, Harris ML, Tuckerman R, et al. Online and offline recruitment of young women for a longitudinal health survey: findings from the Australian Longitudinal Study on Women's Health 1989-95 cohort. *J Med Internet Res* 2015 May 04;17(5):e109 [FREE Full text] [doi: [10.2196/jmir.4261](https://doi.org/10.2196/jmir.4261)] [Medline: [25940876](https://pubmed.ncbi.nlm.nih.gov/25940876/)]
38. Hernandez-Romieu AC, Sullivan PS, Sanchez TH, Kelley CF, Peterson JL, Del RC, et al. The comparability of men who have sex with men recruited from venue-time-space sampling and Facebook: a cohort study. *JMIR Res Protoc* 2014 Jul 17;3(3):e37 [FREE Full text] [doi: [10.2196/resprot.3342](https://doi.org/10.2196/resprot.3342)] [Medline: [25048694](https://pubmed.ncbi.nlm.nih.gov/25048694/)]
39. Wilkerson JM, Patankar P, Rawat SM, Simon RB, Shukla KM, Rhoton J, et al. Recruitment strategies of Indian men who have sex with men in the State of Maharashtra into an online survey. *Int J Sex Health* 2016;28(3):221-227 [FREE Full text] [doi: [10.1080/19317611.2016.1193079](https://doi.org/10.1080/19317611.2016.1193079)] [Medline: [27668029](https://pubmed.ncbi.nlm.nih.gov/27668029/)]

40. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. *J Med Internet Res* 2012 Feb 01;14(1):e20 [FREE Full text] [doi: [10.2196/jmir.1978](https://doi.org/10.2196/jmir.1978)] [Medline: [22297093](https://pubmed.ncbi.nlm.nih.gov/22297093/)]
41. Hosmer D, Lemeshow S. Applied logistic regression, 2nd edition. New York: Wiley; 2000.
42. Krogstad J. Pew Research Center. Social media preferences vary by race and ethnicity URL: <http://www.pewresearch.org/fact-tank/2015/02/03/social-media-preferences-vary-by-race-and-ethnicity/> [accessed 2018-04-04] [WebCite Cache ID [6yQRBtuMR](https://www.webcitation.org/6yQRBtuMR)]
43. Smith A. Pew Research Center. 2014. African Americans and technology use URL: <http://www.pewinternet.org/2014/01/06/methods-23/> [accessed 2018-04-30] [WebCite Cache ID [6z4Fs8j1q](https://www.webcitation.org/6z4Fs8j1q)]
44. Caspi A, Gorsky P. Online deception: prevalence, motivation, and emotion. *Cyberpsychol Behav* 2006 Feb;9(1):54-59. [doi: [10.1089/cpb.2006.9.54](https://doi.org/10.1089/cpb.2006.9.54)] [Medline: [16497118](https://pubmed.ncbi.nlm.nih.gov/16497118/)]
45. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *Am Psychol* 2004;59(2):93-104. [doi: [10.1037/0003-066X.59.2.93](https://doi.org/10.1037/0003-066X.59.2.93)] [Medline: [14992636](https://pubmed.ncbi.nlm.nih.gov/14992636/)]
46. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. *Am Psychol* 2015 Sep;70(6):543-556. [doi: [10.1037/a0039210](https://doi.org/10.1037/a0039210)] [Medline: [26348336](https://pubmed.ncbi.nlm.nih.gov/26348336/)]
47. Arayasirikul S, Chen Y, Jin H, Wilson E. A Web 2.0 and epidemiology mash-up: using respondent-driven sampling in combination with social network site recruitment to reach young transwomen. *AIDS Behav* 2016 Jun;20(6):1265-1274 [FREE Full text] [doi: [10.1007/s10461-015-1234-4](https://doi.org/10.1007/s10461-015-1234-4)] [Medline: [26499337](https://pubmed.ncbi.nlm.nih.gov/26499337/)]
48. Cohen D. AdWeek. Instagram expands availability of ads, adds new features URL: <http://www.adweek.com/digital/instagram-ads-expansion-new-features/> [accessed 2017-08-25] [WebCite Cache ID [6sxtljNb6](https://www.webcitation.org/6sxtljNb6)]
49. Greenwood S, Perrin A, Duggan M. Pew Research Center. 2016. Social media update URL: <http://www.pewinternet.org/2016/11/11/social-media-update-2016/> [accessed 2017-07-02] [WebCite Cache ID [6sxuzkahx](https://www.webcitation.org/6sxuzkahx)]

Abbreviations

AOR: adjusted odds ratio
CTP: Center for Tobacco Products
DMA: designated market area
FDA: US Food and Drug Administration
GPS: global positioning system
IP: internet protocol
IRR: incidence rate ratio
LGBT: lesbian, gay, bisexual, and transgender
OR: odds ratio
REF: reference category

Edited by G Eysenbach; submitted 20.11.17; peer-reviewed by R Kayrouz, S Rose; comments to author 11.01.18; revised version received 10.04.18; accepted 12.04.18; published 18.06.18.

Please cite as:

Guillory J, Wiant KF, Farrelly M, Fiocco L, Alam I, Hoffman L, Crankshaw E, Delahanty J, Alexander TN
 Recruiting Hard-to-Reach Populations for Survey Research: Using Facebook and Instagram Advertisements and In-Person Intercept
 in LGBT Bars and Nightclubs to Recruit LGBT Young Adults
J Med Internet Res 2018;20(6):e197
 URL: <http://www.jmir.org/2018/6/e197/>
 doi: [10.2196/jmir.9461](https://doi.org/10.2196/jmir.9461)
 PMID: [29914861](https://pubmed.ncbi.nlm.nih.gov/29914861/)

©Jamie Guillory, Kristine F Wiant, Matthew Farrelly, Leah Fiocco, Ishrat Alam, Leah Hoffman, Erik Crankshaw, Janine Delahanty, Tesfa N Alexander. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 18.06.2018. 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.

Original Paper

Connecting With Your Dentist on Facebook: Patients' and Dentists' Attitudes Towards Social Media Usage in Dentistry

Nilesh Parmar¹, MBA, MSc, MSc, BDS; Lin Dong², MSc, BA (Hons); Andreas Benedikt Eisingerich², PhD, Mphil, BSC

¹Nilesh Parmar Dentistry, London, United Kingdom

²Imperial College Business School, Imperial College London, London, United Kingdom

Corresponding Author:

Lin Dong, MSc, BA (Hons)

Imperial College Business School

Imperial College London

Ayrton Road, Kensington

London, SW7 2AZ

United Kingdom

Phone: 44 020 7589 5111

Email: lin.dong13@imperial.ac.uk

Abstract

Background: Social media has begun to proliferate across medical areas and transformed how medical professionals serve and interact with their patients. It offers a new communication avenue that has the potential to engage patients and, hence, may be used to create value for both medical professionals and patients. In dentistry, even though patients and dentists frequently use social media in their personal lives, little is known about their attitudes and expectations toward using social media for professional interactions.

Objective: In this paper, we focus on the role of social media in dentistry. Specifically, we explore patients' and dentists' attitudes toward social media usage and their current online behaviors in this context. Furthermore, we examine potential challenges and opportunities regarding dentists' adoption of social media practices.

Methods: This research employed a large-scale online survey of 588 patients and 532 dental professionals. We assessed the attitudes, expectations, and social media behaviors from both patients' and dentists' perspectives.

Results: We found that more than 55% (290/532) of dentists in our sample have accounts for their dental practice on various social media platforms. Interestingly, while 73% (374/511) of patients did not expect their dental practice to have a social media presence, and 44% (207/468) thought that establishing a friendship with their dentists is not appropriate, the findings show that 36% (164/460) of patients had searched for their dentists, and 44% (207/470) of them were happy to establish contacts with dentists on social media. Furthermore, the findings highlight that patients were interested in exploring additional information such as online reviews and the qualifications of their dentists on Facebook pages. For dentists, more than half (375/432, 83%) of them in our sample thought that social media marketing is more efficient compared to traditional marketing.

Conclusions: Our findings revealed some key challenges and opportunities to utilize social media in dentistry. For both patients and dentists, the role of social media in dental services remains vague, and both parties still share concerns about connecting with each other on social media platforms. However, there also exists a sizeable number of patients who are already comfortable to connect with their dentists on social media sites such as Facebook. The current findings show that there is an opportunity for dental practices to trade upon a more active social media presence for enhanced patient interaction and engagement.

(*J Med Internet Res* 2018;20(6):e10109) doi:[10.2196/10109](https://doi.org/10.2196/10109)

KEYWORDS

social media, medical communication, dental practices, dental anxiety, Facebook, Twitter

Introduction

Background

The rise of social media has transformed how medical professionals interact with their patients and deliver different types of health care services [1,2,3]. Dentistry has been no exception [4]. The UK dental market was valued at £5.8bn and rising [5]. As with numerous other health profession contexts, social media has been noted to play an increasingly important role in dentistry [4]. This may create various challenges because, from a professional point of view, dentists must uphold the established image, principles, and procedures [6].

However, some of the critical traits of social media, such as self-exposure and self-disclosure, make established norms inadequate or outdated to help navigate daily interactions between dentists and their patients [7]. To this day, we still lack a clear understanding of how dentists and patients think of social media usage in dentistry. To what extent do patients and dentists feel comfortable to connect with one another on social sites such as Facebook? This research explores the attitudes of dentists and patients toward social media as well as their current online behaviors to uncover key challenges and potential opportunities for using social media in this professional context.

On social media, individuals construct public or semipublic profiles that enable them to create, circulate, share, and exchange information with their connections [8,9]. In dentistry, social media has been used in training and development of dental professionals for some time [10]. Dental education, for example, has relied on online communication to help professionals to develop clinical skills [11]. Previous studies have also highlighted that social media assists dentists to share domain knowledge with each other and facilitate professional networks [12]. It allows professionals from different locations to connect with one another and to discuss important issues and obtain feedback. These activities benefit lifelong learning and professional development [13]. That said, while social media has been hailed for effective peer-to-peer exchange, the role of social media for dentist-patient exchanges has been less than clear.

Interestingly, even though the literature has advocated the use of social media for interactions between medical professionals and patients, so far, professionals have been somewhat conservative and reluctant to involve social media in their workplace [10]. This is perhaps not too surprising. Medical services are associated with professional, formal practices while social media is usually depicted as an informal and entertaining platform [14]. Besides, personal information and communication on social media are visible to the public [15]. As a result, inappropriate social media practices might threaten medical professionals' identities and image [15,16]. This might also create concerns for patients about privacy issues such as whether dentists use their information on social media [17]. Surprisingly, to this day, little research exists on patients' and dentists' attitudes toward social media usage in the context of dentistry.

Lack of insights into the attitudes of dentists and patients toward using social media in dentistry may prevent dentists from taking

advantage of social media activities that might benefit both dentists and patients. For instance, social media offers a space to communicate and reach out. It empowers both dentists and patients to connect with each other without much time and distance limitations [9].

As an effective and relatively inexpensive means of communication, dental professionals can use social media for marketing activities [2]. Traditionally, dental practices would primarily advertise their services using local yellow pages (ie, telephone directory of local businesses and services) or put the promotion information on a small note in the window of dental surgeries. In the recent twenty years of digital transformation, social media has offered a forum for public communication that dentists can leverage to conduct marketing activities in a timely and cost-efficient way [4,16].

Furthermore, the communication space provided by social media allows dentists to diversify the traditional dental services by, for example, disseminating dental health information or offering online consultation [12,13,17]. Importantly, dentists can play a more active and significant role to cope with some enduring challenges in dentistry such as dental anxiety, a strong negative feeling toward visiting dental surgery that prevents patients to anticipate proper treatments [16].

Acknowledging the importance of social media in marketing and service delivery in dentistry, however, we still lack a clear understanding of patients' and dentists' attitudes toward social media usage [4]. Broad and accelerated access to information creates challenges for dental professionals to manage professional image and their relationship with patients [10,18]. Even though the governing institutions have realized this challenge, the published guidelines on how medical professionals interact with patients on social media are broad and, thus, may not provide feasible solutions [19].

Without a grounded understanding of dentists' and patients' attitudes and current usage behaviors, it is difficult to develop viable guidance that can help dentists to explore and trade upon the potential opportunities that social media platforms offer. We lack extensive knowledge about social media usage among patients and dentists. Therefore, the current research aimed to complement and extend existing knowledge and to address some of these critical issues.

Objective

The objectives of this study were to address the following two critical points:

1. To explore patients' and dentists' attitudes toward social media usage in dentistry.
2. To discover potential opportunities and challenges for dentists to adopt social media practices.

Methods

As part of this research, we conducted a large-scale online survey of patients and dentists. The dentists we target were from a general dental council or the equivalent registration. There were no specific criteria for selecting the patients. We designed separate surveys for patients and dentists with similar themes.

The principal goal was to compare the responses of patients and dentists regarding their attitudes and behaviors related to the social media usage in dentistry.

The dentist survey consisted of 27 questions, with the estimated time to complete the survey to be around 15 minutes. The patient survey had 20 questions with some being split into sub-questions. The estimated time to finish the survey was around 12 minutes. We used the software Qualtrics survey platform because it is accessible with no requirement for the participants to register. Furthermore, it is available on multiple platforms including smartphones, tablets and desktop computers and, thus, helped us to encourage more participation in this study.

The survey for dentists was distributed via a dentist-only Facebook group called “For Dentists, by Dentists”. This group has over 4,500 dentists as members and currently is the most active online forum for dentists in the UK. Also, the first author promoted the survey through various avenues such as professional blogs and websites such as dentistry.co.uk, where members of the “For Dentists, by Dentists” Facebook group often visit. Specifically, the survey was aimed at the dental community, which comprises all those with a General Dental Council registration. It encompasses dentists, nurses, dental technicians, and hygienists. The key aim here was to assess dentists’ participation in and attitudes about social media and to assess to what extent this participation extends to their patients.

We promoted the survey for patients mainly through social media platforms. The patient survey was distributed via Facebook, LinkedIn, and Twitter messages by the first author, with all recipients being asked to share the survey with their connections. Furthermore, in the first author’s dental practice, the survey was uploaded onto an iPad, and patients were encouraged to fill in the survey while they waited for their appointments. The principal goal here was (1) to assess patients’ expectations of the social media presence of their chosen dental practice, (2) to determine what content on a Facebook dental practice page that patients find relevant and important, and (3) to explore patients’ view toward communicating with their dentists via social media. Both surveys ran for 6 weeks with regular updates on social media at premeditated times to enhance visibility within individuals’ timelines and increase effective response rates. This study obtained ethics approval by Imperial College London.

Results

Overview

The data collection efforts resulted in 588 patients’ responses and 532 dentists’ responses. Because some respondents skipped a few questions, we highlighted the number of responses for each question when discussing the detailed results to ensure accuracy. [Figure 1](#) summarizes some of the key insights that

emerged from the data. Specifically, 77% (399/515) of patients expected their dental practice to have a website, but the majority (374/511, 73%), did not expect their dental practice to have a social media presence. That said, 44% (207/470) of patients noted to be happy to be contacted by their dentist through social media and 36% (164/460) have searched for their dentist on social media. Interestingly, while 74% (333/448) of dentists and 44% (207/468) of patients state that it is not appropriate for dentists and their patients to be friends on social media platforms, 29% (112/382) of dentists and 17% (76/460) of patients have accepted friend requests have added their dentist on social media already. We elaborate on additional vital insights in the following section.

Platform Usage by Both Patients and Dentists

Facebook was the most popular platform to use for both patients (482/492, 98%), and dentists (290/377, 77%). For patients, Twitter, Instagram, and LinkedIn were popular choices in their personal lives. For dentists, except for Facebook, it is noteworthy that 37% (138/377) of them opened a business account on Twitter and 27% (103/377) on Google+ (see [Figure 1](#)). In general, we had a balanced response regarding the gender of both patients (321/572, 56% were female) and dentists (245/532, 46% were female) in our survey. [Table 1](#) shows the demographic and summary statistics of survey respondents. Most of the responses were over 25 years old and owned at least one personal social media account.

Patients’ Attitudes Toward Social Media Usage in Dentistry

The findings show that 47% (238/508) of patients have visited their dental surgery’s Facebook page or website. A total of 17% (76/460) searched their medical doctors or dentists on social media and added them as friends. Of patients, 19% (88/460) searched for their medical doctors and dentists on social media but did not add them as friends on social media. A total of 64% (296/460) neither searched nor added their medical doctors and dentists as friends on social media (see [Table 1](#)). Furthermore, the findings reveal that 79% (391/493) of the patients agreed dental surgeries should have an online presence of sort (see [Figure 2](#)).

Moreover, for patients, our results show that social media does not play a significant role in their decision-making process when selecting a dental practice (see [Figure 3](#)). More specifically, patients valued recommendations from friends and family, facilities and technologies, online reviews and quality of websites more than they did a dental practice’s social media presence.

Regarding dentists’ Facebook page, patients ranked qualifications as the most important content to be displayed. Besides qualifications, for some patients, positive reviews, awards, and original content were also appreciated on dentists’ Facebook pages (see [Figure 4](#)).

Figure 1. A summary of findings between dentistry and social media.

Table 1. Patient survey and dentist survey: demographics and personal social media usage.

Survey Variables	Dentist Survey, n (%)	Patient Survey, n (%)
Demographic		
Gender	N=532	N=572
Male	287 (54%)	251 (44%)
Female	245 (46%)	321 (56%)
Age group	N=529	N=588
18-25	9 (2%)	65 (11%)
26-35	331 (62%)	234 (40%)
36-45	130 (25%)	172 (29%)
46+	59 (11%)	117 (20%)
Social media		
Do you have a personal social media account?	N=451	N=499
Yes	421 (93%)	484 (97%)
No	30 (7%)	15 (3%)
Patients' behaviors on social media		
Have you visited your dental surgery's Facebook page or website?	N/A ^a	N=508
Yes	N/A	238 (47%)
No	N/A	270 (53%)
Have you ever added any of your medical doctors or dentists on social media?	N/A	N=460
Yes, searched and added	N/A	76 (17%)
Yes, but only searched	N/A	88 (19%)
Neither	N/A	296 (64%)

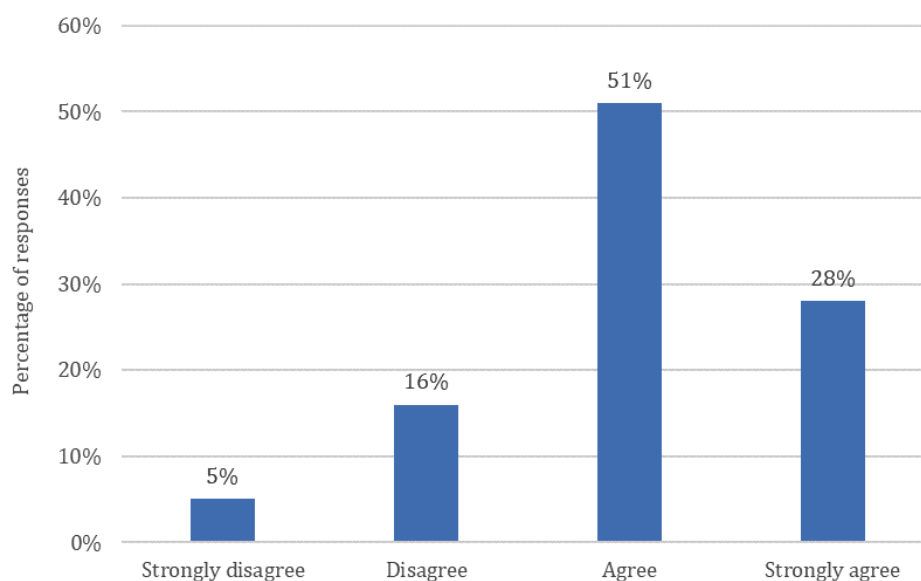
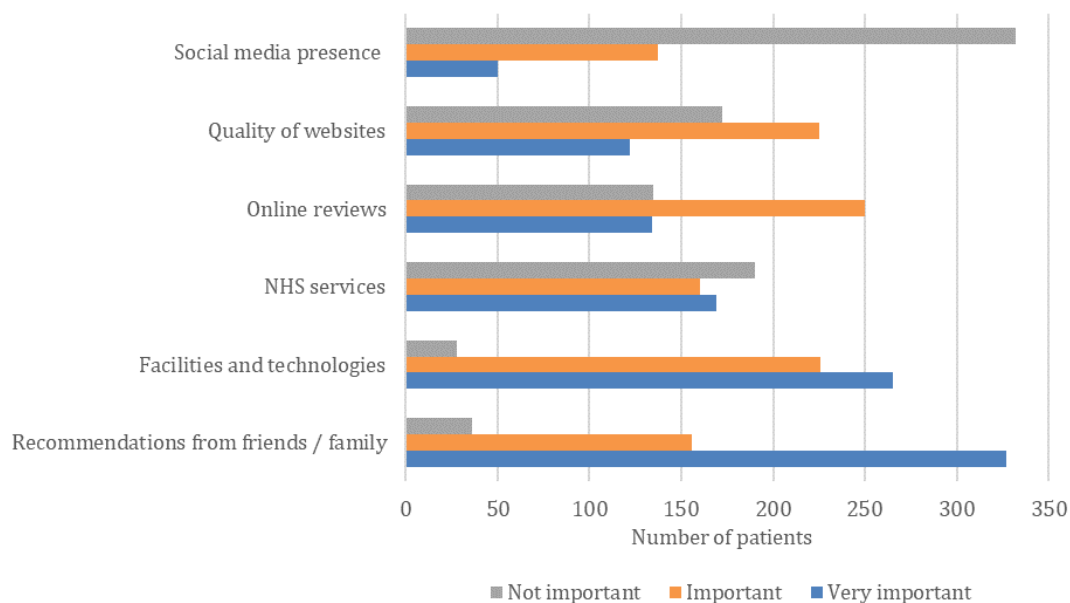
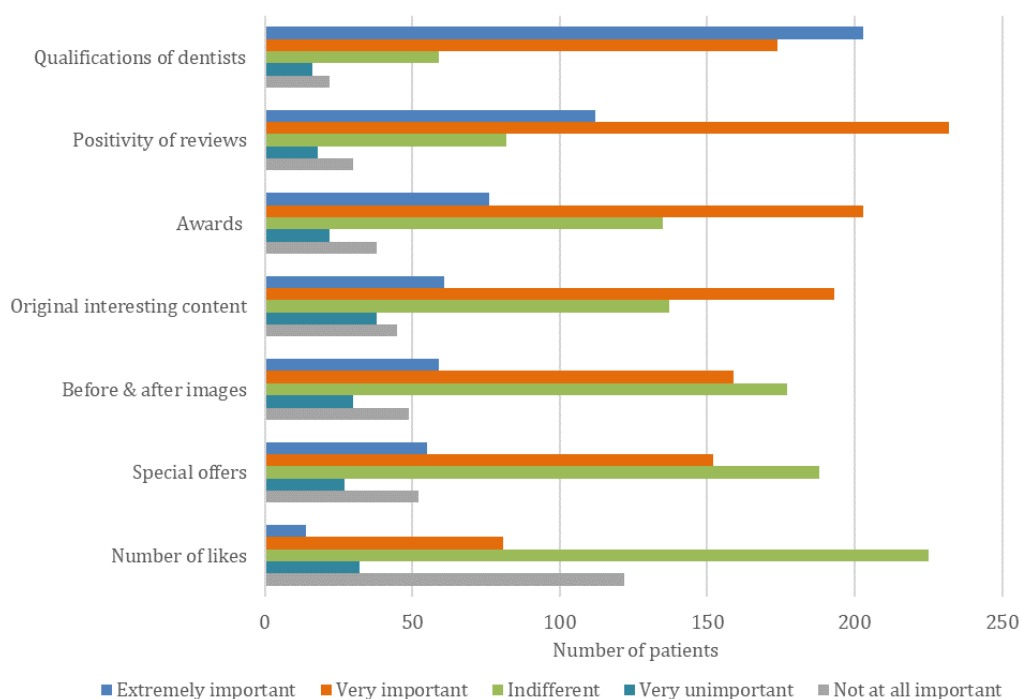
^aN/A: not applicable.**Figure 2.** Responses from patient survey prompt: A modern-day dental practice should have an online presence (N=493).

Figure 3. The results of a patient survey. Factors that are most important to patients when choosing a dental practice.**Figure 4.** Results of a patient survey. Factors that are most important to patients when looking at a Facebook page for a dental practice.

An interesting finding is that when asked about the objection to visiting a dental practice, patients showed greater pain-related anxiety caused by, for example, injection and drill, compared with financial-related anxiety (see Figure 1). There seems to be an opportunity for dentists to help alleviate patients' concerns and anxieties. Social media might offer a useful tool in helping patients access critical and helpful information about dental solutions and access these in a convenient and non-threatening way. Next, we explore the insights we obtained from our survey with dentists.

Dentists' Attitudes Toward Social Media as a Marketing Tool

The findings reveal that dentists hold a positive attitude toward social media as an effective tool to reach new patients (see Figure 5). Critically, the return on investment (ROI) of social media marketing was noted as higher than for traditional marketing efforts. Hence, potentially, social media can improve the financial performance of dental practices (see Figure 6).

Figure 5. Responses from patient survey prompt: How effective is a social media presence for a dental practice to engage and obtain new patients (N=432)?

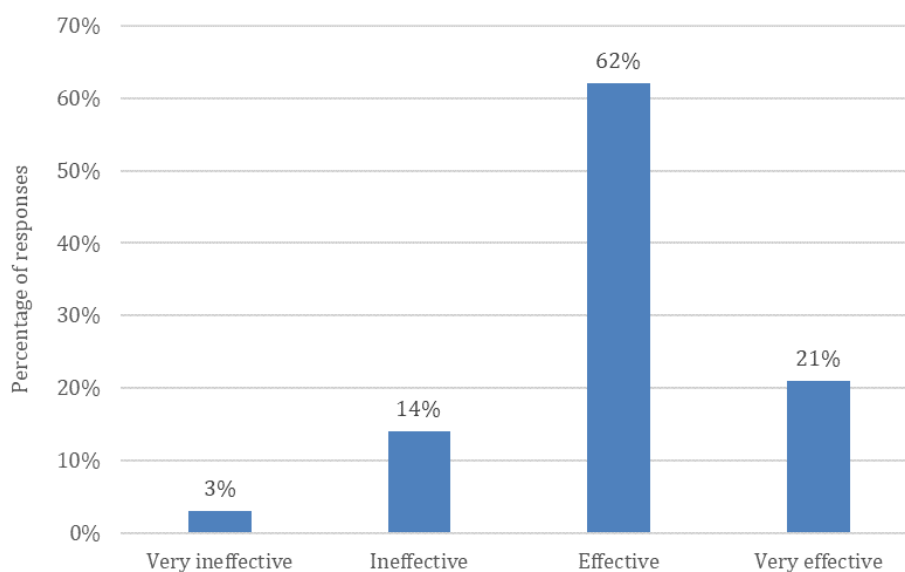
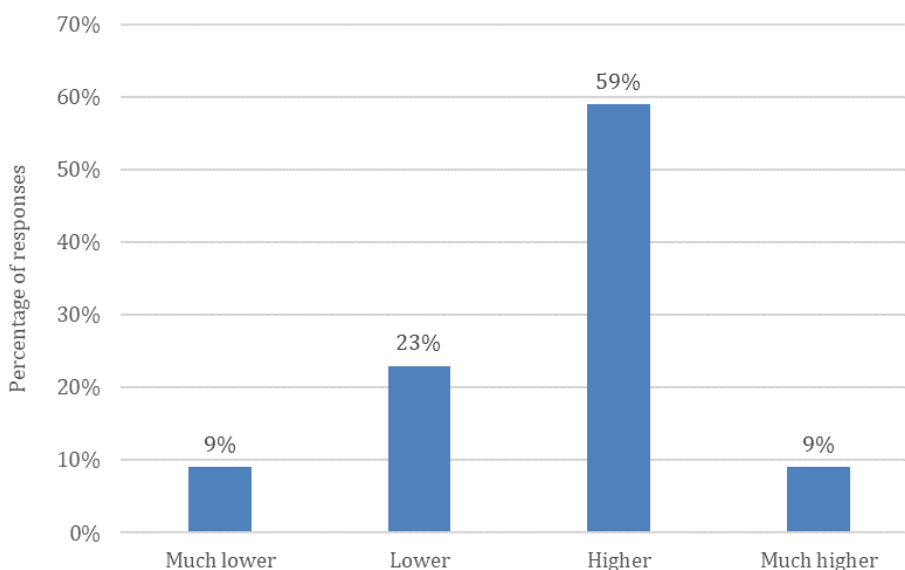


Figure 6. Responses from patient survey prompt: What do you think is the return-on-investment of social media marketing when compared to conventional marketing for a dental practice (N=422)?



Discussion

Principal Findings

The use of social media as a means to facilitate communication between dentists and their patients has been somewhat limited [10]. Social media, however, may offer critical opportunities for dentists to facilitate patient-dentist relationship outside the surgery [20]. That said, we did not know much about how patients and dentists feel about direct communication with each other on social media. Before this study, we conducted a PubMed search, which resulted in a total of 179 articles with the search term of “social media and dentistry” and the conditions of “published in the recent ten years” and “full text available”. Only a few articles discuss the role of social media in dentistry, and the empirical samples of these studies are relatively small. The present research thus complements and

extends the extant body of work by surveying the attitudes and preferences of 588 patients and 532 dentists. It reveals interesting and some surprising insights into attitudes toward social media and dentistry.

In line with prior work, our results show that both patients and dentists are already active on various social media platforms to some extent [4,15]. However, they share some concerns about connecting with each other as online friends. Although dental professionals see social media as an essential opportunity to improve marketing efficiency, we find that patients have unclear expectations about how social media adds value to them. Despite these challenges, we have also uncovered many key opportunities that offer critical insights for dental professionals.

Patients' Relationship Status With Dentists on Facebook: It's complicated

The current findings showcase that both dentists and patients remain somewhat hesitant to connect with each other as friends on social media. This finding is in line with previous work, which noted that the trait of self-disclosure might increase both parties' concerns about their privacy [1,10]. Furthermore, social media, such as Facebook, is typically used to connect with friends and family [21]. Hence, relevant social media practices and routines have been developed outside the professional context—that is, what is appropriate to be shared on social media may not necessarily be appropriate in a more professional context [9]. That said, a surprisingly high number of patients already connect with dentists on Facebook. For instance, 44% (207/470) of patients indicated to be happy to be contacted by their dentists on social media. Thus, avenues to explore this exciting issue further do exist.

When moving from interpersonal to professional interactions, intuitively, the relationship between patient and dentist often does not equal to “friendship”. Thus, the routines on social media such as commenting on and liking each other's posts may not be easily applied in a professional context. It may confuse patients and dentists about the type of relationships they should hold and how to interact with each other on social media. That said, without the capability to connect with patients on social media, it would be challenging for dentists to leverage social media as an effective marketing tool. Patients perhaps also miss the opportunity to have better access to their dentists and obtain relevant and timely information about dental services that can help alleviate their anxieties and fears about their next visit to the dentist.

Reaching Out to Patients Via Social Media

The findings highlight that dentists value social media as useful marketing tool. On the contrary, as the target audience, most patients do not consider social media as part of their decision-making process. However, 76/460 (17%) of patients in our sample declared that they had added their dentist on social media. Thus, while some patients do not seem to see the value of a social media presence by dental practices, there also appears to be an opportunity for dentists to trade upon a more active social media presence.

Critically, patients appear not to know why they should use social media when selecting a dental practice. Such ambiguity and mismatched attitudes present a barrier for dentists to take advantage of social media fully. Dentists need to think about and explicitly suggest how social media offers exclusive services and information for patients that are unavailable on other channels (eg, website or phone call). For example, YouTube is perceived and widely acknowledged as a convenient educational content sharing platform for dentists [22]. Dentists can communicate with their patients about what information or services patients can obtain on Facebook, Twitter or other social media platforms.

Opportunities for Dentists and Patients to Connect on Social Media

The findings indicate many opportunities for dentists and patients to connect on social media. First, dental professionals may use LinkedIn to showcase their expertise. Based on our findings, patients would like to know dental professionals' qualifications before they visit dental practices. As one of the popular social media platforms that patients use, LinkedIn operates the world's largest professional network [23]. It allows dentists to display their expertise including educational background, work experience, skills, and endorsements. The norms and codes of this particular platform are consistent with the professional image of dentists [24]. It also minimizes dentists' concerns about invasions of their personal lives. Thus, LinkedIn naturally serves as a professional space that dentists can use to demonstrate their expertise and capabilities, as part of social media marketing activities.

Second, dentists may connect with patients on social media to encourage online word of mouth. As the most critical factor that impacts patients' decision-making, recommendations from friends and family could be operated on social media platforms. Compared to offline word of mouth, recommending dental practices on social media allows patients to share more information with their peers when necessary. For example, a patient may send the website of a recommended dental practice to her peers directly and encourage them to obtain more information by visiting the website. Social media also allows various forms of word of mouth such as reviewing dental practice's service and reposting its advertisements. Also, activities such as commenting on the posts and simply “liking” the posts increase a surgery's awareness within its patients' social network [21].

Third, since patients are aware of and visit the websites of their dental practices, dentists might introduce their social media account on the website. They should state the functions of each social media platform explicitly. For example, some banks guide their customers to use Twitter for online help by adding the link to their Twitter page under the tab of “Help” [25]. That is, customers see the bank's Twitter page as a functional place to solve their questions. By introducing the functions of social media platform, patients would have a clear idea of how social media creates value for them (eg, a place for online Q&A or a place for finding helpful or encouraging daily dental health suggestions). Also, integrating social media platforms with a website also increases the chance for patients to be aware of the social media presence of their dental practice.

Limitations and Future Research

We note the following limitations of our study, which also offer promising avenues for future work. First, in the data collection process, asking some of the respondents to use a tablet to fill in the questionnaires might introduce a particular bias. For example, respondents may not be used to these devices and skip questions. That said, the answers from respondents who filled out the survey on a tablet did not differ from respondents who filled out the survey through other means in our study. Second, we did not collect qualitative data. For example, 74% (333/448) of the dentists agreed that connecting as friends on social media

is inappropriate. Integrating a qualitative method such as in-depth interviews would allow us to explore dentist's concerns in more detail.

Also, a qualitative method may allow us to analyze successful cases of social media usage in dentistry. It would provide some specific suggestions for dental professionals. Moreover, the analyses of our data suggest that social media usage in the context of dentistry and patients did not significantly differ for female versus male or older versus younger patients. This is intriguing and warrants further investigation. What other patient characteristics may influence attitudes toward social media usage in the context of dentistry? We invite additional research to address these relevant issues. Our research thus suggests several promising avenues for future studies.

Social media provides opportunities to tackle some issues such as after-clinic care and dental anxiety. Dental anxiety, an enduring challenge in dentistry, often causes sleeplessness, reluctance to form close interpersonal relationships and problems in workplaces [22,26,27]. It acts as a barrier for patients to actively seek necessary dental treatments [28]. Unfortunately, despite advances in technology, the prevalence of dental phobias has not been changed in the last twenty years [26]. Coping with it requires professional help, especially before clinic visits. Social media may allow dental professionals to deliver the help outside a dental practice environment. Notably, some features of social media such as gamification [29,30] and engaging website design [31] may be particularly relevant to address and inform patients about critical health-related issues and educate them, which can also help increase their trust and appreciate the level of service received [32-34]. Research studying the role of how social media may help reduce patients' dental anxiety is a worthwhile and exciting endeavor.

Furthermore, future research may also consider the role of professional and regulatory bodies. Social media usage might threaten privacy and bring new challenges for both patients and dental professionals. For dental professionals, the principles, practices, and procedures related to social media communication are not clear [15]. For example, dentists may use social media to demonstrate their expertise by uploading clinical photographs of before and after treatment. Is it acceptable? If so, when? Moreover, unprofessional audiences may challenge dentists' expertise and capabilities in a public place [15,35]. These potential violations need to be considered carefully and managed by professional and regulatory bodies to ensure that social media is a safe place for dental professionals in which to engage.

Currently, the General Dental Council (GDC), the regulatory body for dentistry in the UK, has published a guideline in 2013 titled "Guidance on Using Social Media". It takes a hardline approach to its registrants [19]. For example, the document states: "maintain appropriate boundaries in the relationships you have with patients" and "you should think carefully before accepting friend requests from patients" [19]. The British Medical Association (BMA) also suggests that accepting friendship requests from patients might be inappropriate [36]. Further research may explore the type of guidance that professional and regulatory bodies should provide. That is, they should navigate dentists' daily social media activities that not

only ensure the alignment with ethical codes but also leave sufficient room for them to use social media to create value for their patients [36].

Future research may explore some of the critical skills and mindset that young professionals need to flourish in the digital era. Medical and pharmaceutical students are not always clear about the long-term repercussions associated with their online behaviors [37]. Merely listing acceptable social media practices may not work, as unfavorable outcomes can be caused for various reasons [38]. For example, users can upload photographs and thoughts on social media within seconds, without evaluating whether the message aligns with their professional image. These messages are also visible electronic footprints that might have unintended and negative consequences [15]. Social media usage may also create an environment where patients feel more comfortable to participate and proactively ask questions [39,40] or even challenge others [41]. Thus, while social media can facilitate transparency and allow people to access information conveniently [42,43], which allows service providers to be seen as more helpful and hence play a more critical role in people's lives [44-46], true engagement online and offline remains a challenge [47,48]. Future research that studies how the current medical education may better help professionals and equip them with the skills needed to reach out to patients and manage the relationships with their patients in the context of social media is richly deserving.

Also, advances in technologies such as artificial intelligence might impact patients' trust and behaviors significantly or even shift the extant business models for high credence services such as dentistry. Given the importance of educating one's customer [49,50], it is critical to start to think about how young professionals can develop a technology-sensitive mindset as a long-term professional asset from their professional education and training to share critical information with and earn stronger trust from their patients. Such technology-sensitive mindset can help medical professionals navigate a hard-to-predict environment [51,52] and engage in responsible behavior while trading upon social media offerings (eg, visually pleasing profile with enticing logo) to explore new opportunities to connect with their patients more effectively [53-55]. As current research shows, people can become instantaneously attached to, and find it hard to give up, the digital services they use [56]. Future work that explores how dentists can make use of social media and other mHealth solutions to further their self-identification with their profession and pride [57] and ensure patients lead healthier lives [58] is richly deserving.

Conclusion

A total of 207/470 (44%) of patients noted to be happy to be contacted on social media by their dentists, while 164/460 (36%) of patients have already searched for their dentist on social media. Social media offers opportunities for dental professionals to improve the efficiency of marketing activities, and to provide additional services such as offering online diagnosis and online Q&A. However, within the professional context, the norms and procedures related to social media usage in patient-dentist communication remain vague and underdeveloped. Moreover, while 74% (333/448) of dentists agreed that social media

friendship was not appropriate, 112/382 (29%) accepted friend requests from their patients on social media. Our study represents an important step to unpack patient and dentist expectations about, and attitudes toward, social media usage in

dentistry. We hope additional work can build on some of the current findings and shed additional light on the role of social media in dentist-patient interactions and relationships.

Acknowledgments

The dissemination of current research findings is funded by the Economic and Social Research Council (ESRC) (grant reference number ES/M500562/1).

Conflicts of Interest

None declared.

Authors' Contributions

NP, LD, ABE reviewed the literature; NP and ABE designed the study. NP collected and analyzed the data; NP, LD, ABE wrote the article.

References

- 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;15(8):e184 [FREE Full text] [doi: [10.2196/jmir.2708](https://doi.org/10.2196/jmir.2708)] [Medline: [23985172](https://pubmed.ncbi.nlm.nih.gov/23985172/)]
- Shibasaki S, Gardner K, Sibthorpe B. Using knowledge translation to craft “sticky” Social Media health messages that provoke interest, raise awareness, impart knowledge, and inspire changes. *JMIR Mhealth Uhealth* 2016 Oct 05;4(4):e115 [FREE Full text] [doi: [10.2196/mhealth.5987](https://doi.org/10.2196/mhealth.5987)] [Medline: [27707685](https://pubmed.ncbi.nlm.nih.gov/27707685/)]
- Kazley AS, Hamidi B, Balliet W, Baliga P. Social Media use among living kidney donors and recipients: Survey on current practice and potential. *J Med Internet Res* 2016 Dec 20;18(12):e328 [FREE Full text] [doi: [10.2196/jmir.6176](https://doi.org/10.2196/jmir.6176)] [Medline: [27998880](https://pubmed.ncbi.nlm.nih.gov/27998880/)]
- Melkers J, Hicks D, Rosenblum S, Isett KR, Elliott J. Dental blogs, podcasts, and associated social media: Descriptive mapping and analysis. *J Med Internet Res* 2017 Jul 26;19(7):e269 [FREE Full text] [doi: [10.2196/jmir.7868](https://doi.org/10.2196/jmir.7868)] [Medline: [28747291](https://pubmed.ncbi.nlm.nih.gov/28747291/)]
- LaingBuisson. Dentistry UK summary market report 2014 URL: <https://www.practiceplan.co.uk/resource-library/the-dental-industry-environment/dentistry-uk-summary-market-report-2014> [accessed 2018-02-09] [WebCite Cache ID [6x71dj13Q](https://www.webcitation.org/6x71dj13Q)]
- Evetts J. Short note: The sociology of professional groups. *Current Sociology* 2016 Jun 29;54(1):133-143. [doi: [10.1177/0011392106057161](https://doi.org/10.1177/0011392106057161)]
- Green T, Hartley N, Gillespie N. Service provider's experiences of service separation: The case of Telehealth. *Journal of Service Research* 2016 Nov 11;19(4):477-494. [doi: [10.1177/1094670516666674](https://doi.org/10.1177/1094670516666674)]
- Boyd DM, Ellison N. Social network sites: Definition, history, and scholarship. *Journal of Computer Mediated Communication* 2008;13:210-230 [FREE Full text] [doi: [10.1111/j.1083-6101.2007.00393.x](https://doi.org/10.1111/j.1083-6101.2007.00393.x)]
- Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. *Business Horizons* 2010;53 (1):59-68 [FREE Full text]
- Neville P, Waylen A. Social media and dentistry: Some reflections on e-professionalism. *Br Dent J* 2015 Apr 24;218(8):475-478. [doi: [10.1038/sj.bdj.2015.294](https://doi.org/10.1038/sj.bdj.2015.294)] [Medline: [25908363](https://pubmed.ncbi.nlm.nih.gov/25908363/)]
- GOV.UK. Digital strategy: Leading the culture change in health and care. UK URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213222/final-report1.pdf [accessed 2018-01-27] [WebCite Cache ID [70R8ELvEB](https://www.webcitation.org/ID70R8ELvEB)]
- Sugawara Y, Narimatsu H, Tsuya A, Tanaka A, Fukao A. Medical institutions and twitter: A novel tool for public communication in Japan. *JMIR Public Health Surveill* 2016;2(1):e19 [FREE Full text] [doi: [10.2196/publichealth.4831](https://doi.org/10.2196/publichealth.4831)] [Medline: [27227154](https://pubmed.ncbi.nlm.nih.gov/27227154/)]
- Keller B, Labrique A, Jain KM, Pekosz A, Levine O. Mind the gap: Social Media engagement by public health researchers. *J Med Internet Res* 2014;16(1):e8 [FREE Full text] [doi: [10.2196/jmir.2982](https://doi.org/10.2196/jmir.2982)] [Medline: [24425670](https://pubmed.ncbi.nlm.nih.gov/24425670/)]
- Maben-Feaster RE, Stansfield RB, Opipari A, Hammoud MM. Evaluating patient perspectives of provider professionalism on Twitter in an academic obstetrics and gynecology clinic: Patient survey. *J Med Internet Res* 2018 Mar 12;20(3):e78 [FREE Full text] [doi: [10.2196/jmir.8056](https://doi.org/10.2196/jmir.8056)] [Medline: [29530838](https://pubmed.ncbi.nlm.nih.gov/29530838/)]
- Holden ACL. Social Media and professionalism: Does the profession need to rethink the parameters of professionalism within Social Media? *Aust Dent J* 2017 Mar;62(1):23-29. [doi: [10.1111/adj.12445](https://doi.org/10.1111/adj.12445)] [Medline: [27468729](https://pubmed.ncbi.nlm.nih.gov/27468729/)]
- Gao X, Hamzah SH, Yiu CKY, McGrath C, King NM. Dental fear and anxiety in children and adolescents: Qualitative study using YouTube. *J Med Internet Res* 2013;15(2):e29 [FREE Full text] [doi: [10.2196/jmir.2290](https://doi.org/10.2196/jmir.2290)] [Medline: [23435094](https://pubmed.ncbi.nlm.nih.gov/23435094/)]

17. Sinha A, Porter T, Wilson A. The use of online health forums by patients with chronic cough: Qualitative study. *J Med Internet Res* 2018 Jan 24;20(1):e19 [FREE Full text] [doi: [10.2196/jmir.7975](https://doi.org/10.2196/jmir.7975)] [Medline: [29367181](#)]
18. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: An integrative review. *J Med Internet Res* 2016 Jun 16;18(6):e166 [FREE Full text] [doi: [10.2196/jmir.5312](https://doi.org/10.2196/jmir.5312)] [Medline: [27328967](#)]
19. Royal College of General Practitioners. 2013. Social media highway code URL: <http://www.foundation.severndeanery.nhs.uk/assets/Foundation/Info-for-incoming-and-current-trainees/C62-RCGP-Social-Media-Highway-Code.pdf> [accessed 2018-06-19] [WebCite Cache ID 70HIXXt0l]
20. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: Future prospects and the research agenda. *J Med Internet Res* 2001;3(2):E15 [FREE Full text] [doi: [10.2196/jmir.3.2.e15](https://doi.org/10.2196/jmir.3.2.e15)] [Medline: [11720957](#)]
21. Eisingerich AB, Chun H, Liu Y, Jia HM, Bell SJ. Why recommend a brand face-to-face but not on Facebook? How word-of-mouth on online social sites differs from traditional word-of-mouth. *Journal of Consumer Psychology* 2015;25(1):120-128 [FREE Full text]
22. Knösel M, Jung K, Bleckmann A. YouTube, dentistry, and dental education. *J Dent Educ* 2011 Dec;75(12):1558-1568 [FREE Full text] [Medline: [22184594](#)]
23. LinkedIn. About LinkedIn URL: <https://about.linkedin.com/> [accessed 2018-03-20] [WebCite Cache ID 70HlpFuGd]
24. Rooderkerk RP, Pauwels KH. No comment?! The drivers of reactions to online posts in professional groups. *Journal of Interactive Marketing* 2016 Aug;35:1-15. [doi: [10.1016/j.intmar.2015.12.003](https://doi.org/10.1016/j.intmar.2015.12.003)]
25. Santander. Contact us URL: <https://www.santander.co.uk/uk/help-support/contact-us> [accessed 2018-06-19] [WebCite Cache ID 70HlvNEGM]
26. Boyle CA, Newton T, Milgrom P. Who is referred for sedation for dentistry and why? *Br Dent J* 2009 Mar 28;206(6):e12 [FREE Full text] [doi: [10.1038/sj.bdj.2009.251](https://doi.org/10.1038/sj.bdj.2009.251)] [Medline: [19329945](#)]
27. Ellison NB, Steinfield C, Lampe C. The benefits of Facebook "Friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication* 2007;12(4):1143-1168.
28. Sohn W, Ismail AI. Regular dental visits and dental anxiety in an adult dentate population. *J Am Dent Assoc* 2005 Jan;136(1):58-66. [Medline: [15693497](#)]
29. El-Hilly AA, Iqbal SS, Ahmed M, Sherwani Y, Muntasir M, Siddiqui S, et al. Game on? Smoking cessation through the gamification of mHealth: A Longitudinal qualitative study. *JMIR Serious Games* 2016 Oct 24;4(2):e18 [FREE Full text] [doi: [10.2196/games.5678](https://doi.org/10.2196/games.5678)] [Medline: [27777216](#)]
30. Tudor-Sfetea C, Rabee R, Najim M, Amin N, Chadha M, Jain M, et al. Evaluation of two mobile health apps in the context of smoking cessation: Qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. *JMIR Mhealth Uhealth* 2018 Apr 18;6(4):e98 [FREE Full text] [doi: [10.2196/mhealth.9405](https://doi.org/10.2196/mhealth.9405)] [Medline: [29669708](#)]
31. Eisingerich AB, Kretschmer T. In e-commerce, more is more. *Harvard Business Review* 2008 Mar;20-21 [FREE Full text]
32. Bell SJ, Auh S, Eisingerich AB. Unraveling the customer education paradox: When, and how, should firms educate their customers? *Journal of Service Research* 2017; 20(3). *Journal of Service Research* 2017;20(3):306-321 [FREE Full text]
33. Eisingerich AB, Bell SJ. Perceived service quality and customer trust: Does enhancing customers' service knowledge matter? *Journal of Service Research* 2008; 10(3). *Journal of Service Research* 2008;10(3):256-268 [FREE Full text]
34. Bell SJ, Eisingerich AB. Work with me. *Harvard Business Review* 2007 Jun;32 [FREE Full text]
35. Boyt TE, Lusch RF, Naylor G. The role of professionalism in determining job satisfaction in professional services: A study of marketing researchers. *Journal of Service Research* 2016 Jun 29;3(4):321-330. [doi: [10.1177/109467050134005](https://doi.org/10.1177/109467050134005)]
36. BMA. 2017. Social media guidance for doctors URL: <https://www.bma.org.uk/advice/employment/ethics/social-media-guidance-for-doctors> [accessed 2018-06-19] [WebCite Cache ID 70HmXMImG]
37. Chretien KC, Goldman EF, Beckman L, Kind T. It's your own risk: Medical students' perspectives on online professionalism. *Acad Med* 2010 Oct;85(10 Suppl):S68-S71. [doi: [10.1097/ACM.0b013e3181ed4778](https://doi.org/10.1097/ACM.0b013e3181ed4778)] [Medline: [20881708](#)]
38. Third A, Richardson I, Collin P, Rahilly K, Bolzan N. Inspire Foundation. 2011. Intergenerational attitudes towards social networking and cybersafety: A living lab URL: <http://researchdirect.westernsydney.edu.au/islandora/object/uws:11518> [accessed 2018-02-09] [WebCite Cache ID 70RAGwCOd]
39. Eisingerich AB, Auh S, Merlo O. Acta non verba? The role of customer participation and word of mouth in the relationship between service firm's customer satisfaction and sales performance. *Journal of Service Research* 2014;17(1):40-53 [FREE Full text]
40. Merlo O, Eisingerich AB, Auh S. Why customer participation matters. *MIT Sloan Management Review* 2014;55(2):81-88 [FREE Full text]
41. Eisingerich AB, Bhardwaj G, Miyamoto Y. Behold the extreme consumers and learn to embrace them. *Harvard Business Review* 2010;30-31 [FREE Full text]
42. Foscht T, Lin Y, Eisingerich AB. Blinds up or down? The influence of transparency, future orientation, and CSR on sustainable and responsible behavior. *European Journal of Marketing* 2018;52(3/4):476-498 [FREE Full text]
43. Liu Y, Eisingerich AB, Auh S, Merlo O, Chun HE. Service firm performance transparency: How, when, and why does it pay off? *Journal of Service Research* 2015(4). *Journal of Service Research* 2015;18(4):451-467 [FREE Full text]

44. Park CW, MacInnis D, Priester J, Eisingerich AB, Iacobucci D. Brand attachment and brand attitude strength. *Journal of Marketing* 2010;74(6):1-17 [[FREE Full text](#)]
45. Park CW, MacInnis D, Eisingerich AB. *Brand admiration: Building a business people love*. New York: Wiley; 2016.
46. Park CW, Eisingerich AB, Park J. Attachment-aversion (AA) model of customer-brand relationships. *Journal of Consumer Psychology* 2013;23(2):229-248 [[FREE Full text](#)]
47. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: Systematic review. *J Med Internet Res* 2017 Nov 28;19(11):e399 [[FREE Full text](#)] [doi: [10.2196/jmir.8598](#)] [Medline: [29187342](#)]
48. Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, et al. Prevalence of health app use among older adults in Germany: National survey. *JMIR Mhealth Uhealth* 2018 Jan 23;6(1):e26 [[FREE Full text](#)] [doi: [10.2196/mhealth.8619](#)] [Medline: [29362211](#)]
49. Eisingerich AB, Bell SJ. Maintaining customer relationships in high credence services. *Journal of Services Marketing* 2007;21(4):8559-8262. [doi: [10.1108/08876040710758559](#)]
50. Eisingerich AB, Bell SJ. Customer education increases trust. *MIT Sloan Management Review* 2008;50(1):10-11 [[FREE Full text](#)]
51. Seifert M, Siemsen E, Hadida AL, Eisingerich AB. Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management* 2015;36:33-45. [doi: [10.1016/j.jom.2015.02.001](#)]
52. Seifert M, Siemsen E, Hadida AL, Eisingerich AB. When to forget the rearview mirror. *Harvard Business Review* 2015;30 [[FREE Full text](#)]
53. Eisingerich AB, Bhardwaj G. Does corporate social responsibility help protect a company's reputation? *MIT Sloan Management Review* 2011;52(3):18 [[FREE Full text](#)]
54. Chun HEH, Park CW, Eisingerich AB, MacInnis D. Strategic benefits of low fit brand extensions: When and why? *Journal of Consumer Psychology* 2015; 25(4). doi.org/10.1016/j.jcps.12.2014;25(4):003-595. [doi: [10.1016/j.jcps.2014.12.003](#)]
55. Park CW, Eisingerich AB, Pol G. The power of a good logo. *MIT Sloan Management Review*; 55(2) 2014;55(2):10-12 [[FREE Full text](#)]
56. Fritze MP, Eisingerich AB, Benkenstein M. Digital transformation and possession attachment examining the endowment effect for consumers? Relationships with hedonic and utilitarian digital service technologies. *Electronic Commerce Research* 2018;1-27. [doi: [10.1007/s10660-018-9309-8](#)]
57. Gill-Simmen L, MacInnis D, Eisingerich AB, Park CW. Brand-self connections and brand prominence as drivers of employee brand attachment. *AMS Review* 2018;1-19. [doi: [10.1007/s13162-018-0110-6](#)]
58. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mhealth app exploring the impact of hedonic well-being, psychological empowerment and inspiration. *JMIR Mhealth Uhealth* 2018;6(6):0024. [doi: [10.2196/10024](#)]

Abbreviations

BMA: British Medical Association
GDC: General Dental Council
ROI: return on investment

Edited by G Eysenbach; submitted 13.02.18; peer-reviewed by T Cruvinel, P Ayala Aguirre; comments to author 01.03.18; revised version received 18.04.18; accepted 18.06.18; published 29.06.18.

Please cite as:

Parmar N, Dong L, Eisingerich AB

Connecting With Your Dentist on Facebook: Patients' and Dentists' Attitudes Towards Social Media Usage in Dentistry
J Med Internet Res 2018;20(6):e10109

URL: <http://www.jmir.org/2018/6/e10109/>

doi: [10.2196/10109](#)

PMID: [29959108](#)

©Nilesh Parmar, Lin Dong, Andreas Benedikt Eisingerich. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 29.06.2018. 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.

Original Paper

Increasing the Availability of Psychological Treatments: A Multinational Study of a Scalable Method for Training Therapists

Marianne O'Connor^{1*}, BA; Katy E Morgan², PhD; Suzanne Bailey-Straebler¹, DPhil, MSN; Christopher G Fairburn¹, FMedSci, FRCP; Zafra Cooper^{1,3*}, DPhil, Dip Clin Psych

¹Centre for Research on Dissemination at Oxford, Department of Psychiatry, University of Oxford, Oxford, United Kingdom

²Department of Medical Statistics, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom

³Department of Psychiatry, Yale School of Medicine, Yale University, New Haven, CT, United States

*these authors contributed equally

Corresponding Author:

Zafra Cooper, DPhil, Dip Clin Psych

Centre for Research on Dissemination at Oxford

Department of Psychiatry

University of Oxford

Warneford Lane

Headington

Oxford, OX3 7JX

United Kingdom

Phone: 44 2038094213

Email: zafra.cooper@psych.ox.ac.uk

Abstract

Background: One of the major barriers to the dissemination and implementation of psychological treatments is the scarcity of suitably trained therapists. A highly scalable form of Web-centered therapist training, undertaken without external support, has recently been shown to have promise in promoting therapist competence.

Objective: The aim of this study was to conduct an evaluation of the acceptability and effectiveness of a scalable independent form of Web-centered training in a multinational sample of therapists and investigate the characteristics of those most likely to benefit.

Methods: A cohort of eligible therapists was recruited internationally and offered access to Web-centered training in enhanced cognitive behavioral therapy, a multicomponent, evidence-based, psychological treatment for any form of eating disorder. No external support was provided during training. Therapist competence was assessed using a validated competence measure before training and after 20 weeks.

Results: A total of 806 therapists from 33 different countries expressed interest in the study, and 765 (94.9%) completed a pretraining assessment. The median number of training modules completed was 15 out of a possible 18 (interquartile range, IQR: 4-18), and 87.9% (531/604) reported that they treated at least one patient during training as recommended. Median pretraining competence score was 7 (IQR: 5-10, range: 0-19; N=765), and following training, it was 12 (IQR: 9-15, range: 0-20; N=577). The expected change in competence scores from pretraining to posttraining was 3.5 (95% CI 3.1-3.8; $P<.001$). After training, 52% (300/574) of therapists with complete competence data met or exceeded the competence threshold, and 45% (95% CI 41-50) of those who had not met this threshold before training did so after training. Compliance with training predicted both an increase in competence scores and meeting or exceeding the competence threshold. Expected change in competence score increased for each extra training module completed (0.19, 95% CI 0.13-0.25), and those who treated a suitable patient during training had an expected change in competence score 1.2 (95% CI 0.4-2.1) points higher than those who did not. Similarly, there was an association between meeting the competence threshold after training and the number of modules completed (odds ratio, OR=1.11, 95% CI 1.07-1.15), and treating at least one patient during training was associated with competence after training (OR=2.2, 95% CI 1.2-4.1).

Conclusions: Independent Web-centered training can successfully train large numbers of therapists dispersed across a wide geographical area. This finding is of importance because the availability of a highly scalable method of training potentially increases the number of people who might receive effective psychological treatments.

KEYWORDS

internet; web-centered; eating disorders; cognitive therapy; effective treatment

Introduction

In recent years, there has been considerable progress in developing evidence based psychological treatments. This progress has highlighted two important issues: the need to disseminate these treatments developed in controlled settings to routine clinical care [1,2] and the need to ensure that more people in need of care actually receive effective treatment [3].

Psychological treatments are difficult to disseminate [3,4] and implement widely. One of the major barriers to their dissemination and implementation is the scarcity of suitably trained therapists [5]. The currently accepted method of training typically involves attending a specialist workshop, reading relevant texts, and a subsequent period of supervision from someone expert in the treatment [6]. As this method is both labor-intensive and costly [7,8] and vulnerable to the shortage of treatment experts, it limits the number of therapists that can be trained, and therefore, the number of people who might potentially receive effective treatment.

There has been growing interest in addressing the problem of the scalability of training by using the internet to train therapists [9,10]. A recent systematic review has provided preliminary support for Web-based training methods while noting significant methodological shortcomings in many studies [11], particularly the use of outcome measures developed specifically for a particular study without the assessment and reporting of their psychometric properties [12].

Web-based training methods have a number of potential advantages. Training can be offered simultaneously to large numbers of geographically dispersed trainees with training materials that can be accessed at any time and from any place. Trainees can review and revisit material in a way that potentially reinforces learning and prevents subsequent therapist drift [13,14]. In addition, trainees can view detailed clinical illustrations of treatment interventions and complete interactive formative assessments such as knowledge tests that facilitate experiential learning. Furthermore, the training program can be updated regularly to incorporate new information. Finally, the training process can be informed by and improved upon by the collection of data on website usage.

We have developed and tested a specific form of Web-centered therapist training [15,16]. It differs from conventional training in that the training is fully automated with the expertise residing within the program rather than being provided by an outside expert. Web-centered training can be undertaken completely autonomously (independent training) or with support from a nonspecialist support worker whose role is to encourage the trainee to follow the program. An initial proof-of-concept study of the supported form of Web-centered training found that the method was feasible and acceptable to trainees across Ireland and was effective in promoting therapist competence, with 43% of trainee therapists scoring above the threshold of a validated

measure of therapist competence after training [16]. This finding was replicated and extended in a randomized controlled comparison that recruited therapists across the United States and Canada and compared the independent and supported form of training [17]. Both methods of training were effective in improving therapist competence, with 45% of therapists becoming competent after training and no clear evidence of a change in competence scores between the end of training and 6-month follow-up.

This study had two aims: (1) to determine whether our previous finding concerning acceptability and efficacy of the most highly scalable form of training (independent training) could be replicated in a multinational sample of therapists who wished to be trained and (2) to investigate the characteristics of the therapists who were most likely to benefit from this form of training.

Methods

Design

A cohort of eligible therapists was offered access to the 20-week, Web-centered training program. No external support was provided during the training. Therapist competence was assessed before and after the training.

The research protocol was submitted to Oxford University Central Research Ethics Committee. As the intervention was judged to be educational rather than clinical, the committee decided that formal ethical approval was not required.

Recruitment

Participants were recruited internationally through a posting on the research group's website offering free training in enhanced cognitive behavioral therapy (CBT-E), a multicomponent evidence-based psychological treatment for any form of eating disorder [18,19]. The posting included a link to an online description of the training and the research study. Potential participants had to be licensed mental health professionals who were prepared to take part in the research evaluating this training and who provided informed digital consent.

It was strongly recommended that participants met the following eligibility criteria: had been previously trained in delivering a short-term psychological treatment, were currently working with people with eating disorders, were willing to devote at least 9 hours to the training program, and were able to treat one or more suitable patients using CBT-E during the 20-week period of training. In the information provided to participants, it was stressed that clinical responsibility for their patients remained with their local clinical team and not with the researchers.

Participants were asked to complete a brief online questionnaire asking about their professional background, age, gender, and clinical experience. They also completed an online therapist

competence assessment measure (see below). They were subsequently sent a link to the training website together with instructions about how to use the training program. In addition, they were sent brief information about the minimum technical specifications for accessing and using the website.

Recruitment took place over 18 months from August 2016 to January 2017.

The Training Program

The CBT-E Web-centered training program has two main parts: The Course and The Library. Fuller details are provided elsewhere [16], and a summary of the content of the training is provided in [Multimedia Appendix 1](#). Briefly, the Course is linear in nature and takes between 8 and 9 hours to complete. It is a detailed practical description of how to implement the main focused form of CBT-E given by an expert on the treatment (CGF). This description is delivered in the form of multiple brief video presentations accompanied by handouts and interspersed with formative learning exercises, video recordings of acted illustrations of the treatment, and tests of knowledge together with feedback. While working through The Course, trainees are encouraged to read relevant sections from the treatment manual [20] and treat one or more patients.

The second part of the training website, The Library, contains all the material in The Course, including the handouts, learning exercises, and clinical illustrations in indexed form. In addition, there is a large amount of supplementary material on how to use CBT-E with specific subgroups of patients including adolescents; those who are severely underweight; and those with clinical perfectionism, core low self-esteem, or marked interpersonal difficulties. Participants were granted access to The Course and core Library material from the start of training. They only had access to the supplementary Library material once they had completed the study.

Trainees were not provided with any external support while completing the training. They received reminder emails 6, 10, 14, and 18 weeks after starting, informing them of the number of weeks of training that had elapsed and the number of weeks remaining.

Assessment

Participants' competence at delivering CBT-E, ie, their knowledge of how to deliver the treatment skilfully [7], was assessed before training and immediately following training. It was measured using a scalable online measure with sound psychometric properties that had previously been developed independently of the creation of the training website. Its development and validation are described in detail in a separate publication [21]. This included detailed blueprinting, state-of-the-art item writing, independent item review, and initial field-testing, followed by formal Rasch analysis to test for good model fit. Strict criteria of unidimensionality were met by stepwise exclusion of misfitting items until there was no individual item misfitting at $P < .01$. The resulting measure consists of 22 items addressing trainees' knowledge and understanding of CBT-E and its implementation (ie, applied knowledge). The instrument generates a total score (out of a possible 22), and trainees can be classified as scoring at or above

the previously established cut point. This was established using receiver operator characteristic analyses to determine the "best cut point" from the values of sensitivity and specificity calculated at increasing test score cut points. Three equivalent versions of the competence measure were developed so that different versions could be used on different assessment occasions as was done in this study. Those who score above the competence threshold (an equated score of 12 or above) have also been shown to be competent at implementing CBT-E when systematically observed and rated using a performance based measure [22].

Data Analysis

Missing values were imputed using multiple imputation (see [Multimedia Appendix 2](#)). Those with a nonmissing pretraining score, gender, and age were used for all analyses based on imputed data ($N=760$). Fifty imputations were created, and regression models were applied to each imputed dataset, and the estimates from each imputation were combined using Rubin's rules [23]. Imputed values for posttraining score were derived from the imputed values for change in score.

Linear regression models were used to investigate changes in score after training and logistic regression models to explore scoring above the competence cut point after training.

To examine predictors of change in score and competence levels after training, linear and logistic models were fitted adjusting for age and gender and, additionally, in each case, a second set of models were fitted adjusting for pretraining score. The variables examined, included in the models one at a time, were as follows: trainee characteristics (professional background, years of clinical experience, weekly time spent treating patients having attended a training workshop, and using English to treat patients) and training adherence (number of training modules completed, treating at least one suitable patient during training, and number of patients treated during training). The number of modules completed and treating a suitable patient were also examined together to explore the associations for each of these variables when adjusting for the other. Reported results use imputed data unless otherwise stated. Preliminary analyses explored the impact of country of recruitment on change in score and, as no effect was found, it was not included in further analyses. After exploring various distributional forms to describe the number of training modules completed, it was decided to use this variable as a linear term.

To explore attrition, univariate logistic regression models were fitted to examine associations between a binary variable for having a missing posttraining test score and a range of trainee adherence variables and trainee characteristics as detailed above.

All analyses were conducted in Stata 15 (StataCorp LLC), with multiple imputation performed using the `smcfc` command [24].

Results

Recruitment

A total of 806 therapists expressed an interest in participating in the study and completed the consent form. Of these 765 (94.9%, 765/806) completed the pretraining assessment and

were given access to the training website. [Figure 1](#) shows their progress through the study.

The median age of the participants was 36 years (interquartile range, IQR: 31-44; range: 22-67), and 88.7% were female (691/779). As regards professional training, those with a clinical psychology background were the overwhelming majority at

47.7% (373/782), followed by psychiatric nurses comprising 8.6% (67/782) and those from a social work background at 7.8% (61/782). The remainder came from a variety of other backgrounds including counselling, CBT therapist training, family therapy, psychiatry, nursing, and dietetics. Further details about their professional background, clinical experience, current clinical practice, and training can be seen in [Table 1](#).

Figure 1. Flow of the participants through the study.

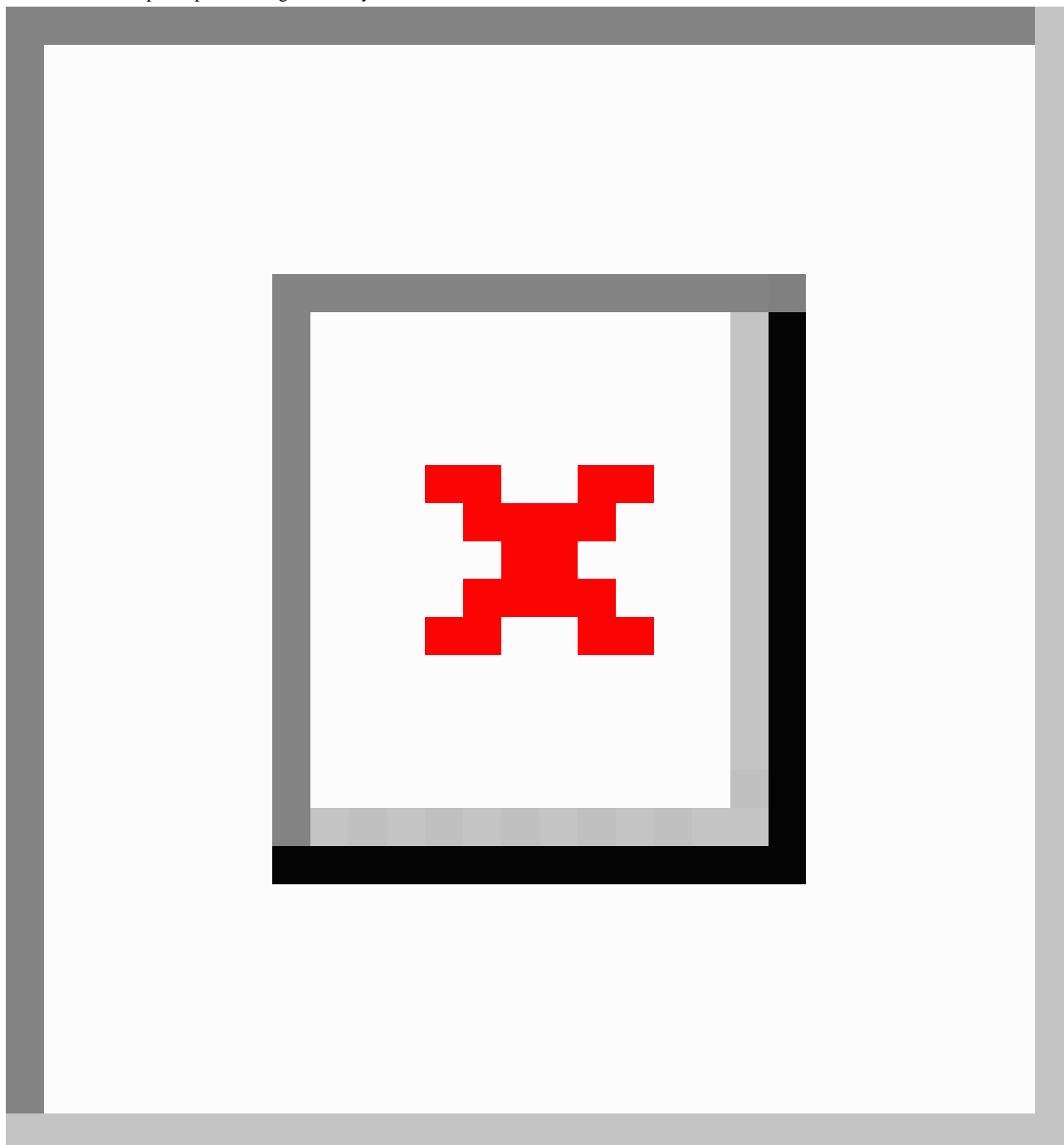


Table 1. Background characteristics of trainee therapists.

Characteristics	Statistics
Gender (N=779), n (%)	
Female	691 (88.7)
Male	88 (11.3)
Age in years (N=780), median (IQR ^a ; range)	36 (31-44; 22-67)
Professional background (N=782), n (%)	
Clinical psychologist PhD	99 (12.7)
Clinical psychologist PsyD	165 (21.1)
Clinical psychologist Masters	109 (13.9)
Psychiatric nurse	67 (8.6)
Social worker	61 (7.8)
Other	281 (35.9)
Hours per week seeing patients (N=729), median (IQR; range)	10 (4-17; 0-40)
Years of clinical experience (N=718), median (IQR; range)	6 (3-10; 0-45)
Previously attended a CBT-E^b workshop (N=778), n (%)	
Yes	212 (27.3)
No	566 (72.8)
Primary language for treating patients (N=766), n (%)	
English	580 (75.7)
Other	186 (24.3)

^aIQR: interquartile range.

^bCBT-E: enhanced cognitive behavioral therapy.

The trainee therapists were located in 33 different countries. The majority, 76.6% (617/806), came from English-speaking countries (United Kingdom, United States, Australia, Canada, New Zealand, and Ireland, respectively), followed by 5.8% (47/806) from the Netherlands, 4.1% (33/806) from Sweden, 3.7% (30/806) from Italy, 2.6% (21/806) from Norway, and 7.2% (58/806) from other countries.

Training Completion

The median number of modules of the training program completed was 15 out of a possible 18 (IQR: 4-18), and 87.9% (531/604) of the trainees reported that they treated at least one suitable patient during the training with 42.9% (259/604) reporting that they treated two or more.

Outcome of Training

The median pretraining competence score was 7 (IQR: 5-10, range: 0-19; N=765), and following training, it was 12 (IQR: 9-15, range: 0-20; N=577). The expected change in competence scores from pretraining to posttraining was 3.5 (95% CI 3.1-3.8; $P<.001$).

After training, 52% (300/574) of therapists with complete competence data met or exceeded the competence threshold. Of those with complete data, 45% (213/471, 95% CI 41-50%) who did not meet the competence threshold before training did so after training, and 16% (16/103, 95% CI 9-24%) who did meet it before training no longer met it after training.

Predictors of Training Outcome

Change in Competence Score

As can be seen in Table 2 age, gender, years of clinical experience, time spent treating patients each week, and professional background were not associated with changes in expected competence scores whether adjusting for pretraining score or not.

When adjusting for pretraining scores, neither prior workshop attendance nor using English to treat patients was associated with expected change in competence scores. When unadjusted for pretraining scores prior workshop attendance was associated with a reduction of 1.6 points in expected change (95% CI 0.9-2.3, $P<.001$) and using English to treat patients was associated with a 0.9 increase in expected change (95% CI 0.2-1.6, $P=.02$), suggesting that these differences were primarily due to an association with pretraining score. Similarly, when adjusting for pretraining score, the association between number of patients treated while training was not statistically significant ($P=.16$), although when unadjusted, there was borderline evidence supporting this association with change in score ($P=.03$).

There was strong evidence of an increase in expected change in competence score for each extra training module completed, whether or not adjustments were made for pretraining score (0.14 [95% CI 0.07-0.20] and 0.19 [95% CI 0.13-0.25] points

per module, respectively). Treating a suitable patient while training was associated with change in competence score when adjusting for pretraining score with those who treated such a patient having a change in score 1.2 (95% CI 0.4-2.1) points higher than those who did not. The effect sizes did not appear to change when the number of modules completed and treating a suitable patient were examined together (0.18 [95% CI 0.13-0.23, $P<.001$] and 1.3 [95% CI 0.5-2.2, $P=.001$], respectively; $N=601$ therapists with nonmissing data).

Scoring Above the Competence Threshold

An examination of the associations between trainee characteristics and training adherence variables and scoring above the competence threshold yielded a pattern of results similar to those for the predictors of expected change in score.

Age, gender, years of clinical experience, and time spent treating patients each week were not associated with competence after training, whether adjusting for pretraining score or not (see [Table 3](#)).

When adjusting for pretraining score, there was no evidence for an association between professional background and competence

($P=.22$). When unadjusted for pretraining score, professional background was associated with competence after training ($P<.001$), suggesting that this may be due to differences in pretraining score by professional background. Similarly, there was no evidence of an association between prior workshop attendance or using English to treat patients and posttraining competence when adjusting for pretraining score. When unadjusted, attending a workshop increased the odds of achieving competence after training by 3.2 (95% CI 2.2-4.6), and using English decreased the odds by 0.5 (95% CI 0.4-0.8).

Again, as for the predictors of change in competence score, there was strong evidence of an association between being above the competence threshold after training and the number of modules completed with the odds ratio (OR) when adjusting for pretraining scores being very similar to that when unadjusted for pretraining score (OR=1.11, 95% CI 1.07-1.15 and OR=1.11, 95% CI 1.08-1.15, respectively; see [Table 3](#)). When adjusting for pretraining score, there was evidence that treating at least one suitable patient while training was associated with competence after training (OR=2.2, 95% CI 1.2-4.1) and only borderline evidence for an association with number of patients treated ($P=.05$).

Table 2. Predictors of change in competence score after training.

Trainee characteristic or training compliance	Unadjusted for pretraining score		Adjusted for pretraining score	
	Difference in expected change score (95% CI)	P value	Difference in expected change score (95% CI)	P value
Age (years)	0.00 (−0.03 to 0.03)	.98	−0.01 (−0.04 to 0.02)	.52
Gender				
Female	—	—	—	—
Male	0.1 (−0.9 to 1.1)	.87	0.1 (−0.9 to 1.0)	.91
Clinical experience (years)	0.04 (−0.02 to 0.09)	.24	0.04 (−0.01 to 0.10)	.11
Weekly time treating patients (hours)	−0.03 (−0.06 to 0.01)	.10	−0.01 (−0.04 to 0.02)	.54
Professional background^a				
Clinical psychology				
PsyD	—	—	—	—
PhD	1.3 (0.2-2.4) ^b	.02	0.7 (−0.2 to 1.7) ^b	.14
Masters	0.3 (−0.8 to 1.3) ^b	.64	−0.3 (−1.3 to 0.6) ^b	.51
Psych nurse	0.3 (−1.0 to 1.6) ^b	.62	−0.9 (−2.1 to 0.2) ^b	.11
Social work	0.4 (−0.9 to 1.7) ^b	.56	−0.9 (−2.1 to 0.3) ^b	.14
Other	0.5 (−0.3 to 1.4) ^b	0.21	−0.4 (−1.2 to 0.4) ^b	.30
Treating patients using English	0.9 (0.2-1.6)	.02	−0.1 (−0.8 to 0.5)	.72
Previous attendance at workshop	−1.6 (−2.3 to −0.9)	<.001	0.2 (−0.5 to 0.9)	.54
Number of patients treated during training (n=572)^{c,d}				
No patients	—	—	—	—
One patient	−0.1 (−1.4 to 1.3) ^e	.94	0.7 (−0.4 to 1.9) ^e	.21
Two or more patients	−0.5 (−1.8 to 0.7) ^e	.40	0.9 (−0.2 to 2.0) ^e	.10
One or more part cases	0.6 (−0.7 to 1.9) ^e	.36	1.3 (0.1-2.4) ^e	.03
Number of training modules completed ^f	0.14 (0.07-0.20)	<.001	0.19 (0.13-0.25)	<.001
Treating at least one suitable patient (n=572) ^c	0.4 (−0.6 to 1.4)	.43	1.2 (0.4-2.1)	.004

^aOverall P value for including professional background when unadjusted for pretraining score, $P=.34$ and when adjusted for pretraining score, $P=.06$.

^bEffect relative to clinical psychology (PsyD).

^cAnalyses conducted on complete cases only.

^dOverall P value for including number of patients; $P=.03$ when unadjusted for pretraining score and $P=.16$ when adjusted.

^eEffect relative to treating no patients.

^fNumber of modules completed included as a linear term.

Predictors of Training and Study Dropout

About a quarter of therapists who completed initial study assessments (186/760, 24.5%) did not complete competence assessments after training. Not surprisingly, the number of training modules completed was strongly associated with the odds of having a missing posttraining assessment, decreasing by over 25% (OR=0.73, 95% CI 0.70-0.76; $P<.001$) for every extra module viewed. Pretraining competence score was also strongly associated with having a missing posttraining score, with the odds decreasing by around 10% for every extra point

scored at pretraining (OR=0.89, 95% CI 0.85-0.94, $P<.001$). When adjusting the model to take account of the number of modules completed, the OR for the number of training modules completed remains almost identical (OR=0.73, 95% CI 0.70-0.77; $P<.001$), whereas the OR for pretraining score is greatly attenuated (OR=0.94, 95% CI 0.88-1.00; $P=.07$), suggesting that the decrease in missing posttraining scores for those with higher pretraining scores may be partially mediated through the number of modules completed. For further details of trainee background characteristics associated with missing posttraining scores, see [Multimedia Appendix 3](#).

Table 3. Predictors of competence after training.

Trainee characteristic or training compliance	Unadjusted for pretraining score		Adjusted for pretraining score	
	OR ^a (95% CI)	P value	OR (95% CI)	P value
Age (years)	0.99 (0.97-1.01)	.23	0.99 (0.98-1.01)	.57
Gender				
Female	—	—	—	—
Male	0.96 (0.59-1.58)	.89	0.94 (0.53-1.67)	.84
Clinical experience (years)	1.03 (1.00-1.06)	.10	1.03 (0.99-1.06)	.12
Weekly time treating patients (hours)	1.01 (1.00-1.03)	.13	1.00 (0.98-1.02)	.80
Professional background^b				
Clinical psychology				
PsyD	—	—	—	—
PhD	0.8 (0.4-1.3) ^c	.37	1.0 (0.5-1.9) ^c	.92
Masters	0.5 (0.3-0.9) ^c	.03	0.7 (0.3-1.2) ^c	.18
Psych nurse	0.3 (0.1-0.6) ^c	<.001	0.5 (0.2-1.1) ^c	.07
Social worker	0.3 (0.2-0.6) ^c	.001	0.5 (0.2-1.1) ^c	.09
Other	0.5 (0.3-0.8) ^c	.002	0.7 (0.5-1.2) ^c	.25
Treating patients using English	0.5 (0.4-0.8)	.001	0.9 (0.6-1.4)	.72
Previous attendance at workshop	3.2 (2.2-4.6)	<.001	1.3 (0.8-2.0)	.25
Number of patients treated during training (n=572)^{d,e}				
No patients	—	—	—	—
One patient	1.8 (0.8-3.9) ^f	.13	1.2 (0.5-2.7) ^f	.67
Two or more patients	3.5 (1.7-7.2) ^f	.001	1.9 (0.9-4.1) ^f	.11
One or more part cases	2.7 (1.3-5.7) ^f	.009	2.2 (1.0-4.8) ^f	.06
Number of training modules completed ^g	1.11 (1.08-1.15)	<.001	1.11 (1.07-1.15)	<.001
Treating at least one suitable patient (n=572) ^d	3.0 (1.7-5.2)	<.001	2.2 (1.2-4.1)	.009

^aOR: odds ratio.^bOverall value for professional background when unadjusted for pretraining score: $P<.001$ and when adjusted for pretraining score: $P=.22$.^cEffect relative to clinical psychology (PsyD).^dAnalyses conducted on complete cases only.^eOverall value for number of patients seen when unadjusted for pretraining score: $P<.001$ and when adjusted for pretraining score: $P=.05$.^fEffect relative to treating no patients.^gNumber of modules completed included as a linear term.

Discussion

Principal Findings

The study had two aims. The first was to determine whether earlier findings concerning the acceptability and effectiveness of a highly scalable form of Web-centered training (independent training) could be replicated in a multinational sample of therapists. This was found to be the case with the present findings replicating the earlier ones. Over 70% of the therapists enrolled in the study provided end of training competence data, the great majority of the training modules were completed, and most therapists treated a suitable patient while training in

compliance with the recommendations of the program. Scores on the validated competence measure increased significantly, and 45% of the trainee therapists who had not previously been competent achieved competence scores indicative of a good level of competence. This figure is identical to that obtained in the study of therapists in the United States and Canada [17].

The second aim was to investigate the characteristics of the therapists who were most likely to benefit from this form of training. Compliance with training recommendations was a predictor of both an increase in competence score and becoming competent. There was a linear relationship between the number of training modules completed and changes in competence score,

and the likelihood of therapists scoring above the competence threshold increased with the number of modules they completed. Similarly, treating at least one patient while training was associated with both an increased change in competence score and the likelihood of scoring over the competence threshold.

It was less clear from these findings why some therapists complied with the training recommendations, thereby deriving benefit, while others did not. An examination of those who were study noncompleters seems to suggest that any efforts to improve training should focus on ways of increasing training completion in those groups at risk of not doing so. It is possible that for these groups, the addition of a more generic introduction to the treatment method to complement the focus on specific specialist skills related to providing evidence-based CBT-E might be beneficial. Such an addition might also allow the training to be extended to an even wider range of potential therapists.

Comparisons With Other Studies

There has been limited research on the outcome of therapist training against which to compare the present findings [8,13,25]. In general, therapist training has been relatively neglected as a research topic until recently [2,26], and Web-based training is an even more novel focus of interest [11].

Competence figures that have been reported following more conventional training in psychological treatment for depression range from 21% after attending a training workshop to 96% after extensive consultation with an expert including treatment session review and feedback [27]. A study of community clinicians receiving training in transdiagnostic cognitive behaviour therapy reported 59.5% of clinicians competent after training [28]. However, the latter training also involved extensive expert consultation and session review of a kind that is not scalable; thus, the findings are not directly comparable. Although positive changes in knowledge and skill have been reported for Web-based training, studies have employed widely varying methods, and few have employed standardized outcome measures, making it difficult to draw firm conclusions [11,12]. One study of Web-based training in transdiagnostic cognitive behavior therapy for community clinicians that used an established measure reported that 48% reached competency, but the training also included 7.5 months of face-to-face consultation with previously trained peers and 6.5 hours of expert instructor time [29]. Clearly, these findings are not

directly comparable with the independent training investigated in this study.

Study Strengths

The study had a number of strengths. First, a large number of trainees from 33 countries were recruited and trained. Second, nearly three-quarters of those who consented to participate in the study completed competence assessments at the end of training. Third, the study used a previously validated measure of therapist competence that had an empirically established competence threshold. Fourth, the study was able to explore predictors of change in competence scores and the achievement of competence, as well as factors associated with the noncompletion of the study and training.

Study Limitations

The study also had certain limitations. First, it did not include a no-training control condition or a delayed training group. Thus, we cannot discount the possibility that competence scores would have increased over time without any training, but this seems unlikely especially given the strong association between module completion and competence scores. Second, despite good retention in the study, there was still a significant amount of missing data, necessitating the use of multiple imputation that relies on certain assumptions about the nature of the missing data and, in some cases, the use of analyses based on complete data only. Attrition, both in the form of participants ceasing to use an intervention and not completing study assessments, has been noted as a particular problem in internet interventions [30]. Third, as clinicians chose to undertake the training, generalization to those less motivated may not be warranted. Fourth, although we have reported competence data concerning clinicians' ability to provide CBT-E treatment, we do not have data concerning the quality of the treatment they actually provide.

Conclusions

This study confirms that independent Web-centered training can successfully train a large number of therapists dispersed across a wide geographical area. This finding is of great practical importance because it indicates that this form of independent Web-centered training is able to overcome a major barrier to the dissemination and implementation of psychological treatments. The availability of a highly scalable training method potentially greatly increases the number of people who might have access to effective psychological treatments.

Acknowledgments

The research was funded by a strategic award from the Wellcome Trust (094585: CGF and ZC). CGF held a Principal Research Fellowship from the Wellcome Trust (046386). The Wellcome Trust had no involvement with the design or conduct of the study or the preparation of this manuscript.

Conflicts of Interest

CGF receives royalties from the treatment guide.

Multimedia Appendix 1

Enhanced cognitive behavioral therapy (CBT-E) training program.

[[PDF File \(Adobe PDF File\), 24KB - jmir_v20i6e10386_app1.pdf](#)]

Multimedia Appendix 2

Data imputation.

[[PDF File \(Adobe PDF File\), 12KB - jmir_v20i6e10386_app2.pdf](#)]

Multimedia Appendix 3

Predictors of training and study dropout.

[[PDF File \(Adobe PDF File\), 19KB - jmir_v20i6e10386_app3.pdf](#)]

References

1. Shafran R, Clark DM, Fairburn CG, Arntz A, Barlow DH, Ehlers A, et al. Mind the gap: improving the dissemination of CBT. *Behav Res Ther* 2009 Nov;47(11):902-909. [doi: [10.1016/j.brat.2009.07.003](#)] [Medline: [19664756](#)]
2. Harvey AG, Gumport NB. Evidence-based psychological treatments for mental disorders: modifiable barriers to access and possible solutions. *Behav Res Ther* 2015 May;68:1-12. [doi: [10.1016/j.brat.2015.02.004](#)] [Medline: [25768982](#)]
3. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. *Behav Res Ther* 2017 Jan;88:7-18. [doi: [10.1016/j.brat.2016.06.004](#)] [Medline: [28110678](#)]
4. Fairburn CG, Patel V. The global dissemination of psychological treatments: a road map for research and practice. *Am J Psychiatry* 2014 May;171(5):495-498. [doi: [10.1176/appi.ajp.2013.13111546](#)] [Medline: [24788281](#)]
5. Kazdin AE, Fitzsimmons-Craft EE, Wilfley DE. Addressing critical gaps in the treatment of eating disorders. *Int J Eat Disord* 2017 Mar;50(3):170-189. [doi: [10.1002/eat.22670](#)] [Medline: [28102908](#)]
6. Herschell AD, Kolko DJ, Baumann BL, Davis AC. The role of therapist training in the implementation of psychosocial treatments: a review and critique with recommendations. *Clin Psychol Rev* 2010 Jun;30(4):448-466 [FREE Full text] [doi: [10.1016/j.cpr.2010.02.005](#)] [Medline: [20304542](#)]
7. Fairburn CG, Cooper Z. Therapist competence, therapy quality, and therapist training. *Behav Res Ther* 2011 Jun;49(6-7):373-378 [FREE Full text] [doi: [10.1016/j.brat.2011.03.005](#)] [Medline: [21492829](#)]
8. Martino S, Paris M, Añez L, Nich C, Canning-Ball M, Hunkele K, et al. The effectiveness and cost of clinical supervision for motivational interviewing: a randomized controlled trial. *J Subst Abuse Treat* 2016 Sep;68:11-23. [doi: [10.1016/j.jsat.2016.04.005](#)] [Medline: [27431042](#)]
9. Khanna MS, Kendall PC. Bringing technology to training: web-based therapist training to promote the development of competent cognitive-behavioral therapists. *Cogn Behav Pract* 2015 Aug;22(3):291-301. [doi: [10.1016/j.cbpra.2015.02.002](#)]
10. McMillen JC, Hawley KM, Proctor EK. Mental health clinicians' participation in web-based training for an evidence supported intervention: signs of encouragement and trouble ahead. *Adm Policy Ment Health* 2016 Jul;43(4):592-603. [doi: [10.1007/s10488-015-0645-x](#)] [Medline: [25822326](#)]
11. Jackson CB, Quetsch LB, Brabson LA, Herschell AD. Web-based training methods for behavioral health providers: a systematic review. *Adm Policy Ment Health* 2018 Jan 19:- Epub ahead of print(forthcoming). [doi: [10.1007/s10488-018-0847-0](#)] [Medline: [29352459](#)]
12. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. *Implement Sci* 2014 Sep 04;9:118 [FREE Full text] [doi: [10.1186/s13012-014-0118-8](#)] [Medline: [25185799](#)]
13. Waller G, Turner H. Therapist drift redux: why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behav Res Ther* 2016 Feb;77:129-137. [doi: [10.1016/j.brat.2015.12.005](#)] [Medline: [26752326](#)]
14. Fairburn CG, Wilson GT. The dissemination and implementation of psychological treatments: problems and solutions. *Int J Eat Disord* 2013 Jul;46(5):516-521 [FREE Full text] [doi: [10.1002/eat.22110](#)] [Medline: [23658103](#)]
15. Cooper Z, Bailey-Straebl S. Disseminating evidence-based psychological treatments for eating disorders. *Curr Psychiatry Rep* 2015 Mar;17(3):551 [FREE Full text] [doi: [10.1007/s11920-015-0551-7](#)] [Medline: [25663154](#)]
16. Fairburn CG, Allen E, Bailey-Straebl S, O'Connor ME, Cooper Z. Scaling up psychological treatments: a countrywide test of the online training of therapists. *J Med Internet Res* 2017 Jun 16;19(6):e214 [FREE Full text] [doi: [10.2196/jmir.7864](#)] [Medline: [28623184](#)]
17. Cooper Z, Bailey-Straebl S, Morgan KE, O'Connor ME, Caddy C, Hamadi L, et al. Using the Internet to train therapists: randomized comparison of two scalable methods. *J Med Internet Res* 2017 Oct 18;19(10):e355 [FREE Full text] [doi: [10.2196/jmir.8336](#)] [Medline: [29046265](#)]
18. Fairburn CG. *Cognitive Behavior Therapy and Eating Disorders*. New York: The Guilford Press; 2008.
19. *Eating Disorders: Recognition, Assessment and Treatment*. London, UK: National Institute for Health Care Excellence; May 2017.

20. Fairburn C, Cooper Z, Shafran R, Bohn K, Hawker D, Murphy R. Enhanced cognitive behavior therapy for eating disorders: the core protocol. New York: The Guilford Press; 2008.
21. Cooper Z, Doll H, Bailey-Straebl S, Klucznik D, Murphy R, O'Connor ME, et al. The development of an online measure of therapist competence. *Behav Res Ther* 2015 Jan;64:43-48 [FREE Full text] [doi: [10.1016/j.brat.2014.11.007](https://doi.org/10.1016/j.brat.2014.11.007)] [Medline: [25528502](https://pubmed.ncbi.nlm.nih.gov/25528502/)]
22. Cooper Z, Doll H, Bailey-Straebl S, Bohn K, de Vries D, Murphy R, et al. Assessing therapist competence: development of a performance-based measure and its comparison with a web-based measure. *JMIR Ment Health* 2017 Oct 31;4(4):e51 [FREE Full text] [doi: [10.2196/mental.7704](https://doi.org/10.2196/mental.7704)] [Medline: [29089289](https://pubmed.ncbi.nlm.nih.gov/29089289/)]
23. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons; 1987.
24. Bartlett J, Morris TP. Multiple imputation of covariates by substantive-model compatible fully conditional specification. *Stata J* 2015;15(2):437-456.
25. Waller G. Treatment protocols for eating disorders: clinicians' attitudes, concerns, adherence and difficulties delivering evidence-based psychological interventions. *Curr Psychiatry Rep* 2016 Apr;18(4):36 [FREE Full text] [doi: [10.1007/s11920-016-0679-0](https://doi.org/10.1007/s11920-016-0679-0)] [Medline: [26893234](https://pubmed.ncbi.nlm.nih.gov/26893234/)]
26. Rosen RC, Ruzek JI, Karlin BE. Evidence-based training in the era of evidence-based practice: challenges and opportunities for training of PTSD providers. *Behav Res Ther* 2017 Jan;88:37-48. [doi: [10.1016/j.brat.2016.07.009](https://doi.org/10.1016/j.brat.2016.07.009)] [Medline: [28110675](https://pubmed.ncbi.nlm.nih.gov/28110675/)]
27. Walser RD, Karlin BE, Trockel M, Mazina B, Barr Taylor C. Training in and implementation of acceptance and commitment therapy for depression in the Veterans Health Administration: therapist and patient outcomes. *Behav Res Ther* 2013 Sep;51(9):555-563. [doi: [10.1016/j.brat.2013.05.009](https://doi.org/10.1016/j.brat.2013.05.009)] [Medline: [23851161](https://pubmed.ncbi.nlm.nih.gov/23851161/)]
28. Creed TA, Frankel SA, German RE, Green KL, Jager-Hyman S, Taylor KP, et al. Implementation of transdiagnostic cognitive therapy in community behavioral health: the Beck Community initiative. *J Consult Clin Psychol* 2016 Dec;84(12):1116-1126. [doi: [10.1037/ccp0000105](https://doi.org/10.1037/ccp0000105)] [Medline: [27379492](https://pubmed.ncbi.nlm.nih.gov/27379492/)]
29. German RE, Adler A, Frankel SA, Stirman SW, Pinedo P, Evans AC, et al. Testing a web-based, trained-peer model to build capacity for evidence-based practices in community mental health systems. *Psychiatr Serv* 2018 Mar 01;69(3):286-292. [doi: [10.1176/appi.ps.201700029](https://doi.org/10.1176/appi.ps.201700029)] [Medline: [29137558](https://pubmed.ncbi.nlm.nih.gov/29137558/)]
30. Eysenbach G. The law of attrition. *J Med Internet Res* 2005 Mar 31;7(1):e11 [FREE Full text] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](https://pubmed.ncbi.nlm.nih.gov/15829473/)]

Abbreviations

CBT-E: enhanced cognitive behavioral therapy

IQR: interquartile range

OR: odds ratio

Edited by G Eysenbach; submitted 12.03.18; peer-reviewed by J Ruzek, G Waller; comments to author 29.03.18; revised version received 07.04.18; accepted 08.04.18; published 08.06.18.

Please cite as:

O'Connor M, Morgan KE, Bailey-Straebl S, Fairburn CG, Cooper Z

Increasing the Availability of Psychological Treatments: A Multinational Study of a Scalable Method for Training Therapists

J Med Internet Res 2018;20(6):e10386

URL: <http://www.jmir.org/2018/6/e10386/>

doi: [10.2196/10386](https://doi.org/10.2196/10386)

PMID: [29884606](https://pubmed.ncbi.nlm.nih.gov/29884606/)

©Marianne O'Connor, Katy E Morgan, Suzanne Bailey-Straebl, Christopher G Fairburn, Zafra Cooper. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.06.2018. 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.

Original Paper

Reproductive Health and Medication Concerns for Patients With Inflammatory Bowel Disease: Thematic and Quantitative Analysis Using Social Listening

Michelle Sophie Keller^{1,2,3}, MPH; Sasan Mosadeghi⁴, MS, MD; Erica R Cohen⁵, MD; James Kwan⁶, BS; Brennan Mason Ross Spiegel^{1,5,6}, MSHS, MD

¹Department of Health Policy and Management, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States

²Division of Informatics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, United States

³Division of General Internal Medicine, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, United States

⁴Department of Medicine, University of Arizona, Phoenix, AZ, United States

⁵Division of Gastroenterology, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, United States

⁶Cedars-Sinai Center for Outcomes Research and Education, Cedars-Sinai Medical Center, Los Angeles, CA, United States

Corresponding Author:

Michelle Sophie Keller, MPH

Department of Health Policy and Management

Fielding School of Public Health

University of California, Los Angeles

650 Charles Young Dr S

31-269 CHS Box 951772

Los Angeles, CA, 90095 1772

United States

Phone: 1 310 825 2594

Email: mkeller05@ucla.edu

Abstract

Background: Inflammatory bowel disease (IBD) affects many individuals of reproductive age. Most IBD medications are safe to use during pregnancy and breastfeeding; however, observational studies find that women with IBD have higher rates of voluntary childlessness due to fears about medication use during pregnancy. Understanding why and how individuals with IBD make decisions about medication adherence during important reproductive periods can help clinicians address patient fears about medication use.

Objective: The objective of this study was to gain a more thorough understanding of how individuals taking IBD medications during key reproductive periods make decisions about their medication use.

Methods: We collected posts from 3000 social media sites posted over a 3-year period and analyzed the posts using qualitative descriptive content analysis. The first level of analysis, open coding, identified individual concepts present in the social media posts. We subsequently created a codebook from significant or frequently occurring codes in the data. After creating the codebook, we reviewed the data and coded using our focused codes. We organized the focused codes into larger thematic categories.

Results: We identified 7 main themes in 1818 social media posts. Individuals used social media to (1) seek advice about medication use related to reproductive health (13.92%, 252/1818); (2) express beliefs about the safety of IBD therapies (7.43%, 135/1818); (3) discuss personal experiences with medication use (16.72%, 304/1818); (4) articulate fears and anxieties about the safety of IBD therapies (11.55%, 210/1818); (5) discuss physician-patient relationships (3.14%, 57/1818); (6) address concerns around conception, infertility, and IBD medications (17.38%, 316/1818); and (7) talk about IBD symptoms during and after pregnancy and breastfeeding periods (11.33%, 206/1818).

Conclusions: Beliefs around medication safety play an important role in whether individuals with IBD decide to take medications during pregnancy and breastfeeding. Having a better understanding about why patients stop or refuse to take certain medications during key reproductive periods may allow clinicians to address specific beliefs and attitudes during office visits.

(*J Med Internet Res* 2018;20(6):e206) doi:[10.2196/jmir.9870](https://doi.org/10.2196/jmir.9870)

KEYWORDS

pregnancy; breastfeeding; reproductive health; social media; medication adherence; infodemiology; pharmacovigilance

Introduction**Inflammatory Bowel Disease and Reproductive Health**

Inflammatory bowel disease (IBD) is a chronic, relapsing, and remitting autoimmune disorder comprising Crohn disease (CD) and ulcerative colitis (UC). There are over 1.6 million Americans living with IBD and as many as 70,000 new cases are diagnosed annually [1]. IBD is frequently diagnosed in the second to fourth decades of life, with the highest incidence between 20 and 29 years of age, during peak female reproductive years [2].

Treatment of IBD patients during their reproductive years provides a unique clinical challenge for clinicians, who seek to maintain remission during conception and pregnancy while ensuring the health of the fetus. Adverse outcomes such as preterm delivery, low birth weight, and increased risk of miscarriage are all associated with the degree of disease activity at the time of conception [3,4]. Women who plan their pregnancies while in remission are less likely to experience disease flares during pregnancy and have similar fertility rates as the general population [5]. A growing body of literature now suggests that the majority of IBD medications are low risk for use during pregnancy and even during lactation [6]. Despite this, observational studies note that women—particularly those with CD—have higher rates of voluntary childlessness due to concerns surrounding medication side effects, passing the disease to their offspring, fear of infertility, and advice given by treating physicians [7,8]. Moreover, many women perceive medications to be unsafe during pregnancy [9-11]. Consequently, women are uncertain about taking medications during pregnancy and many believe that medication use should be highly restricted during pregnancy, even when discontinuing medication use might threaten the health of the mother [10]. In one study, one-third of women believed that all medication use should be halted during pregnancy, and 20% of patients surveyed said they would stop medications even if they were advised by a physician to continue medication use [12]. Of the women surveyed, 68% reported anxiety or worries related to the effects of drugs on their pregnancy [12].

Social Listening and Pharmacovigilance

Social media use is widespread in the United States. More than two thirds of Americans report using Facebook every day, and the typical American is on 3 social media platforms [13]. Additionally, millions of individuals around the world use the internet to describe and share their health care experiences while seeking a sense of community for information-seeking purposes. A study reported that 34% of Americans reported reading about someone else's health care experience and 15% consulted online reviews of medical facilities [14]. Social listening, or examining social media and forum posts, can reveal trends in public attitudes and behaviors about various health care topics. For example, our group has used social listening to examine patient beliefs about the use of opioid medications and their gastrointestinal side effects, public opinions about the use of

virtual reality in health care, and patient beliefs about biologic medications in IBD [15-17]. Researchers have used social listening methods to examine e-cigarette attitudes and posts in the United States, estimate trends in influenza epidemics, and monitor the use and misuse of antidepressants [18-22]. This approach offers unique advantages compared with traditional interviews or focus groups, including the ability to capture discussions among social media users without a moderator present. As a result, social listening may capture opinions from individuals on sensitive topics or from those who might not necessarily choose to participate in a research study [23]. Participants also create communities within social media and network forums that draw patients from many different countries; gathering such a diverse and large group of participants for a focus group or interview-based study would be both logistically challenging and prohibitively expensive.

The percentage of adults who use at least 1 social media site has increased steadily since 2006. Individuals of certain demographic groups are more likely to use and engage in social media compared with other groups. One exception is race—approximately the same proportion of black, white, and Hispanic Americans use at least one social media site [24]. However, income, gender, and education have been shown to be associated with different levels of social media use. A larger proportion of individuals making more than \$75,000 per year (77%) reported using at least 1 social media site, compared with 63% of Americans making less than \$30,000 per year [24]. Women are also more likely to use social media compared with men [24]. Education also plays an important role—with 79% of Americans with a college degree reporting use of at least 1 social media site compared with 60% of Americans with high school or less education [24]. Age is also an important factor. Although older Americans are increasingly using social media, a greater proportion of younger adults are on social media compared with older adults. A total of 88% of 18- to 29-year-olds in the United States report using some form of social media compared with 37% of Americans 65 years and older [13]. These numbers have some implications for social listening, with a potential bias toward capturing voices from younger individuals with higher incomes and more education, and potentially capturing more female voices.

In this analysis, we examined social media and health forum posts to explore how patients with IBD understand, discuss, and act on the perceived risks and benefits associated with taking IBD medications during conception, pregnancy, and breastfeeding outside of the physicians' office. The objective of this study was to gain a more thorough understanding of how individuals taking IBD medications during key reproductive periods make decisions about their medications to inform clinical practitioners. Having a better understanding about why patients stop or refuse to take certain medications during key reproductive periods may allow clinicians to address specific beliefs and attitudes during office visits.

Methods

Social Media Data Mining

We collaborated with researchers from Treato, a social media data mining service, to find relevant patient experience data. Treato uses Natural Language Processing computer algorithms to collect and index patient, family, and caregiver content from over 3000 websites and social media sites such as Facebook, Healingwell, and TheBump. Posts are indexed using a lexicon of over 100,000 medical terms based on the Unified Medical Language System.

To find IBD-related posts, we extracted relevant data from an e-forum and social media database, using a set of relevant keywords developed through a literature review (Textbox 1). We sought to identify posts about medication use among individuals with IBD and developed 2 category lists—the Conditions category, which included names and abbreviations for IBD, UC, and CD; and the Medications category, which included brand and generic names and abbreviations for medications indicated to treat IBD, including biologic therapies. Posts were selected for analysis if they contained a keyword from the Conditions category as well as a keyword from the Medications category. We further culled posts by using pregnancy and fertility keywords to identify discussions relevant to this study. We analyzed English language posts published online between June 27, 2012, and June 27, 2015. Textbox 1 provides the list of the keywords used.

Qualitative Content Analysis

After identifying relevant posts, we coded the resulting text corpus using ATLAS.ti (Berlin, Scientific Software Development), a qualitative coding software program. We used qualitative content analysis methods to code the data [25,26]. The coding group was composed of a multidisciplinary team, including MSK, a social science researcher; SM, a

board-certified internist; ERC, a board-certified internist and gastroenterology fellow; and JLK, a research intern. The multidisciplinary composition of the team allowed for rich and varied discussions about the codes and the organization of the codes into larger thematic categories. The first level of analysis, open coding, identified individual concepts present in the social media posts. The unit of analysis was the entire post. We coded individual user's actions and focused on what users were *doing* (eg, sharing personal experiences with medications, expressing fear about medications, not taking medications due to beliefs about harm). We subsequently created a codebook from significant or frequently occurring codes in the data. After creating the codebook, we reviewed the data and coded using our focused codes. Posts were assigned a code if they mentioned experiences regarding pregnancy, breastfeeding, infertility, and experiences with IBD, IBD medications, or IBD-related complications. Each major theme was constructed by aggregating several focused codes identified through the open coding process. For example, focused codes for the major theme "Fear, Anxiety, and Uncertainty" included the following: fear about transferring IBD through genetics, fear of birth defects, fear of delivery issues with IBD, fear of fertility issues with IBD, fear of risks from medications during breastfeeding, fear of risks from medications during pregnancy, fear that pregnancy would worsen IBD symptoms, and general fears and anxieties.

The data collected contained no personal identifiers, and posts included only the website where the comments were posted. Although we were not able to ascertain the gender of the user, given that most social media posts referred to personal experiences with pregnancy, we assumed that most posts were written by women. However, we did find a few exceptions where partners posted on behalf of their wives or where men posted about male fertility issues related to IBD. The social media data were considered part of the public domain. The study was reviewed and approved by the Cedars-Sinai Medical Center IRB. No individual subjects were contacted.

Textbox 1. Keyword searches used to capture social media posts related to reproductive health and inflammatory bowel disease (IBD). The asterisk indicates a search term which was broadened by finding all words that start with the same letters.

- **Fertility:** fertil*|sperm|embryo|fetus|tcc|infert*|in vitro|in vitro|
- **Breastfeeding:** breastfeed*|breast milk|breast*|formula*|breastmilk
- **Pregnancy:** preg*|birth|conception|obgyn|preterm|thebump*|conceive|conception|labor|delivery|trimester|womb|c-section|csection|childbirth|prenatal|morning sickness|twins|babies|newborn*| infant*|placenta*|trisomy|pg
- **Birth defects:** Birth defect|defect*
- **Biologics:** Remicade|humira|cimizia|cimzia
- **Ulcerative colitis:** ulcerative colitis|UC (match case)

Results

Principal Results

We identified 1818 unique posts discussing the use of medication in IBD related to fertility, pregnancy, birth outcomes, or breastfeeding. We identified 7 major themes through the open coding process, presented in Table 1. Individuals used social media to (1) seek and share information and advice about medication use during pregnancy, infertility, and breastfeeding,

13.92% (252/1818); (2) express beliefs about the safety of therapies during pregnancy and breastfeeding, 7.43% (135/1818); (3) discuss decisions around personal experiences with medication use, including changes in medications or decisions to stop using medications during pregnancy and breastfeeding, 16.72% (304/1818); (4) articulate fears and anxieties about the safety and use of therapies during pregnancy and breastfeeding, 11.55% (210/1818); (5) discuss physician-patient relationships and related issues of trust and distrust, 3.14% (57/1818) and recommendations from providers

about medication use during pregnancy and breastfeeding 3.74% (68/1818); (6) share issues around infertility, including concerns about male infertility and IBD medications, 16.97% (309/1818); and (7) discuss issues around pregnancy and breastfeeding health, including IBD symptoms and medication side effects during pregnancy and post-pregnancy, 11.33% (206/1818).

Our analysis revealed that many factors influence the development of beliefs about IBD medications, including

anxieties and fears about taking medications during pregnancy and breastfeeding, the individuals' trust in their health care providers, beliefs about taking medications associated with infertility, and information gathered from online sources and online communities. Beliefs about IBD medications, in turn, are associated with decisions to start, stop, or change IBD medications during pregnancy, breastfeeding, and infertility treatments.

Table 1. Major themes identified in qualitative analysis of social media regarding reproductive health and Inflammatory Bowel Disease (IBD).

Themes ^a	n ^b (%)
Expressing beliefs about safety	135 (7.43)
Belief that benefits of IBD medications outweigh risks	32 (1.76)
Belief that IBD medication is not safe	35 (1.93)
Belief that IBD medication is safe	68 (3.74)
Sharing decisions about taking, changing, or stopping IBD medications	304 (16.72)
Changed, changing, or will change medication during pregnancy or while breast feeding	13 (0.72)
Took, is taking, or will take medication during pregnancy or while breast feeding	198 (10.89)
Stopped, stopping, or will stop medication during pregnancy or while breastfeeding	93 (5.12)
Expressing fear, anxiety, and uncertainty	210 (11.55)
Fear and anxiety surrounding medication and pregnancy and or breastfeeding	156 (8.58)
Uncertainty about use of IBD medication during pregnancy and or breastfeeding	54 (2.97)
Receiving recommendations from providers	68 (3.74)
Provider advised that IBD medications are safe	35 (1.76)
Provider advised that IBD medications are unsafe	23 (1.27)
Provider unsure about safety of IBD medications	10 (0.55)
Discussing patient-physician relationships	57 (3.14)
Poor communication and or evidence of distrust between patient and provider	38 (2.09)
Good communication or user expresses trust in provider	19 (1.05)
Giving and seeking advice experiences	252 (13.86)
Seeking advice about IBD and reproductive health	156 (8.58)
Giving advice about IBD and reproductive health	97 (5.34)
Reporting health during pregnancy	202 (11.11)
Reported healthy pregnancy	65 (3.58)
Reported flare-ups during pregnancy	65 (3.58)
Reported flare-ups postpartum	64 (3.58)
Onset of IBD coincided with pregnancy	37 (2.04)
Discussing infertility	309 (16.97)
General infertility issues	292 (16.06)
Male infertility	19 (1.04)

^aSubtheme totals may not add up to main theme totals due to the co-occurrence of codes. Percentages are calculated using the 1818 total coded posts.

^bThe unit of analysis was the social media post. The data did not include personal identifiers, so it is possible that the same individual posted multiple times on different websites.

Giving and Seeking Advice Around Inflammatory Bowel Disease and Medications

We identified 252 posts (13.92%) where individuals sought and shared advice with others on taking various types of medications, treating flare-ups, and addressing labor complications during pregnancy and breastfeeding. Many individuals expressed distress about recommendations from physicians or about information they had gathered online, particularly with regards to medication changes. For example, a woman 20 weeks into her pregnancy wrote about continuing Remicade infusions (a biologic treatment for IBD) until the 32-week mark and then discontinuing infusions and starting prednisone. She wrote:

They don't know the full effects of Remicade and breast milk so I will not be having anymore infusions after my infusion in September...My Perinatologist said most likely the best option to control it will be Prednisone...Is Prednisone really my only/the best option until I'm done nursing?

The woman in the post above expressed uncertainty in both the medical and the scientific community and the clinicians' recommendations. Hesitation in trusting clinical findings and the medical community as a whole recurred throughout the posts in our analysis.

Individuals also expressed anguish about the lack of definitive safety information available for newer medications, particularly biologics, and sought advice on these medications. For instance, 1 individual seeking advice about his/her partner's treatment wrote:

Any advice or questions we should ask our doctor would be great. Humira is so new that most Dr's don't have a clue when we ask about complications.

The individual noted that his or her partner had received conflicting information from various providers on medications and delivering via Caesarian section and shared the frustrations with the online community.

Beliefs About Safety of Inflammatory Bowel Disease Medications

Individuals also expressed their implicit and explicit beliefs about taking IBD medications during pregnancy, breastfeeding, and infertility treatments. Despite many concerns shared about the safety of IBD medications, there were twice as many posts expressing beliefs that IBD medications are safe during pregnancy and breastfeeding ($n=68$, 3.74%) as compared with beliefs that IBD medications are not safe ($n=35$, 1.93%). When sharing their experiences, many individuals explicitly noted that their medications were "very safe" and mentioned whether their providers agreed with their beliefs. One individual emphasized that she had consulted with 3 different clinicians when considering the safety of the IBD medications she was taking while pregnant, sharing both her personal opinion and viewpoints of clinicians in her life:

I personally felt it was quite safe to breastfeed while on mesalamine, as did my OBGYN, GI and sister [she's an MD].

Other social media users expressed doubts or firm beliefs that IBD medications were unsafe during pregnancy, echoing other qualitative work that has found that many women are uncertain about taking any medications during pregnancy.

Some individuals expressed beliefs that conflicted with recommendations they had received from their providers. Many of these patients expressed distrust in the evidence supporting the safety of IBD medications. For example, 1 patient posted a comment about a bad experience she had heard of regarding an individual using a biologic medication during pregnancy:

Humira has not been fully studied in pregnant women. There is even a registry to have to go on if you plan on being pregnant while you take humira. I know of a horror story and pregnancy and humira [I am not sure it was the humira that caused it or just bad luck.].

We also identified 32 (1.76%) posts where individuals weighed the benefits and risks of taking IBD medication and concluded that the benefits outweighed the risks of a severe flare-up during pregnancy. Avoiding flare-ups by taking IBD medications was a common narrative. One 29-year-old patient with CD discussed taking azathioprine for 8 years and staying on it throughout her pregnancy. She wrote:

If you come off aza before getting preg it is more dangerous for the baby in case you relapse whilst carrying the child and get really poorly! It's safer for yourself to be [healthy] and well whilst pregnant. I have heard about the birth defect thing (not through professionals though) but it was highly advised that I stay on aza.

She notes the various sources of information she consulted throughout her pregnancy, including information from social networks and formal sources (ie, professionals), highlighting the fact that many individuals seek multiple sources of information about medical decisions and must weigh the validity of these sources when making medication-related decisions.

Sharing Decisions About Taking, Changing, or Stopping Inflammatory Bowel Disease Medications During Pregnancy and Breastfeeding

Many individuals reported changing or stopping medications during pregnancy and breastfeeding, either as a result of the advice of a clinician or based on personal beliefs about the safety of taking IBD medications. There were 198 posts (10.89%) where individuals specifically cited taking or planning to take IBD medications during pregnancy or breastfeeding, 93 (5.12%) posts where individuals said they stopped or planned to stop medications during pregnancy or breastfeeding, and 13 (0.72%) posts where individuals said they planned on changing or changed medications during these reproductive periods.

Individuals cited reasons such as beliefs about the safety of medications, advice from providers, or advice from fellow sufferers for their decisions. As an example, one social media user discussed her decision to avoid breastfeeding while taking Remicade, a biologic therapy for IBD:

I do not plan on [breastfeeding] while being on remicade. We will formula feed. But I know A LOT of people that do bf, I'm just not comfortable with it.

This individual expressed that although she understood that other people had made the decision to breastfeed while taking these medications, she preferred to take the personal risk of going off the medication to avoid any perceived risk to her baby. This example highlights a tension that many individuals expressed between caring for their own bodies and avoiding perceived harm for the baby. The desire to avoid medications during pregnancy extended to individuals who had already given birth and taken medications while pregnant but wanted to avoid all medications for the next pregnancy. In one post, a woman expressed how she took 2 different IBD medications—Azulfidine and Asacol—during her first pregnancy, which she mentions was unplanned. However, for future planned pregnancies, she describes that she would:

...go off my meds for that 9 months. I would take all types of prenatals and omega and iron, and I would hopefully stay healthy for that baby. That would be MY choice for MY baby.

The individual posting here emphasizes her desire to use nonmedication alternatives (supplements), perceived as safer, to stay “healthy” and thus potentially be able to avoid taking medications she perceives as harmful.

Among those who cited either a personal reason or provider advice to stop taking a medication, nearly 3 times as many patients who stopped taking a medication did so as a result of personal reasons (N=25, 1.38%) versus those who stated they stopped based on the advice of a provider (N=8, 0.44%). One post exemplifies this decision-making process:

I am supposed to take Asacol HD. However I haven't taken it with my pregnancy for my own personal reasons especially since I do feel pretty good.

Expressing Fear, Anxiety, and Uncertainty

Patients frequently expressed fear, anxiety, and uncertainty around the use of IBD medications during periods of infertility, pregnancy, and breastfeeding. A major concern was the fear of birth defects and higher risk of miscarriage due to taking any medications during pregnancy. Although many patients had received medical advice, they frequently looked to others online to validate their concerns and gather more information. For instance, an individual with UC who was 7 months pregnant wrote:

My doctor now wants me to take Asacol HD and I'm very hesitant to take any medication while pregnant for fear that it may cause some kind of issue or birth defect with my baby.

Patients were conflicted between their fears of potential harms to the baby and the benefits of taking medications during pregnancy and breastfeeding, despite reassurances from clinicians. In many cases, providers appeared to express uncertainty with regards to the safety of the medication, leading the patient to also feel a heightened degree of uncertainty. One individual wrote:

My doc said I could do Rowasa and he thought I could still nurse “with caution” and watch the baby's weight gain and BM's, or I would need to cease nursing and take Canasa and Lialda. I hate taking pills. I only nurse my son once per day now because I am afraid of the effects of the Rowasa on him, but hate to give up that one session.

Discussing Patient-Physician Relationships and Medication Choices

Patients discussed beliefs that their physicians expressed about the safety and appropriate use of IBD medications during pregnancy and breastfeeding. There were 35 reports (1.93%) of providers stating that IBD medications were safe during pregnancy/breastfeeding, 23 (1.27%) where patients said the provider indicated that IBD medications were unsafe during these periods, and 10 (0.55%) where patients said their provider was unsure about the safety of IBD medications during pregnancy and breastfeeding.

In this example, an individual discusses her experiences with her maternal-fetal medicine specialist, the beliefs her provider has expressed, and the relationship she has with this provider:

...My maternal-fetal medicine doctor says Humira is perfectly safe for pregnancy and breastfeeding (class B drug)...I really recommend seeing an MFM doctor if you haven't already, even just for consultation. Mine knew a LOT about this stuff. More than my OBGYN, and even more than my IBD specialist. It was a very reassuring experience and has allowed me to enjoy my pregnancy!

In the post above, the individual expressed how the maternal-fetal medicine doctor was the one who was able to bridge the information divide between the 2 other specialists.

This experience echoes that of many individuals who posted on social media, many of whom expressed frustration that their obstetricians were not well informed on how to counsel women with IBD and how their gastroenterologists did not provide sufficient counseling during pregnancy. Many patients wrote that they were frustrated with their providers' unwillingness to discuss pregnancy and fertility issues. One patient expressed anguish regarding her conversations with physicians about IBD medications and pregnancy:

I'm so confused. None of my docs even want to talk about pregnancy! They say, well you really want to discuss this when you're on all these meds? Hmm, yes! Docs can be so frustrating at times. Possibly should I see an OBGYN?

We found that even when patients discussed provider recommendations, there were elements of distrust or uncertainty regarding the physicians' advice. One individual noted:

My doctors were scaring me saying it was bad but I've been reading online that a lot of women actually stay on it.

Expressing Concerns About Infertility and Inflammatory Bowel Disease Medications

We identified a group of posts focused on the effects of IBD and IBD medications on infertility ($n=316$, 17.38%). In these posts, social media users also discussed the effect of IBD or IBD medications on infertility in general or whether these affected the effectiveness of in vitro fertilization. Although most posts were related to female infertility and IBD, we identified 19 posts (1.05%) discussing male infertility or risks of birth defects due to medications taken by male IBD patients. Several patients reported confusion or surprise that medications used to manage IBD could affect male fertility. One patient with CD taking Cimzia and prednisone with a persistent flare-up sought advice on taking methotrexate while he and his wife attempted to conceive:

...I have never done anything like this before but my wife and I need help....My wife and I are most worried about having children soon or in the future but based on my research, you should not try while on the medication,. Does anyone know any info on this? Please help!

As with other posts, the individual decided to turn to social media to seek information rather than turn to a medical professional.

Communicating About General Pregnancy Health and Inflammatory Bowel Disease

Patients used social media to discuss IBD symptoms and potential causes during pregnancy. We found 202 posts (11.11%) that discussed flare-ups during and after pregnancy/during breastfeeding. Patients discussed how changes during pregnancy and breastfeeding affected their IBD symptoms; 65 posts (3.57%) discussed an IBD flare-up during pregnancy, 65 posts (3.57%) discussed complete remission during pregnancy, and 64 posts reported postpartum flare-ups. We also found 37 posts (2.04%) where patients reported that the onset of IBD coincided with their pregnancy, underscoring the importance of patient education on IBD and reproductive issues during the reproductive years.

Discussion

Principal Findings

We found that social media users often expressed significant tension between taking IBD medications for their own health and fears about potential birth defects. Our findings are in line with another research that finds that individuals often overestimate the teratogenic effects of medications during pregnancy [10]. A study using traditional survey methods also found that one quarter of participants thought it was more important to tolerate flares than take medications during pregnancy and 36% expressed beliefs that IBD medications are harmful to the fetus [27]. Although we were not able to measure education in our study, other studies have found that education level and health literacy are inversely associated with beliefs about medication safety during pregnancy, with less educated women/women with lower health literacy stating beliefs that medications are harmful during pregnancy [10,28].

Additionally, other studies have also found that adherence to medications for chronic conditions is low during pregnancy and is strongly related to women's beliefs about medication safety [29]. Our findings demonstrate that social media users indeed express that they stop taking medications for IBD during pregnancy when they are feeling better to reduce perceived risks.

Discussions with gastroenterologists about reproductive health in IBD are vital to bridge information gaps. Women with IBD are more likely to seek information about contraception and reproductive health from their gastroenterologist than from a family care physician, although most of this counseling is initiated by patients [29]. Yet, most patients seeing a gastroenterologist have no documented reproductive counseling, and 41% of general practitioners report that they never raise the issue of family planning with their IBD patients [12]. A timely discussion with a provider can alleviate some of these anxieties and help women make better decisions about managing their IBD during pregnancy. Discussion of family planning with a physician is associated with higher IBD-related pregnancy knowledge scores and lower odds of voluntary childlessness [8]. Gastroenterologists can play an important role in transmitting information about reproductive health and IBD during office visits to increase adherence to medications and reduce flare-ups.

One potential issue raised by social media users in this study may be the dearth of evidence-based information about medication safety during pregnancy, particularly for newer medications such as biologics. A 2010 study of 305 obstetricians and gynecologists found that 42% selected lack of sufficient information on medications as a significant barrier to discussing medication safety with pregnant women [30]. Increasing pregnancy registry participation could improve the quality and quantity of information available to clinicians and patients.

Strengths and Limitations

Strengths of this research include the ability to analyze the varied perspectives of hundreds of patients from all over the globe, which would be logistically unfeasible with traditional focus groups or interviews. However, there are also several limitations to this analysis. First, we cannot verify the identities of the patients posting on social media, particularly because many post using screen names. As a result, we cannot definitely confirm whether all patients have clinically verified diagnoses or are self-diagnosed. Another limitation is that not all patients are equally likely to write about their medical experiences online. As a result, we may only have the opportunity to view and analyze the experiences of a certain segment of the population.

As other researchers have found, social media use is associated with female gender, higher income, higher level of education, and younger age [13,24]. Moreover, these same factors are associated with characteristics of individuals who seek health information online [31]. Health status may be associated with online information-seeking, but the findings are conflicting—one study identified that healthier individuals are more likely to seek information online, whereas another study found that health status was not associated with seeking online information

[31,32]. Thus, it is possible that the social media posts included in this analysis reflect a younger, more educated, and higher-income cohort, and thus, the results may be more applicable to individuals with similar backgrounds. As with traditional focus groups, another limitation is that certain individuals may be more likely to engage and participate in the conversations [33]. However, we examined more than 1800 posts from a variety of different websites, which should allow for discussions and posts from a variety of different communities and individuals, which strengthens the findings from our analyses.

Conclusions

In this study, we used social listening methods to survey how individuals with IBD think about and make decisions about medications during key reproductive periods. We aimed to understand the types of beliefs that individuals have about IBD medications and birth defects, fertility, or outcomes in children. One important finding of our study is that individuals' needs may not be adequately addressed in clinical practice and that many patients are frustrated with their providers for failing to

discuss the effects of their pharmacological regimens on fertility, pregnancy, and breastfeeding. We found that both men and women have concerns about the safety and effects of IBD therapies during reproductive periods, and that there are significant fears and anxieties about the use of these therapies that are largely not addressed during office visits. As result, individuals may turn to online groups for information and emotional support. We also found that individuals often questioned the recommendations of their clinicians and sought second opinions via online forums. However, as these online forums may not provide evidence-based or updated medical information, individuals might receive erroneous or dangerous information. Future research should examine why clinicians might be hesitant to bring up reproductive concerns during office visits for individuals with IBD. Future research might also elucidate successful strategies that individuals can use to raise these topics that foster effective shared decision making in the IBD and reproductive health settings. Improving patient-provider communication around this issue may help improve the health of individuals with IBD during important reproductive periods.

Conflicts of Interest

The data for this study was acquired from another study that received support from Takeda Pharmaceuticals. MSK has research support from the NIH/National Center for Advancing Translational Science (NCATS) UCLA CTSI Grant Number TL1TR000121. BS has received grant funding from Pfizer, Takeda Pharmaceuticals, AstraZeneca, Ironwood Pharmaceuticals, Salix Pharmaceuticals, Nestle Health Sciences, and Amgen. BS has also received consulting fees, served on advisory committees, and/or served on review panels from Ironwood Pharmaceuticals, Commonwealth Labs, Valeant Pharmaceuticals, and Synergy Pharmaceuticals. MSK and BS declare no conflicts of interest in preparing this article. SM, EC, and JK report no relevant conflicts of interest.

References

1. Crohn's & Colitis Foundation of America. 2014 Nov. The Facts about Inflammatory Bowel Diseases URL: <http://www.crohnscolitisfoundation.org/assets/pdfs/updatedibdifactbook.pdf> [accessed 2018-05-03] [WebCite Cache ID 6z9Lfjq20]
2. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012 Jan;142(1):46-54.e42; quiz e30. [doi: [10.1053/j.gastro.2011.10.001](https://doi.org/10.1053/j.gastro.2011.10.001)] [Medline: [22001864](https://pubmed.ncbi.nlm.nih.gov/22001864/)]
3. Fonager K, Sørensen HT, Olsen J, Dahlerup JF, Rasmussen SN. Pregnancy outcome for women with Crohn's disease: a follow-up study based on linkage between national registries. *Am J Gastroenterol* 1998 Dec;93(12):2426-2430. [doi: [10.1111/j.1572-0241.1998.00698.x](https://doi.org/10.1111/j.1572-0241.1998.00698.x)] [Medline: [9860403](https://pubmed.ncbi.nlm.nih.gov/9860403/)]
4. Gawron LM, Goldberger AR, Gawron AJ, Hammond C, Keefer L. Disease-related pregnancy concerns and reproductive planning in women with inflammatory bowel diseases. *J Fam Plann Reprod Health Care* 2015 Oct;41(4):272-277. [doi: [10.1136/jfprhc-2014-101000](https://doi.org/10.1136/jfprhc-2014-101000)] [Medline: [25902816](https://pubmed.ncbi.nlm.nih.gov/25902816/)]
5. Heetun ZS, Byrnes C, Neary P, O'Morain C. Review article: Reproduction in the patient with inflammatory bowel disease. *Aliment Pharmacol Ther* 2007 Aug 15;26(4):513-533 [FREE Full text] [doi: [10.1111/j.1365-2036.2007.03397.x](https://doi.org/10.1111/j.1365-2036.2007.03397.x)] [Medline: [17661756](https://pubmed.ncbi.nlm.nih.gov/17661756/)]
6. Damas OM, Deshpande AR, Avalos DJ, Abreu MT. Treating Inflammatory Bowel Disease in Pregnancy: The Issues We Face Today. *J Crohns Colitis* 2015 Jun 30;9(10):928-936. [doi: [10.1093/ecco-jcc/jjv118](https://doi.org/10.1093/ecco-jcc/jjv118)] [Medline: [26129693](https://pubmed.ncbi.nlm.nih.gov/26129693/)]
7. Zelinkova Z, Mensink PB, Dees J, Kuipers EJ, van der Woude CJ. Reproductive wish represents an important factor influencing therapeutic strategy in inflammatory bowel diseases. *Scand J Gastroenterol* 2010;45(1):46-50. [doi: [10.3109/00365520903362628](https://doi.org/10.3109/00365520903362628)] [Medline: [19883275](https://pubmed.ncbi.nlm.nih.gov/19883275/)]
8. Huang VW, Chang H, Kroeker KI, Goodman KJ, Hegadoren KM, Dieleman LA, et al. Does the level of reproductive knowledge specific to inflammatory bowel disease predict childlessness among women with inflammatory bowel disease? *Can J Gastroenterol Hepatol* 2015 Mar;29(2):95-103. [Medline: [25803020](https://pubmed.ncbi.nlm.nih.gov/25803020/)]
9. Nordeng H, Ystrøm E, Einarson A. Perception of risk regarding the use of medications and other exposures during pregnancy. *Eur J Clin Pharmacol* 2010 Feb;66(2):207-214. [doi: [10.1007/s00228-009-0744-2](https://doi.org/10.1007/s00228-009-0744-2)] [Medline: [19841915](https://pubmed.ncbi.nlm.nih.gov/19841915/)]
10. Nordeng H, Koren G, Einarson A. Pregnant women's beliefs about medications--a study among 866 Norwegian women. *Ann Pharmacother* 2010 Sep;44(9):1478-1484. [doi: [10.1345/aph.1P231](https://doi.org/10.1345/aph.1P231)] [Medline: [20736425](https://pubmed.ncbi.nlm.nih.gov/20736425/)]

11. Petersen I, McCrea RL, Lupattelli A, Nordeng H. Women's perception of risks of adverse fetal pregnancy outcomes: a large-scale multinational survey. *BMJ Open* 2015 Jun 01;5(6):e007390 [FREE Full text] [doi: [10.1136/bmjopen-2014-007390](https://doi.org/10.1136/bmjopen-2014-007390)] [Medline: [26033946](https://pubmed.ncbi.nlm.nih.gov/26033946/)]
12. Toomey D, Waldron B. Family planning and inflammatory bowel disease: the patient and the practitioner. *Fam Pract* 2013 Feb;30(1):64-68. [doi: [10.1093/fampra/cms035](https://doi.org/10.1093/fampra/cms035)] [Medline: [22843639](https://pubmed.ncbi.nlm.nih.gov/22843639/)]
13. Smith A, Anderson M. Washington, DC: Pew Research Center; 2018 Mar 01. Social Media Use in 2018: Demographics and Statistics URL: <http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/> [accessed 2018-05-03] [WebCite Cache ID 6z9L1ooBm]
14. Fox S. Pew Internet. 2011 May 12. The social life of health information, 2011 URL: <http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/> [accessed 2018-05-03] [WebCite Cache ID 6z9LAJwMv]
15. Whitman CB, Reid MW, Arnold C, Patel H, Ursos L, Sa'adon R, et al. Balancing opioid-induced gastrointestinal side effects with pain management: insights from the online community. *J Opioid Manag* 2015;11(5):383-391. [doi: [10.5055/jom.2015.0288](https://doi.org/10.5055/jom.2015.0288)] [Medline: [26535966](https://pubmed.ncbi.nlm.nih.gov/26535966/)]
16. Keller MS, Park HJ, Cunningham ME, Fouladian JE, Chen M, Spiegel BMR. Public Perceptions Regarding Use of Virtual Reality in Health Care: A Social Media Content Analysis Using Facebook. *J Med Internet Res* 2017 Dec 19;19(12):e419 [FREE Full text] [doi: [10.2196/jmir.7467](https://doi.org/10.2196/jmir.7467)] [Medline: [29258975](https://pubmed.ncbi.nlm.nih.gov/29258975/)]
17. Martinez B, Dailey F, Almario CV, Keller MS, Desai M, Dupuy T, et al. Patient Understanding of the Risks and Benefits of Biologic Therapies in Inflammatory Bowel Disease: Insights from a Large-scale Analysis of Social Media Platforms. *Inflamm Bowel Dis* 2017 Jul;23(7):1057-1064. [doi: [10.1097/MIB.0000000000001110](https://doi.org/10.1097/MIB.0000000000001110)] [Medline: [28410343](https://pubmed.ncbi.nlm.nih.gov/28410343/)]
18. Lazard AJ, Wilcox GB, Tuttle HM, Glowacki EM, Pikowski J. Public reactions to e-cigarette regulations on Twitter: a text mining analysis. *Tob Control* 2017 Dec;26(e2):e112-e116. [doi: [10.1136/tobaccocontrol-2016-053295](https://doi.org/10.1136/tobaccocontrol-2016-053295)] [Medline: [28341768](https://pubmed.ncbi.nlm.nih.gov/28341768/)]
19. Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter Data to Gain Insights into E-cigarette Marketing and Locations of Use: An Infoveillance Study. *J Med Internet Res* 2015 Nov 06;17(11):e251 [FREE Full text] [doi: [10.2196/jmir.4466](https://doi.org/10.2196/jmir.4466)] [Medline: [26545927](https://pubmed.ncbi.nlm.nih.gov/26545927/)]
20. 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](https://doi.org/10.2196/jmir.4969)] [Medline: [26508089](https://pubmed.ncbi.nlm.nih.gov/26508089/)]
21. Anderson LS, Bell HG, Gilbert M, Davidson JE, Winter C, Barratt MJ, et al. Using Social Listening Data to Monitor Misuse and Nonmedical Use of Bupropion: A Content Analysis. *JMIR Public Health Surveill* 2017 Feb 01;3(1):e6 [FREE Full text] [doi: [10.2196/publichealth.6174](https://doi.org/10.2196/publichealth.6174)] [Medline: [28148472](https://pubmed.ncbi.nlm.nih.gov/28148472/)]
22. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS One* 2011;6(8):e23610 [FREE Full text] [doi: [10.1371/journal.pone.0023610](https://doi.org/10.1371/journal.pone.0023610)] [Medline: [21886802](https://pubmed.ncbi.nlm.nih.gov/21886802/)]
23. Langer R, Beckman SC. Sensitive research topics: netnography revisited. *Qualitative Market Research: An International Journal* 2005;8(2):189-203. [doi: [10.1108/13522750510592454](https://doi.org/10.1108/13522750510592454)]
24. Pew Internet. 2018 Feb 05. Social Media Fact Sheet URL: <http://www.pewinternet.org/fact-sheet/social-media/> [accessed 2018-05-03] [WebCite Cache ID 6z9M6EoV0]
25. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nurs Health Sci* 2013 Sep;15(3):398-405. [doi: [10.1111/nhs.12048](https://doi.org/10.1111/nhs.12048)] [Medline: [23480423](https://pubmed.ncbi.nlm.nih.gov/23480423/)]
26. Corbin J, Strauss A, Strauss AL. Basics of Qualitative Research. Thousand Oaks: SAGE Publishing; 2015.
27. Selinger CP, Eaden J, Selby W, Jones DB, Katelaris P, Chapman G, et al. Inflammatory bowel disease and pregnancy: lack of knowledge is associated with negative views. *J Crohns Colitis* 2013 Jul;7(6):e206-e213. [doi: [10.1016/j.crohns.2012.09.010](https://doi.org/10.1016/j.crohns.2012.09.010)] [Medline: [23040449](https://pubmed.ncbi.nlm.nih.gov/23040449/)]
28. Lupattelli A, Picinardi M, Einarson A, Nordeng H. Health literacy and its association with perception of teratogenic risks and health behavior during pregnancy. *Patient Educ Couns* 2014 Aug;96(2):171-178. [doi: [10.1016/j.pec.2014.04.014](https://doi.org/10.1016/j.pec.2014.04.014)] [Medline: [24862909](https://pubmed.ncbi.nlm.nih.gov/24862909/)]
29. Mountfield R, Bampton P, Prosser R, Muller K, Andrews JM. Fear and fertility in inflammatory bowel disease: a mismatch of perception and reality affects family planning decisions. *Inflamm Bowel Dis* 2009 May;15(5):720-725. [doi: [10.1002/ibd.20839](https://doi.org/10.1002/ibd.20839)] [Medline: [19067431](https://pubmed.ncbi.nlm.nih.gov/19067431/)]
30. Morgan MA, Cragan JD, Goldenberg RL, Rasmussen SA, Schulkin J. Obstetrician-gynaecologist knowledge of and access to information about the risks of medication use during pregnancy. *J Matern Fetal Neonatal Med* 2010 Oct;23(10):1143-1150. [doi: [10.3109/14767051003653252](https://doi.org/10.3109/14767051003653252)] [Medline: [20218819](https://pubmed.ncbi.nlm.nih.gov/20218819/)]
31. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. *Soc Sci Med* 2004 Nov;59(9):1795-1806. [doi: [10.1016/j.socscimed.2004.02.020](https://doi.org/10.1016/j.socscimed.2004.02.020)] [Medline: [15312915](https://pubmed.ncbi.nlm.nih.gov/15312915/)]
32. Xiao N, Sharman R, Rao HR, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. *Decis Support Syst* 2014 Jan;57:417-427 [FREE Full text]
33. Stewart DW, Shamdasani PN. Focus Groups: Theory and Practice. Thousand Oaks: Sage Publications; 2015.

Abbreviations**IBD:** inflammatory bowel disease**CD:** Crohn disease**UC:** ulcerative colitis

Edited by G Eysenbach; submitted 17.01.18; peer-reviewed by R Randell, S Holton; comments to author 28.03.18; revised version received 10.04.18; accepted 10.04.18; published 11.06.18.

Please cite as:

Keller MS, Mosadeghi S, Cohen ER, Kwan J, Spiegel BMR

Reproductive Health and Medication Concerns for Patients With Inflammatory Bowel Disease: Thematic and Quantitative Analysis Using Social Listening

J Med Internet Res 2018;20(6):e206

URL: <http://www.jmir.org/2018/6/e206/>

doi: [10.2196/jmir.9870](https://doi.org/10.2196/jmir.9870)

PMID: [29891471](https://pubmed.ncbi.nlm.nih.gov/29891471/)

©Michelle Sophie Keller, Sasan Mosadeghi, Erica R Cohen, James Kwan, Brennan Mason Ross Spiegel. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.06.2018. 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.

Review

eHealth as the Next-Generation Perinatal Care: An Overview of the Literature

Josephus FM van den Heuvel^{1*}, MSc, MD; T Katrien Groenhof^{1*}, MSc, MD; Jan HW Veerbeek¹, MD, PhD; Wouter W van Solinge², PhD; A Titia Lely¹, MD, PhD; Arie Franx¹, MD, PhD; Mireille N Bekker¹, MD, PhD

¹Division of Woman and Baby, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

²Department of Clinical Chemistry and Hematology, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

*these authors contributed equally

Corresponding Author:

Josephus FM van den Heuvel, MSc, MD

Division of Woman and Baby

University Medical Center Utrecht

Utrecht University

Lundlaan 6

Utrecht, 3584 AB

Netherlands

Phone: 31 887554913

Fax: 31 887555320

Email: jheuve10@umcutrecht.nl

Abstract

Background: Unrestricted by time and place, electronic health (eHealth) provides solutions for patient empowerment and value-based health care. Women in the reproductive age are particularly frequent users of internet, social media, and smartphone apps. Therefore, the pregnant patient seems to be a prime candidate for eHealth-supported health care with telemedicine for fetal and maternal conditions.

Objective: This study aims to review the current literature on eHealth developments in pregnancy to assess this new generation of perinatal care.

Methods: We conducted a systematic literature search of studies on eHealth technology in perinatal care in PubMed and EMBASE in June 2017. Studies reporting the use of eHealth during prenatal, perinatal, and postnatal care were included. Given the heterogeneity in study methods, used technologies, and outcome measurements, results were analyzed and presented in a narrative overview of the literature.

Results: The literature search provided 71 studies of interest. These studies were categorized in 6 domains: information and eHealth use, lifestyle (gestational weight gain, exercise, and smoking cessation), gestational diabetes, mental health, low- and middle-income countries, and telemonitoring and teleconsulting. Most studies in gestational diabetes and mental health show that eHealth applications are good alternatives to standard practice. Examples are interactive blood glucose management with remote care using smartphones, telephone screening for postnatal depression, and Web-based cognitive behavioral therapy. Apps and exercise programs show a direction toward less gestational weight gain, increase in step count, and increase in smoking abstinence. Multiple studies describe novel systems to enable home fetal monitoring with cardiotocography and uterine activity. However, only few studies assess outcomes in terms of fetal monitoring safety and efficacy in high-risk pregnancy. Patients and clinicians report good overall satisfaction with new strategies that enable the shift from hospital-centered to patient-centered care.

Conclusions: This review showed that eHealth interventions have a very broad, multilevel field of application focused on perinatal care in all its aspects. Most of the reviewed 71 articles were published after 2013, suggesting this novel type of care is an important topic of clinical and scientific relevance. Despite the promising preliminary results as presented, we accentuate the need for evidence for health outcomes, patient satisfaction, and the impact on costs of the possibilities of eHealth interventions in perinatal care. In general, the combination of increased patient empowerment and home pregnancy care could lead to more satisfaction and efficiency. Despite the challenges of privacy, liability, and costs, eHealth is very likely to disperse globally in the next decade, and it has the potential to deliver a revolution in perinatal care.

KEYWORDS

pregnancy; eHealth; telemedicine; pregnancy complications; fetal monitoring; patient-centered care; pregnancy, high risk; diabetes, gestational; remote consultation; ambulatory monitoring; obstetrics; perinatal care; antenatal care

Introduction

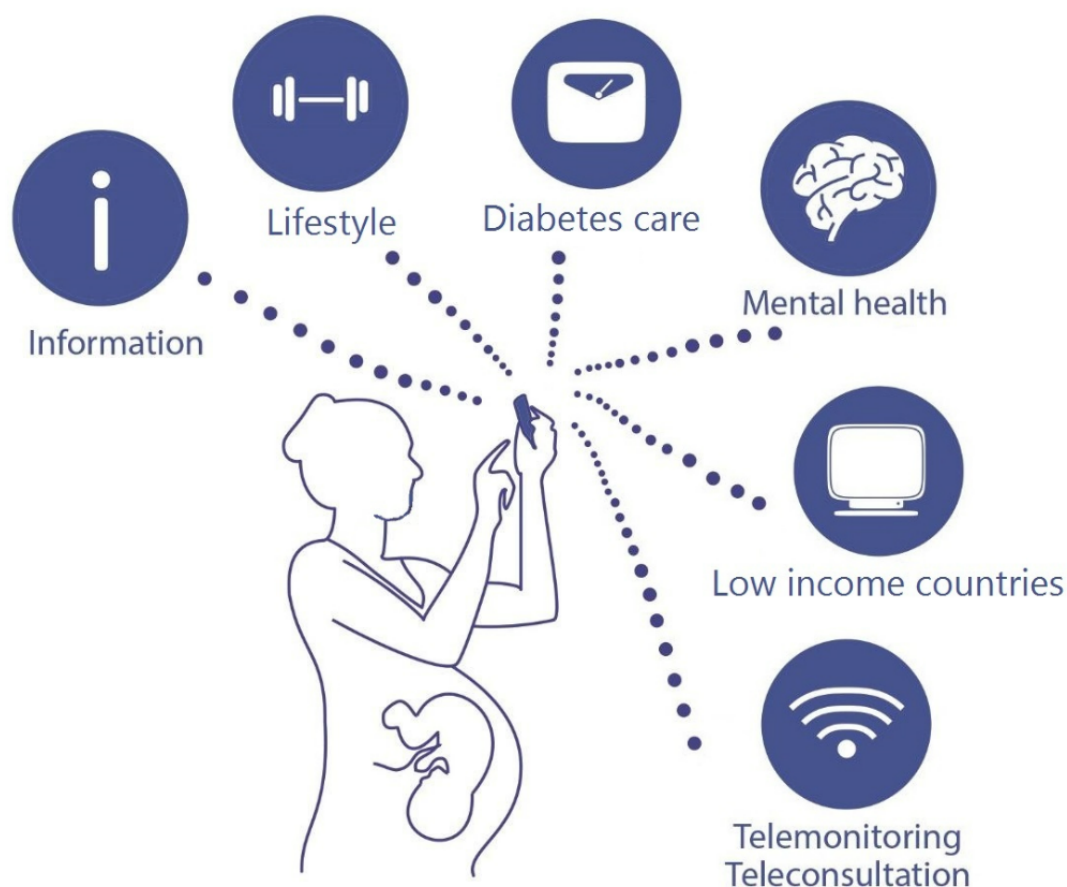
Electronic Health—A New Opportunity?

Health care is facing the emergence of a new range of systems, services, and applications using electronic communication. *Electronic health* (eHealth) is the network of technology applications regarding health issues, including, for example, Web-based informative programs, remote monitoring, teleconsultation, and mobile device-supported care [1]. As the health care costs in developed countries continue to increase, policies for cost reduction without concessions to the quality of care are being imposed. Unrestricted by time and place, eHealth applications also provide solutions for patient empowerment and value-based health care [2]. Patient empowerment is assumed to improve patient participation in medical decision making, commitment to treatment, and thus, health outcomes [3-5]. The boost in patient engagement can be an important factor for the improvement of quality of care and patient safety [6].

Young women in their reproductive years are frequent users of internet, social media, and smartphone apps [7]. The internet is ever more utilized for the search of health information on prenatal, perinatal, and postnatal topics [8]. Furthermore, the Web is also used as a forum for the exchange of experiences and peer support [9]. Figure 1 shows multiple domains of perinatal care in which eHealth is already being used by patients and health care providers.

Protocols of professionals' associations and institutions contain little communication regarding eHealth. No statements are made regarding eHealth in guidelines from the British Royal College of Obstetricians and Gynaecologists, the National Institute for Health and Care Excellence, and the American College of Obstetrics and Gynecology. The Dutch Association of Obstetrics and Gynecology notes that *developments in eHealth should be actively implemented in obstetric healthcare to induce the shift of scheduled care to the home setting and thus lower the in-hospital care burden* [10].

Figure 1. Electronic health (eHealth) solutions in 6 domains of perinatal care.



Objective

eHealth has the potential to fulfill a key role in the transformation of the health care system for both patients and caregivers. However, questions are raised if eHealth can deliver the quality of care that is required to remain or even improve health outcomes. It is evident that there is a need for guidance and management of quality standards. Issues of costs and reimbursement; safety of data collection; and storage, privacy, and reliability of information on websites and in apps should also be taken into account.

Our aim is to provide a comprehensive and contemporary overview of the literature on eHealth in perinatal care and assess the applicability, advantages, limitations, and future of this new generation of pregnancy care.

Methods

A systematic literature search was performed in PubMed and EMBASE in June 2017, combining various synonyms for perinatal care and telemedicine and eHealth (see [Multimedia Appendix 1](#) for the search strategy). Studies reporting the use of eHealth during prenatal, perinatal, and postnatal care were included. Due to the rapid developments in this field and our contemporary scope, we excluded articles describing outdated technologies, for example, fax communication, phonocardiography, and home visits or home care. Screening and reviewing the abstracts and full articles was done by 2 independent authors (JH and KG). Given the heterogeneity in study methods, used technologies, and outcome measurements, results were analyzed and presented in a narrative overview of the literature.

Results

Study Selection

Literature search and reference screening provided 71 studies of interest (see [Multimedia Appendix 1](#) for the flow diagram of selection of studies). All articles were categorized in 6 domains, which will be addressed accordingly: information and eHealth use, lifestyle (gestational weight gain, exercise, and smoking cessation), gestational diabetes, mental health, low- and middle-income countries, and telemonitoring/teleconsulting (see also [Figure 1](#)). [Tables 1-3](#) show the overview of 71 publications in 6 domains of perinatal care in which eHealth use in patient care was described, implemented, or compared with standard care.

Information and eHealth Use in Pregnancy

In 15 studies, the characteristics of eHealth users in the perinatal period were described ([Table 1](#)). Around 88% (31/35) of participants owned a smartphone [11]. Usage of websites and pregnancy apps for medical information varies from 50% to 98% [7,11-14]. Online information-seeking behavior is common in pregnant women in general and it is not restricted to women with a special profile based on age, education, or social support [7]. Increased knowledge on pregnancy complications has also shown to reduce maternal anxiety and costly hospital visits [15,16]. Factors associated with app use in pregnancy are

younger age, nulliparity, lower self-rated health, and higher education. Furthermore, 25.6% (56/219) of questioned women showed interest in a tailored pregnancy app initiated by their health care provider [7,14].

The most searched topics are fetal development, pregnancy complications, healthy lifestyle during pregnancy, generic and specific guidance/advice during pregnancy, and lactation [13,17]. Although they value the Web-based medical information as moderately reliable, 71.3%-75.1% (582/800) of the women do not discuss the information found on internet with their gynecologist [17,18]. One study reported that their lifestyle app helped women to initiate the conversation with their health caregiver on this subject [19].

There is an increasing use of internet for health information, including the perinatal period. However, websites are often contradictory and this may lead to confusion [20]. eHealth may be helpful to address questions through informative websites, apps, and peer support platforms designed by health professionals. Furthermore, eHealth may provide possibilities for decision support in more complicated pregnancies [21].

Health Outcome After eHealth Intervention

The effect on health is the most important issue to address in the effective implementation of eHealth in perinatal care. Parameters for quality standards include disease outcomes, enhancing patient adherence to treatment, reducing overuse, and increasing access to care [29]. Results of the search showed that most publications focus on the improvement of lifestyle (gestational weight gain, exercise, smoking cessation), gestational diabetes monitoring, mental health, care in lower- and middle-income countries, and telemonitoring.

Lifestyle

Our search provided 13 publications describing health outcomes for eHealth interventions on lifestyle during pregnancy ([Table 2](#)). Pursuing a healthy lifestyle has proven to be beneficial for pregnancy outcomes such as preterm birth, gestational diabetes, or pre-eclampsia [30-32]. Participant motivation, reducing the dropout rate, and sustainability of long-term results are notoriously difficult in lifestyle studies. Smartphone technologies provide features to overcome these obstacles. Results from feasibility studies show good acceptability, adherence, and engagement for eHealth interventions for healthy gestational weight gain and physical activity, favoring an app over a website [33,34]. Physical activity trials with tailored text messaging (short message service, SMS) services resulted in an increase in step count up to 4 times more than in the control group. In addition, eHealth interventions resulted in better perceived health in pregnancy and lower, healthier gestational weight gain in both nonobese (7.8 kg vs 9.7 kg) and obese women (6.65 kg vs 9.74 kg) [35-37]. Dietary apps directed at healthy gestational weight gain are still in developmental and experimental phase [27,38,39].

Smoking during pregnancy increases the risk of unfavorable pregnancy outcomes. In 2010, approximately 10% of the women smoked cigarettes during pregnancy, especially younger, non-white mothers of a lower social economic status [40,41]. The 2016 review by Heminger et al summarizes the studies

performed on SMS programs and mobile apps for smoking cessation in pregnancy [42]. Women participating in SMS cessation programs report relatively high abstinence of 38% in the first week and 54% in the second week (n=20). Biochemically confirmed abstinence rates were 12.5% in participants compared with 7.8% in controls (n=207). Smartphone apps were preferred over SMS-driven programs, as seen in over 10,000 installations of apps compared with 20-800 registrations in SMS programs.

Gestational Diabetes

About 5% to 7% of all pregnancies are complicated by gestational diabetes mellitus (GDM) in the United Kingdom and United States (range 1%-25%) [43]. Pregnancies with GDM are associated with perinatal complications such as caesarean section, shoulder dystocia, and neonatal hypoglycemia. Extensive glucose monitoring during pregnancy is a burden for both patients and health care budgets. eHealth in GDM care has

evolved most notably of all perinatal appliances of eHealth the last 3 years [44]. We found 13 studies on this topic, including 2 systematic reviews (Table 2). Developments involve smartphone-facilitated remote blood glucose monitoring, management of medication schedules through Web-based or SMS-facilitated feedback systems, and telephone review service to support and supervise glycemic control [45-51]. Overall, studies showed a decrease in planned and unplanned visits by 50% to 66%, whereas no unfavorable differences in glycemic control, maternal, and neonatal outcomes occurred [47-49,52]. Two recent systematic reviews with meta-analysis confirm these results [53,54]. No cost-effectiveness analysis was performed due to insufficient data. There is also increasing evidence of GDM as a risk factor for type 2 diabetes later in life [55]. eHealth programs for follow-up of women with a history of GDM are being developed but need to be examined more thoroughly [45].

Table 1. Information and electronic health (eHealth) use in pregnancy: overview of the literature.

Reference	Methods	N	Technology/eHealth intervention
Sayakhot et al [12]	Systematic review (with 7 cross-sectional studies)	3359	Patients' use of internet for pregnancy information
Ledford et al [22]	RCT ^a pilot	150	App for pregnancy education and record keeping
Walker et al [15]	Prospective cohort	8	Website for education on placental complications
Bush et al [23]	Before-after study	85	Prenatal care app use and user engagement
Wallwiener et al [7]	Cross sectional	220	Surveys and questionnaires on use of eHealth (smartphones, internet, apps) during pregnancy
Scaioli et al [13]	Cross sectional	1347	Surveys and questionnaires on use of eHealth (smartphones, internet, apps) during pregnancy
Peragallo et al [24]	Cross sectional	100	Surveys and questionnaires on use of eHealth (smartphones, internet, apps) during pregnancy
Lee et al [14]	Cross sectional	193	Surveys and questionnaires on use of eHealth (smartphones, internet, apps) during pregnancy
Lupton et al [25]	Cross sectional	410	Surveys and questionnaires on use of eHealth (smartphones, internet, apps) during pregnancy
Narasimhulu et al [17]	Cross sectional	586	Surveys and questionnaires on use of eHealth (smartphones, internet, apps) during pregnancy
Goetz et al [26]	Qualitative research	30	Focus groups and interviews on eHealth use and implementation (in pregnant women, men, and clinicians)
Willcox et al [27]	Qualitative research	27	Focus groups and interviews on eHealth use and implementation (in pregnant women, men, and clinicians)
Rodger et al [11]	Qualitative research	35	Focus groups and interviews on eHealth use and implementation (in pregnant women, men, and clinicians)
Mackert et al [28]	Qualitative research	32	Focus groups and interviews on eHealth use and implementation (in pregnant women, men, and clinicians)
Lupton et al [25]	Qualitative research	36	Focus groups and interviews on eHealth use and implementation (in pregnant women, men, and clinicians)

^aRCT: randomized controlled trial.

Table 2. Health outcome of electronic health (eHealth) use in lifestyle and gestational diabetes mellitus management in pregnancy: overview of the literature.

Study domain and reference	Methods	N	Technology/eHealth intervention
Lifestyle: Gestational weight gain, exercise, smoking cessation (13 studies)			
O'Brien et al [79]	Systematic review (with 7 studies)	33	Technology-supported diet and lifestyle interventions
Pollak et al [80]	RCT ^a	33	SMS ^b programs on healthy lifestyle
Soltani et al [35]	RCT	14	SMS for healthy lifestyle in women with BMI ^c >30
Graham et al [81]	RCT	1335	Internet-based platform to prevent excessive weight gain
Hayman et al [34]	RCT	77	Web-based physical activity intervention
Huberty et al [82]	RCT	80	SMS programs to increase physical activity
Willcox et al [37]	RCT	91	Healthy gestational weight gain for obese pregnancies
Knight et al [19]	One group pilot	10	App with information for lifestyle behavior
Waring et al [33]	Cross sectional	64	Survey on interest in lifestyle app or website
Choi et al [36]	RCT pilot	30	Activity app+pedometer wearable
Lewis et al [83]	Observational cohort	37	Exercise with SMS or app-based support
Guo et al [84]	One group pilot	50	Video program with yoga via Facebook or DVD
Heminger et al [42]	Systematic review (with 7 RCTs)	702	SMS or app support on smoking: quitting date, relapse, information, daily messages
Gestational diabetes mellitus (13 studies)			
Ming et al [54]	Systematic review (with 7 RCTs)	579	Telemedicine for diabetes in pregnancy
Rasekaba et al [53]	Systematic review (with 3 RCTs)	243	Telemedicine for glucose monitoring
Kruger et al [85]	RCT	18	Telemedicine for glucose monitoring
Dalfrà et al [86]	RCT	276	Telemedicine for glucose monitoring
Perez-Ferre et al [52]	RCT	100	Telemedicine for glucose monitoring
Wojcicki et al [87]	RCT	30	Telemedicine for glucose monitoring
Carral et al [49]	Prospective cohort	104	Web-based telemedicine system
Given et al [50]	Feasibility study	50	Web-based telemedicine system
Nicholson et al [88]	Feasibility study	23	Web-based self monitoring, diary
Mackillop et al [51]	Pilot study	48	Smartphone app with blood glucose meter
Ganapathy et al [89]	Pilot study	50	Remote blood pressure measurements
Khorshidi et al [45]	RCT	80	Postpartum screening after GDM ^d
Harrison et al [90]	Survey+interviews	70	Acceptability of telemedicine for GDM patients

^aRCT: randomized controlled trial.^bSMS: short message services.^cBMI: body mass index.^dGDM: gestational diabetes mellitus.

Table 3. Health outcome of electronic health (eHealth) use in electronic mental (e-mental) health, low- and middle-income countries, and telemonitoring and teleconsultation in pregnancy: overview of the literature.

Study domain and reference	Methods	N	Technology/eHealth intervention
E-mental health (16 studies)			
Lau et al [64]	Systematic review (with 8 RCTs ^a)	1523	Therapist-supported internet-based cognitive behavior therapy among postpartum women
Lee et al [61]	Systematic review (with 4 RCTs)	1274	Cognitive behavioral therapy with internet
Ashford et al [63]	Systematic review (with 11 studies)	1537	Web-based perinatal mental health interventions
Milgrom et al [91]	RCT	43	Cognitive behavioral therapy with internet
Ngai et al [92]	RCT	397	Telephone-based cognitive-behavioral Therapy
Shamshiri Milani et al [93]	RCT	54	Telephone-based cognitive-behavioral therapy
Kingston et al [60]	RCT	636	Acceptability of e-screening for mental health
Fontein et al [94]	Before-after study	433	Website for maternal stress prevention
Jimenez-Serrano et al [59]	Prospective cohort	1880	App screening for postpartum depression
Posmontier et al [62]	Prospective cohort	61	Telephone-administered psychotherapy
Letourneau et al [65]	Prospective cohort	64	Telephone-based peer support intervention
Broom et al [95]	Observational	54	Supportive text messaging in postpartum depression
Mitchell et al [58]	Cross sectional	106	Telephone screening for postpartum depression
Figueiredo et al [96]	Cross sectional	90	Telephone screening for postpartum depression
Pugh et al [97]	Case study	1	Therapeutic assistance with email and SMS ^b
Pineros-Leano et al [98]	Qualitative	25	Screening for postpartum depression using mobile health
Low and middle income countries (2 studies)			
Lee et al [67]; Sondaal et al [66]	2 systematic reviews with 18 RCTs and 18 observational studies	34,149	Mobile health interventions for prenatal, birth, and postnatal period in low- and middle-income countries
Telemonitoring and teleconsulting (12 studies)			
Tapia-Conyer et al [75]	RCT	153	Wireless antepartum maternal-fetal monitoring
Pflugeisen et al [74]	Non-RCT	1058	Prenatal care with virtual visits and home measurements
Ivey et al [99]	Prospective cohort	155	Teleconsultation with tertiary center
Cuneo et al [100]	Prospective cohort	125	Home fetal heart monitoring for anti-SSA+ ^c patients
Rauf et al [73]	Prospective cohort	70	Fetal monitoring system for induction of labor
Krishnamurti et al [101]	Prospective cohort	16	Smartphone app with information and symptom scores
Rhoads et al [102]	Non-RCT	50	Telemonitoring of postpartum hypertension
Kerner et al [77]	Feasibility study	36	Self-administered fetal heart rate monitoring
Marko et al [103]	Feasibility study	8	Remote monitored pregnancy care (blood pressure, weight)
Marko et al [76]	Controlled trial	100	Prenatal care with app and telemonitoring
Lanssens et al [104]	Retrospective cohort	166	Remote monitoring of hypertension in pregnancy
Pflugeisen et al [105]	Cross sectional	171	Satisfaction with virtual obstetric care

^aRCT: randomized controlled trial.^bSMS: short message services.^cAnti-SSA: Anti-Sjögren's-syndrome-related antigen A.

Mental Health

Electronic mental health has already proven to be successful in general population mental health management [56]. In 16 studies, the applicability on screening for and treatment of postpartum depression was investigated (Table 3). The

prevalence of postpartum depression is 3%-15%. These women are reluctant to seek medical attention despite the heavy burden of disease, most notably because of the fear of their child being taken away from them [57,58]. Screening with telephone (alpha coefficients of .72-.94), app (sensitivity 72% and specificity 73%), and iPads were found feasible and acceptable [58-60].

eHealth programs (eg, online sessions based on cognitive behavior therapy) effectuate significant reductions in the depression scales and on symptom scores compared with treatment as usual [61-64]. Besides this significant effect size favoring eHealth, in 1 intervention group, the depression scores reduced also more quickly compared with the waiting list comparator group [63]. Perceptions of peer and social support significantly improved, and higher support was significantly related with lower depression symptoms [65]. An antenatal, first trimester eHealth intervention on depressive symptoms showed 80% intervention response and 60% remission (n=12) [63].

Low- and Middle-Income Countries

Limited resources and poor information are still leading to preventable maternal and neonatal deaths in low- and middle-income countries. The availability of mobile phones (in Africa and South-East Asia over 69%-90%) gives rise to the implementation of eHealth interventions and remote care. For more detailed information in this distinct population where eHealth is widely used, we refer to 2 recently published systematic reviews (Table 3). In summary, the interventions did increase antenatal care attendance, facility and service utilization, skilled support at birth, and vaccination rate [66]. Most of the included studies were of poor methodological quality or did not assess health outcomes [67]. Insufficient information was provided to evaluate the impact of eHealth solutions on maternal and fetal outcomes in these countries [67].

Telemonitoring and Teleconsulting

Telemonitoring of pregnancy is perceived to be one of the most promising answers to the possibilities of eHealth in pregnancy. Several hardware and software systems involving more complex remote monitoring are described lately (Table 3). An integrated system for maternal monitoring of glucose, weight, pulse and blood pressure, and a chat feature for clinician-patient contact is now in test [68]. Yi et al developed an Android-based mobile terminal for wireless fetal monitoring and uterine contractions tracking [69]. Using this system, patients in rural areas are provided with telemonitoring without traveling or hospitalization. Several other telemonitoring devices for cardiotocography have been tested in pilot settings or prospective cohorts and found feasible [70-72]. Currently, the effects of maternal and fetal telemonitoring in high-risk pregnancies on outcome, satisfaction, and costs are under research compared with hospital admission (the HOTEL trial, registered under #NTR6076). In a pilot with remote monitoring with transabdominal fetal electrocardiography (f-ECG) after induction with dinoprostone pessaries (n=70), successful monitoring was obtained in 89% [73]. Three women were recalled to the hospital due to suspicious f-ECG, of which in 2 cases caesarean section was indicated. A *Virtual Obstetric Care* program with normal visits combined with teleconferencing visits for low-risk pregnancy showed no increased risks in health outcomes besides an increase in preeclampsia diagnosis [74]. Another demonstration project describes a promising system of a wirelessly enabled maternal-fetal monitoring system *MiBebe*, used for the improvement of perinatal care in rural regions in Mexico. In the group of 153 high-risk pregnancies,

the remote monitoring in 74 patients resulted in markedly increased adherence to antenatal visits with no adverse health outcomes compared with usual care [75]. One pilot study describes an alternative prenatal care schedule, including an integrated technology platform (mobile app, wireless weight scale, and blood pressure cuff), leading to a 43% reduction in outpatient visits (8 vs 14 visits) [76]. There was an increase in satisfaction and patient engagement and no change in perinatal outcome despite the decrease in face-to-face contact [76]. Remote monitoring and consultation can potentially reduce outpatient visits for antenatal consultation as well as hospitalization for certain clinical reasons. We see this in managing gestational diabetes with glucose monitoring but also in fetal monitoring for fetal growth restriction [53,77]. A model of cost-effectiveness analysis in a tertiary hospital (Ghent, Belgium) predicted a cost-reduction of 145,822 euros per year achieved by introducing home monitoring in high-risk pregnancy [78].

Patient and Caregiver Experience

Examining patients' satisfaction with eHealth interventions, users describe high convenience and acceptance resulting in more patient activation and education. Patients report less concerns and anxiety and are comfortable with fewer clinic visits. Satisfaction rates vary between 86% and 95% in e-mental health studies and 90% (46/51) in home-monitored induction patients, who were very glad to stay in their own homely ambience as long as possible [73,79].

On the health care providers' point of view, adaptation of obstetricians and midwives to eHealth solutions has not been widely described. Only 1 qualitative study interviewed 12 health care providers in obstetric departments. Concerns were raised on implementation barriers and potential medico-legal risks, but if addressed properly, implementation was considered feasible. Some clinicians admitted to have insufficient familiarity and skill with eHealth limiting their engagement and comprehension of the possibilities that eHealth technologies can confer to perinatal care. Overall, these clinicians regarded telemedicine as an additional parallel service rather than integrated into the antenatal care model [27].

Discussion

Principal Findings

By providing this overview of the literature, we aimed to assess the applicability, advantages, and limitations of the use of eHealth in perinatal care. This review showed that eHealth interventions have a very broad, multilevel field of application focused on perinatal care in all its aspects. Most of the reviewed 71 articles were published after 2013, suggesting this novel type of care is an important topic of clinical and scientific relevance. Women of reproductive age seem to be interested in eHealth, as shown by their frequent use of smartphone, internet and apps, and searches for pregnancy information. Most health outcomes for perinatal eHealth interventions were generally positive, either resulting in positive effects (lifestyle, mental health) or providing multiple advantages while health outcomes were found equal (diabetes care). The implementation of telemonitoring was not studied extensively, but research

provided important effects and advantages on facilitation of new care models. Patient and care provider satisfaction with eHealth interventions rates are generally good, with rates up to 95%.

Additional Considerations

Despite the promising preliminary results as reviewed above, research in eHealth has progressed much slower than developments in the health technology industry. A great amount of the reviewed articles on this subject addressed more than health outcomes or satisfaction rates alone. Advances in (implementation of) apps and devices and patient-generated data are retained by legal and financial concerns. Possible privacy risks involve a lack of control to collection of data and the use by third parties afterwards.

In the United States, eHealth legislation, secured in the Fair Information Practice principles (part of the Health Insurance Portability and Accountability Act), is lacking protection for endpoint users: the patients. End-to-end data encryption can be used to protect the useful patient data. Combined with authentication and access control mechanisms for patients as well as care providers, eHealth technologies can further enhance final security control [106]. The development of the Telemedicine for Medicare Act of 2015 may accelerate the removal of barriers and limitations regarding use of telehealth between different states in the United States [107].

In the framework of European law, eHealth is simultaneously a health care service and an information service with corresponding legislation [108]. eHealth developers have to mind general legislation regarding privacy protection (Dir 95/46/EC, Arts 8-12), electronic identification services, e-Commerce directive (eg, online contracting), safety requirements of medical devices, and general product safety and liability requirements. In answer to the interstate

developments in eHealth care, the Cross Border Directive was initiated in 2011 in the European Union (EU). The objective of the initiatives within this directive is to turn telemedicine into a standard medical service, accessible to every European patient and fully covered by the respective social security system. Difficulties arise on liability and creating uniform rules in the EU, as member states have very intrinsic differences in national rules on health care, privacy, and liability. One advice would be for each member state to provide a legal framework for telemedicine, whereas the role of the EU would be limited to regulation [108].

The costs associated with development, purchase, and maintenance of eHealth equipment have dropped in recent years due to technological advancements [107]. Primary investments to implement eHealth in perinatal care are now attributed to personnel costs for both providers and technical support. However, to deliver care with the help of eHealth can also create savings on personnel costs and clinic visits. A systematic review of economic evaluation in telehealth solutions concluded that 29 out of 39 studies (74%) reported cost-effective, economically beneficial eHealth interventions in different conditions and diseases. The conclusion highlighted the fact that many studies did not report all recommended economic outcome items, leading to inconsistent analyses [109].

The challenges for reimbursement are delaying the widespread adoption of eHealth in all ranges of sections of hospital care. Coverage is fragmented, varying at level of country, within hospitals in the same country and within different specialties of health care [29]. Health insurance companies seem to be inclined to cover only well-researched eHealth interventions with according economic evaluations. The use of low-risk, inexpensive care models can operate as opportunities to objectify possible reduction in health care costs.

Textbox 1. Advantages and disadvantages of eHealth implementation in perinatal care.

Advantages

- Patient satisfaction
- Patient engagement
- Fewer clinic visits
- Clinician satisfaction
- Remote monitoring
- Access to care in low- and middle-income countries

Disadvantages

- Reimbursement
- Legal issues
- Technical issues

Indistinct

- Impact on health outcome
- Impact on costs
- Limited A-level evidence

Successes will motivate policy makers and drive the insurance market for additional coverage. Rigorous medical evidence can act as an extra stimulant; however, the duration and costs of designs and trials need to be taken into consideration [107].

Conclusion and Future Perspectives

This review provided an overview of eHealth as the next-generation perinatal care. [Textbox 1](#) provides a condensed summary of the advantages (as described in Principal Findings) and disadvantages (as described in Additional Considerations) of the implementation of eHealth in perinatal care. If eHealth is to achieve its full potential, it should attain all domains of quality in care including safety, timeliness, effectiveness, efficiency, and patient centeredness. Cost-effectiveness assessment is needed to rationalize embracement and reimbursement. Policy makers should consider the international frameworks of legislation to support and implement this new form of care.

We accentuate that more research is needed, including economic evaluation of eHealth interventions. Growing engagement of calls for funding have responded: more large funding associations focus on the use of eHealth, warranting the qualitative impact of the studies in the application designs [110]. In addition, the potential of technology raised a nearly quadrupled amount of money in venture capital funding, from US \$1.1 billion in 2011 to US \$4.3 billion in 2015 [111].

Despite the challenges of privacy, liability, and costs, eHealth is very likely to disperse globally in the next decade. Some even state health care is approaching a tipping point [112]. The current shift to patient-centered care and increased patient empowerment underlines the need for revising current medical practice. eHealth has the potential to be integrated into standard care and deliver a revolution in perinatal health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy and flow diagram.

[[PDF File \(Adobe PDF File\), 123KB - jmir_v20i6e202_app1.pdf](#)]

References

1. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Global Observatory for eHealth Series, Volume 3. Geneva, Switzerland: WHO; 2011.
2. Lettieri E, Fumagalli LP, Radaelli G, Bertele P, Vogt J, Hammerschmidt R, et al. Empowering patients through eHealth: a case report of a pan-European project. BMC Health Serv Res 2015 Aug 05;15:309 [[FREE Full text](#)] [doi: [10.1186/s12913-015-0983-0](#)] [Medline: [26242863](#)]
3. Fumagalli LP, Radaelli G, Lettieri E, Bertele P, Masella C. Patient empowerment and its neighbours: clarifying the boundaries and their mutual relationships. Health Policy 2015 Mar;119(3):384-394. [doi: [10.1016/j.healthpol.2014.10.017](#)] [Medline: [25467286](#)]
4. Jones N, Corrigan PW, James D, Parker J, Larson N. Peer support, self-determination, and treatment engagement: a qualitative investigation. Psychiatr Rehabil J 2013 Sep;36(3):209-214. [doi: [10.1037/prj0000008](#)] [Medline: [23815173](#)]
5. Barelo S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [[FREE Full text](#)] [doi: [10.3389/fpsyg.2015.02013](#)] [Medline: [26779108](#)]
6. Schwappach DL. Review: engaging patients as vigilant partners in safety: a systematic review. Med Care Res Rev 2010 Apr;67(2):119-148. [doi: [10.1177/1077558709342254](#)] [Medline: [19671916](#)]
7. Wallwiener S, Müller M, Doster A, Laserer W, Reck C, Pauluschke-Fröhlich J, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Nov;294(5):937-944. [doi: [10.1007/s00404-016-4093-y](#)] [Medline: [27084763](#)]
8. Bert F, Gualano MR, Brusaferrero S, De Vito E, de Waure C, La Torre G, et al. Pregnancy e-health: a multicenter Italian cross-sectional study on Internet use and decision-making among pregnant women. J Epidemiol Community Health 2013 Dec 01;67(12):1013-1018. [doi: [10.1136/jech-2013-202584](#)] [Medline: [24072743](#)]
9. Johnson SA. 'Intimate mothering publics': comparing face-to-face support groups and Internet use for women seeking information and advice in the transition to first-time motherhood. Cult Health Sex 2015;17(2):237-251. [doi: [10.1080/13691058.2014.968807](#)] [Medline: [25339096](#)]
10. Dutch Society of Obstetrics and Gynaecology. Vision Document 2011: Integrale verloskundige zorg in Nederland. Een stap verder. 2011. URL: https://www.nvog.nl/Sites/Files/0000002181_0000001819_Visiedocument.pdf [accessed 2018-05-03] [[WebCite Cache ID 6z8Dm2hdv](#)]
11. Rodger D, Skuse A, Wilmore M, Humphreys S, Dalton J, Flabouris M, et al. Pregnant women's use of information and communications technologies to access pregnancy-related health information in South Australia. Aust J Prim Health 2013;19(4):308-312. [doi: [10.1071/PY13029](#)] [Medline: [24004661](#)]

12. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. *BMC Pregnancy Childbirth* 2016 Mar 28;16:65 [FREE Full text] [doi: [10.1186/s12884-016-0856-5](https://doi.org/10.1186/s12884-016-0856-5)] [Medline: [27021727](https://pubmed.ncbi.nlm.nih.gov/27021727/)]
13. Scafoli G, Bert F, Galis V, Brusaferrero S, De Vito E, La Torre G, et al. Pregnancy and internet: sociodemographic and geographic differences in e-health practice. Results from an Italian multicenter study. *Public Health* 2015 Sep;129(9):1258-1266. [doi: [10.1016/j.puhe.2015.06.012](https://doi.org/10.1016/j.puhe.2015.06.012)] [Medline: [26210071](https://pubmed.ncbi.nlm.nih.gov/26210071/)]
14. Lee Y, Moon M. Utilization and content evaluation of mobile applications for pregnancy, birth, and child care. *Healthc Inform Res* 2016 Apr;22(2):73-80 [FREE Full text] [doi: [10.4258/hir.2016.22.2.73](https://doi.org/10.4258/hir.2016.22.2.73)] [Medline: [27200216](https://pubmed.ncbi.nlm.nih.gov/27200216/)]
15. Walker MG, Windrim C, Ellul KN, Kingdom JC. Web-based education for placental complications of pregnancy. *J Obstet Gynaecol Can* 2013 Apr;35(4):334-339. [doi: [10.1016/S1701-2163\(15\)30961-0](https://doi.org/10.1016/S1701-2163(15)30961-0)] [Medline: [23660041](https://pubmed.ncbi.nlm.nih.gov/23660041/)]
16. Wade VK, Karnon J, Elshaug AG, Hiller JE. A systematic review of economic analyses of telehealth services using real time video communication. *BMC Health Serv Res* 2010 Aug 10;10:233 [FREE Full text] [doi: [10.1186/1472-6963-10-233](https://doi.org/10.1186/1472-6963-10-233)] [Medline: [20696073](https://pubmed.ncbi.nlm.nih.gov/20696073/)]
17. Narasimhulu DM, Karakash S, Weedon J, Minkoff H. Patterns of internet use by pregnant women, and reliability of pregnancy-related searches. *Matern Child Health J* 2016 Dec;20(12):2502-2509. [doi: [10.1007/s10995-016-2075-0](https://doi.org/10.1007/s10995-016-2075-0)] [Medline: [27456311](https://pubmed.ncbi.nlm.nih.gov/27456311/)]
18. Gao LL, Larsson M, Luo SY. Internet use by Chinese women seeking pregnancy-related information. *Midwifery* 2013 Jul;29(7):730-735. [doi: [10.1016/j.midw.2012.07.003](https://doi.org/10.1016/j.midw.2012.07.003)] [Medline: [22958935](https://pubmed.ncbi.nlm.nih.gov/22958935/)]
19. Knight-Agarwal C, Davis DL, Williams L, Davey R, Cox R, Clarke A. Development and pilot testing of the Eating4two mobile phone app to monitor gestational weight gain. *JMIR Mhealth Uhealth* 2015 Jun 05;3(2):e44 [FREE Full text] [doi: [10.2196/mhealth.4071](https://doi.org/10.2196/mhealth.4071)] [Medline: [26048313](https://pubmed.ncbi.nlm.nih.gov/26048313/)]
20. Buultjens M, Robinson P, Milgrom J. Online resources for new mothers: opportunities and challenges for perinatal health professionals. *J Perinat Educ* 2012;21(2):99-111 [FREE Full text] [doi: [10.1891/1058-1243.21.2.99](https://doi.org/10.1891/1058-1243.21.2.99)] [Medline: [23449623](https://pubmed.ncbi.nlm.nih.gov/23449623/)]
21. Vlemmix F, Warendorf J, Rosman A, Kok M, Mol B, Morris J, et al. Decision aids to improve informed decision-making in pregnancy care: a systematic review. *BJOG* 2013 Feb;120(3):257-266 [FREE Full text] [doi: [10.1111/1471-0528.12060](https://doi.org/10.1111/1471-0528.12060)] [Medline: [23145991](https://pubmed.ncbi.nlm.nih.gov/23145991/)]
22. Ledford CJ, Canzona MR, Cafferty LA, Hodge JA. Mobile application as a prenatal education and engagement tool: a randomized controlled pilot. *Patient Educ Couns* 2016 Apr;99(4):578-582. [doi: [10.1016/j.pec.2015.11.006](https://doi.org/10.1016/j.pec.2015.11.006)] [Medline: [26610389](https://pubmed.ncbi.nlm.nih.gov/26610389/)]
23. Bush J, Barlow DE, Echols J, Wilkerson J, Bellevin K. Impact of a mobile health application on user engagement and pregnancy outcomes among Wyoming Medicaid members. *Telemed J E Health* 2017 Nov;23(11):891-898 [FREE Full text] [doi: [10.1089/tmj.2016.0242](https://doi.org/10.1089/tmj.2016.0242)] [Medline: [28481167](https://pubmed.ncbi.nlm.nih.gov/28481167/)]
24. Peragallo UR, Berger AA, Ivins AA, Beckham A, Thorp Jr JM, Nicholson WK. Internet use and access among pregnant women via computer and mobile phone: implications for delivery of perinatal care. *JMIR Mhealth Uhealth* 2015 Mar 30;3(1):e25 [FREE Full text] [doi: [10.2196/mhealth.3347](https://doi.org/10.2196/mhealth.3347)] [Medline: [25835744](https://pubmed.ncbi.nlm.nih.gov/25835744/)]
25. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. *BMC Pregnancy Childbirth* 2016 Dec 19;16(1):171 [FREE Full text] [doi: [10.1186/s12884-016-0971-3](https://doi.org/10.1186/s12884-016-0971-3)] [Medline: [27435182](https://pubmed.ncbi.nlm.nih.gov/27435182/)]
26. Goetz M, Müller M, Matthies LM, Hansen J, Doster A, Szabo A, et al. Perceptions of patient engagement applications during pregnancy: a qualitative assessment of the patient's perspective. *JMIR Mhealth Uhealth* 2017 May 26;5(5):e73 [FREE Full text] [doi: [10.2196/mhealth.7040](https://doi.org/10.2196/mhealth.7040)] [Medline: [28550005](https://pubmed.ncbi.nlm.nih.gov/28550005/)]
27. Willcox J, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of women and health professionals on mhealth lifestyle interventions in pregnancy: a qualitative investigation. *JMIR Mhealth Uhealth* 2015 Oct 28;3(4):e99 [FREE Full text] [doi: [10.2196/mhealth.4869](https://doi.org/10.2196/mhealth.4869)] [Medline: [26510886](https://pubmed.ncbi.nlm.nih.gov/26510886/)]
28. Mackert M, Guadagno M, Donovan E, Whitten P. Including men in prenatal health: the potential of e-health to improve birth outcomes. *Telemed J E Health* 2015 Mar;21(3):207-212. [doi: [10.1089/tmj.2014.0048](https://doi.org/10.1089/tmj.2014.0048)] [Medline: [25489723](https://pubmed.ncbi.nlm.nih.gov/25489723/)]
29. Schwamm LH, Chumbler N, Brown E, Fonarow GC, Berube D, Nystrom K, et al. Recommendations for the implementation of telehealth in cardiovascular and stroke care: a policy statement from the American Heart Association. *Circulation* 2017 Feb 14;135(7):e24-e44 [FREE Full text] [doi: [10.1161/CIR.0000000000000475](https://doi.org/10.1161/CIR.0000000000000475)] [Medline: [27998940](https://pubmed.ncbi.nlm.nih.gov/27998940/)]
30. Muktabhant B, Lawrie TA, Lumbiganon P, Laopaiboon M. Diet or exercise, or both, for preventing excessive weight gain in pregnancy. *Cochrane Database Syst Rev* 2015 Jun 15(6):CD007145. [doi: [10.1002/14651858.CD007145.pub3](https://doi.org/10.1002/14651858.CD007145.pub3)] [Medline: [26068707](https://pubmed.ncbi.nlm.nih.gov/26068707/)]
31. Fell DB, Joseph KS, Armson BA, Dodds L. The impact of pregnancy on physical activity level. *Matern Child Health J* 2009 Sep;13(5):597-603. [doi: [10.1007/s10995-008-0404-7](https://doi.org/10.1007/s10995-008-0404-7)] [Medline: [18719984](https://pubmed.ncbi.nlm.nih.gov/18719984/)]
32. Di Mascio D, Magro-Malosso ER, Saccone G, Marhefka GD, Berghella V. Exercise during pregnancy in normal-weight women and risk of preterm birth: a systematic review and meta-analysis of randomized controlled trials. *Am J Obstet Gynecol* 2016 Nov;215(5):561-571. [doi: [10.1016/j.ajog.2016.06.014](https://doi.org/10.1016/j.ajog.2016.06.014)] [Medline: [27319364](https://pubmed.ncbi.nlm.nih.gov/27319364/)]
33. Waring ME, Moore Simas TA, Xiao RS, Lombardini LM, Allison JJ, Rosal MC, et al. Pregnant women's interest in a website or mobile application for healthy gestational weight gain. *Sex Reprod Healthc* 2014 Dec;5(4):182-184 [FREE Full text] [doi: [10.1016/j.srhc.2014.05.002](https://doi.org/10.1016/j.srhc.2014.05.002)] [Medline: [25433828](https://pubmed.ncbi.nlm.nih.gov/25433828/)]

34. Hayman M, Reaburn P, Browne M, Vandelanotte C, Alley S, Short CE. Feasibility, acceptability and efficacy of a web-based computer-tailored physical activity intervention for pregnant women - the Fit4Two randomised controlled trial. *BMC Pregnancy Childbirth* 2017 Dec 23;17(1):96 [FREE Full text] [doi: [10.1186/s12884-017-1277-9](https://doi.org/10.1186/s12884-017-1277-9)] [Medline: [28335767](#)]
35. Soltani H, Duxbury AM, Arden MA, Dearden A, Furness PJ, Garland C. Maternal obesity management using mobile technology: a feasibility study to evaluate a text messaging based complex intervention during pregnancy. *J Obes* 2015;2015:814830 [FREE Full text] [doi: [10.1155/2015/814830](https://doi.org/10.1155/2015/814830)] [Medline: [25960889](#)]
36. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. *Matern Child Health J* 2016 May;20(5):1091-1101 [FREE Full text] [doi: [10.1007/s10995-015-1895-7](https://doi.org/10.1007/s10995-015-1895-7)] [Medline: [26649879](#)]
37. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. *BJOG* 2017 Oct;124(11):1718-1728. [doi: [10.1111/1471-0528.14552](https://doi.org/10.1111/1471-0528.14552)] [Medline: [28220604](#)]
38. Agarwal S, Labrique A. Newborn health on the line: the potential mHealth applications. *J Am Med Assoc* 2014 Jul 16;312(3):229-230. [doi: [10.1001/jama.2014.6371](https://doi.org/10.1001/jama.2014.6371)] [Medline: [24953141](#)]
39. Graham ML, Uesugi KH, Niederdeppe J, Gay GK, Olson CM. The theory, development, and implementation of an e-intervention to prevent excessive gestational weight gain: e-Moms Roc. *Telemed J E Health* 2014 Dec;20(12):1135-1142 [FREE Full text] [doi: [10.1089/tmj.2013.0354](https://doi.org/10.1089/tmj.2013.0354)] [Medline: [25354350](#)]
40. Tong VT, Dietz PM, Morrow B, D'Angelo DV, Farr SL, Rockhill KM, Centers for Disease Control and Prevention (CDC). Trends in smoking before, during, and after pregnancy--Pregnancy Risk Assessment Monitoring System, United States, 40 sites, 2000-2010. *MMWR Surveill Summ* 2013 Nov 08;62(6):1-19 [FREE Full text] [Medline: [24196750](#)]
41. McLeod D, Pullon S, Cookson T. Factors that influence changes in smoking behaviour during pregnancy. *N Z Med J* 2003 May 02;116(1173):U418. [Medline: [12740612](#)]
42. Heminger CL, Schindler-Ruwisch JM, Abrams LC. Smoking cessation support for pregnant women: role of mobile technology. *Subst Abuse Rehabil* 2016;7:15-26 [FREE Full text] [doi: [10.2147/SAR.S84239](https://doi.org/10.2147/SAR.S84239)] [Medline: [27110146](#)]
43. National Institute for Health and Care Excellence. 2015. Diabetes in pregnancy: management from preconception to the postnatal period URL: <https://www.nice.org.uk/guidance/ng3> [accessed 2018-05-03] [WebCite Cache ID 6z8LdUImu]
44. Wong VW, Jalaludin B. Gestational diabetes mellitus: who requires insulin therapy? *Aust N Z J Obstet Gynaecol* 2011 Oct;51(5):432-436. [doi: [10.1111/j.1479-828X.2011.01329.x](https://doi.org/10.1111/j.1479-828X.2011.01329.x)] [Medline: [21806589](#)]
45. Khorshidi Roozbahani R, Geranmayeh M, Hantoushzadeh S, Mehran A. Effects of telephone follow-up on blood glucose levels and postpartum screening in mothers with Gestational Diabetes Mellitus. *Med J Islam Repub Iran* 2015;29:249 [FREE Full text] [Medline: [26793640](#)]
46. Chillemi NC, Dalfrà MG, Lapolla A. The emerging role of telemedicine in managing glycemic control and psychobehavioral aspects of pregnancy complicated by diabetes. *Int J Telemed Appl* 2014;2014:621384 [FREE Full text] [doi: [10.1155/2014/621384](https://doi.org/10.1155/2014/621384)] [Medline: [25295059](#)]
47. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. *Diabetes Technol Ther* 2007 Jun;9(3):297-306. [doi: [10.1089/dia.2006.0034](https://doi.org/10.1089/dia.2006.0034)] [Medline: [17561800](#)]
48. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogianis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. *Diabetes Technol Ther* 2012 Jul;14(7):624-629 [FREE Full text] [doi: [10.1089/dia.2012.0010](https://doi.org/10.1089/dia.2012.0010)] [Medline: [22512287](#)]
49. Carral F, Ayala Mdel C, Fernández JJ, González C, Piñero A, García G, et al. Web-based telemedicine system is useful for monitoring glucose control in pregnant women with diabetes. *Diabetes Technol Ther* 2015 May;17(5):349-354. [doi: [10.1089/dia.2014.0223](https://doi.org/10.1089/dia.2014.0223)] [Medline: [25629547](#)]
50. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-Mum: a feasibility study for a randomized controlled trial exploring the potential for telemedicine in the diabetes care of those with gestational diabetes. *Diabetes Technol Ther* 2015 Dec;17(12):880-888. [doi: [10.1089/dia.2015.0147](https://doi.org/10.1089/dia.2015.0147)] [Medline: [26394017](#)]
51. Mackillop L, Loerup L, Bartlett K, Farmer A, Gibson OJ, Hirst JE, et al. Development of a real-time smartphone solution for the management of women with or at high risk of gestational diabetes. *J Diabetes Sci Technol* 2014 Nov;8(6):1105-1114 [FREE Full text] [doi: [10.1177/1932296814542271](https://doi.org/10.1177/1932296814542271)] [Medline: [25004915](#)]
52. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. *Int J Endocrinol* 2010;2010:386941 [FREE Full text] [doi: [10.1155/2010/386941](https://doi.org/10.1155/2010/386941)] [Medline: [20628517](#)]
53. Rasekaba TM, Furler J, Blackberry I, Tacey M, Gray K, Lim K. Telemedicine interventions for gestational diabetes mellitus: a systematic review and meta-analysis. *Diabetes Res Clin Pract* 2015 Oct;110(1):1-9. [doi: [10.1016/j.diabres.2015.07.007](https://doi.org/10.1016/j.diabres.2015.07.007)] [Medline: [26264410](#)]
54. Ming WK, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. *J Med Internet Res* 2016 Nov 09;18(11):e290 [FREE Full text] [doi: [10.2196/jmir.6556](https://doi.org/10.2196/jmir.6556)] [Medline: [27829574](#)]

55. Noctor E, Dunne FP. Type 2 diabetes after gestational diabetes: the influence of changing diagnostic criteria. *World J Diabetes* 2015 Mar 15;6(2):234-244 [[FREE Full text](#)] [doi: [10.4239/wjd.v6.i2.234](https://doi.org/10.4239/wjd.v6.i2.234)] [Medline: [25789105](#)]
56. 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](https://doi.org/10.1371/journal.pone.0013196)] [Medline: [20967242](#)]
57. Beck CT. The lived experience of postpartum depression: a phenomenological study. *Nurs Res* 1992;41(3):166-170. [Medline: [1584660](#)]
58. Mitchell AM, Mittelstaedt ME, Schott-Baer D. Postpartum depression: the reliability of telephone screening. *MCN Am J Matern Child Nurs* 2006;31(6):382-387. [Medline: [17149115](#)]
59. Jiménez-Serrano S, Tortajada S, García-Gómez JM. A mobile health application to predict postpartum depression based on machine learning. *Telemed J E Health* 2015 Jul;21(7):567-574. [doi: [10.1089/tmj.2014.0113](https://doi.org/10.1089/tmj.2014.0113)] [Medline: [25734829](#)]
60. Kingston D, Austin MP, Veldhuyzen van Zanten S, Harvalik P, Giallo R, McDonald SD, et al. Pregnant women's views on the feasibility and acceptability of web-based mental health e-screening versus paper-based screening: a randomized controlled trial. *J Med Internet Res* 2017 Apr 07;19(4):e88 [[FREE Full text](#)] [doi: [10.2196/jmir.6866](https://doi.org/10.2196/jmir.6866)] [Medline: [28389421](#)]
61. Lee EW, Denison FC, Hor K, Reynolds RM. Web-based interventions for prevention and treatment of perinatal mood disorders: a systematic review. *BMC Pregnancy Childbirth* 2016 Feb 29;16:38 [[FREE Full text](#)] [doi: [10.1186/s12884-016-0831-1](https://doi.org/10.1186/s12884-016-0831-1)] [Medline: [26928898](#)]
62. Posmontier B, Neugebauer R, Stuart S, Chittams J, Shaughnessy R. Telephone-administered interpersonal psychotherapy by nurse-midwives for postpartum depression. *J Midwifery Womens Health* 2016 Jul;61(4):456-466. [doi: [10.1111/jmwh.12411](https://doi.org/10.1111/jmwh.12411)] [Medline: [26970401](#)]
63. Ashford MT, Olander EK, Ayers S. Computer- or web-based interventions for perinatal mental health: a systematic review. *J Affect Disord* 2016 Jun;197:134-146. [doi: [10.1016/j.jad.2016.02.057](https://doi.org/10.1016/j.jad.2016.02.057)] [Medline: [26991368](#)]
64. Lau Y, Htun TP, Wong SN, Tam WS, Klainin-Yobas P. Therapist-supported internet-based cognitive behavior therapy for stress, anxiety, and depressive symptoms among postpartum women: a systematic review and meta-analysis. *J Med Internet Res* 2017 Apr 28;19(4):e138 [[FREE Full text](#)] [doi: [10.2196/jmir.6712](https://doi.org/10.2196/jmir.6712)] [Medline: [28455276](#)]
65. Letourneau N, Secco L, Colpitts J, Aldous S, Stewart M, Dennis CL. Quasi-experimental evaluation of a telephone-based peer support intervention for maternal depression. *J Adv Nurs* 2015 Jul;71(7):1587-1599. [doi: [10.1111/jan.12622](https://doi.org/10.1111/jan.12622)] [Medline: [25705786](#)]
66. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. *PLoS One* 2016;11(5):e0154664 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0154664](https://doi.org/10.1371/journal.pone.0154664)] [Medline: [27144393](#)]
67. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. *J Glob Health* 2016 Jun;6(1):010401 [[FREE Full text](#)] [doi: [10.7189/jogh.06.010401](https://doi.org/10.7189/jogh.06.010401)] [Medline: [26649177](#)]
68. Robu A, Gauca B, Crisan-Vida M, Stoicu-Tivadar L. Integrated system for monitoring and prevention in obstetrics-gynaecology. *Stud Health Technol Inform* 2016;221:8-12. [Medline: [27071866](#)]
69. Xi H, Gan G, Zhang H, Chen C. [Design of Smart Care Tele-Monitoring System for Mother and Fetus]. *Zhongguo Yi Liao Qi Xie Za Zhi* 2015 Mar;39(2):102-104. [Medline: [26204738](#)]
70. Hod M, Kerner R. Telemedicine for antenatal surveillance of high-risk pregnancies with ambulatory and home fetal heart rate monitoring--an update. *J Perinat Med* 2003;31(3):195-200. [doi: [10.1515/JPM.2003.026](https://doi.org/10.1515/JPM.2003.026)] [Medline: [12825474](#)]
71. Nițulescu A, Crișan-Vida M, Stoicu-Tivadar L, Bernad E. Integrated wireless sensor network for monitoring pregnant women. *Stud Health Technol Inform* 2015;210:354-358. [Medline: [25991165](#)]
72. Vermeulen-Giovagnoli B, Peters C, van der Hout-van der Jagt MB, Mischi M, van Pul C, Cottaar EJ, et al. The development of an obstetric tele-monitoring system. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015:177-180. [doi: [10.1109/EMBC.2015.7318329](https://doi.org/10.1109/EMBC.2015.7318329)] [Medline: [26736229](#)]
73. Rauf Z, O'Brien E, Stampalija T, Ilioniu FP, Lavender T, Alfirevic Z. Home labour induction with retrievable prostaglandin pessary and continuous telemetric trans-abdominal fetal ECG monitoring. *PLoS One* 2011;6(11):e28129 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0028129](https://doi.org/10.1371/journal.pone.0028129)] [Medline: [22140522](#)]
74. Pflugeisen BM, McCarren C, Poore S, Carlile M, Schroeder R. Virtual visits: managing prenatal care with modern technology. *MCN Am J Matern Child Nurs* 2016;41(1):24-30. [doi: [10.1097/NMC.000000000000199](https://doi.org/10.1097/NMC.000000000000199)] [Medline: [26474477](#)]
75. Tapia-Conyer R, Lyford S, Saucedo R, Casale M, Gallardo H, Becerra K, et al. Improving perinatal care in the rural regions worldwide by wireless enabled antepartum fetal monitoring: a demonstration project. *Int J Telemed Appl* 2015;2015:794180 [[FREE Full text](#)] [doi: [10.1155/2015/794180](https://doi.org/10.1155/2015/794180)] [Medline: [25691900](#)]
76. Marko K, Ganju N, Brown J, Benham J, Gaba ND. Remote prenatal care monitoring with digital health tools can reduce visit frequency while improving satisfaction. *Obstet Gynecol* 2016;127(1):1. [doi: [10.1097/01.AOG.0000483620.40988.df](https://doi.org/10.1097/01.AOG.0000483620.40988.df)]
77. Kerner R, Yogev Y, Belkin A, Ben-Haroush A, Zeevi B, Hod M. Maternal self-administered fetal heart rate monitoring and transmission from home in high-risk pregnancies. *Int J Gynaecol Obstet* 2004 Jan;84(1):33-39. [Medline: [14698827](#)]

78. Buysse H, De Moor MG, Van Maele G, Baert E, Thienpont G, Temmerman M. Cost-effectiveness of telemonitoring for high-risk pregnant women. *Int J Med Inform* 2008 Jul;77(7):470-476. [doi: [10.1016/j.ijmedinf.2007.08.009](https://doi.org/10.1016/j.ijmedinf.2007.08.009)] [Medline: [17923433](https://pubmed.ncbi.nlm.nih.gov/17923433/)]
79. O'Brien E, Rauf Z, Alfirevic Z, Lavender T. Women's experiences of outpatient induction of labour with remote continuous monitoring. *Midwifery* 2013 Apr;29(4):325-331. [doi: [10.1016/j.midw.2012.01.014](https://doi.org/10.1016/j.midw.2012.01.014)] [Medline: [23159160](https://pubmed.ncbi.nlm.nih.gov/23159160/)]
80. Pollak KI, Alexander SC, Bennett G, Lyna P, Coffman CJ, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: a pilot study. *Patient Educ Couns* 2014 Nov;97(2):256-260 [FREE Full text] [doi: [10.1016/j.pec.2014.07.030](https://doi.org/10.1016/j.pec.2014.07.030)] [Medline: [25153313](https://pubmed.ncbi.nlm.nih.gov/25153313/)]
81. Graham ML, Strawderman MS, Demment M, Olson CM. Does usage of an ehealth intervention reduce the risk of excessive gestational weight gain? Secondary analysis from a randomized controlled trial. *J Med Internet Res* 2017 Jan 09;19(1):e6 [FREE Full text] [doi: [10.2196/jmir.6644](https://doi.org/10.2196/jmir.6644)] [Medline: [28069560](https://pubmed.ncbi.nlm.nih.gov/28069560/)]
82. Huberty JL, Buman MP, Leiferman JA, Bushar J, Hekler EB, Adams MA. Dose and timing of text messages for increasing physical activity among pregnant women: a randomized controlled trial. *Transl Behav Med* 2017 Jun;7(2):212-223 [FREE Full text] [doi: [10.1007/s13142-016-0445-1](https://doi.org/10.1007/s13142-016-0445-1)] [Medline: [27800565](https://pubmed.ncbi.nlm.nih.gov/27800565/)]
83. Lewis BA, Martinson BC, Sherwood NE, Avery MD. A pilot study evaluating a telephone-based exercise intervention for pregnant and postpartum women. *J Midwifery Womens Health* 2011;56(2):127-131. [doi: [10.1111/j.1542-2011.2010.00016.x](https://doi.org/10.1111/j.1542-2011.2010.00016.x)] [Medline: [21429076](https://pubmed.ncbi.nlm.nih.gov/21429076/)]
84. Guo SH, Lee CW, Tsao CM, Hsing HC. A social media-based mindful yoga program for pregnant women in Taiwan. *Stud Health Technol Inform* 2016;225:621-622. [Medline: [27332280](https://pubmed.ncbi.nlm.nih.gov/27332280/)]
85. Kruger DF, White K, Galpern A, Mann K, Massirio A, McLellan M, et al. Effect of modem transmission of blood glucose data on telephone consultation time, clinic work flow, and patient satisfaction for patients with gestational diabetes mellitus. *J Am Acad Nurse Pract* 2003 Aug;15(8):371-375. [Medline: [14509102](https://pubmed.ncbi.nlm.nih.gov/14509102/)]
86. Dalfrà MG, Nicolucci A, Lapolla A, TISG. The effect of telemedicine on outcome and quality of life in pregnant women with diabetes. *J Telemed Telecare* 2009;15(5):238-242. [doi: [10.1258/jtt.2009.081213](https://doi.org/10.1258/jtt.2009.081213)] [Medline: [19590029](https://pubmed.ncbi.nlm.nih.gov/19590029/)]
87. Wojcicki JM, Ladyzynski P, Foltynski P. What we can really expect from telemedicine in intensive diabetes treatment: 10 years later. *Diabetes Technol Ther* 2013 Mar;15(3):260-268. [doi: [10.1089/dia.2012.0242](https://doi.org/10.1089/dia.2012.0242)] [Medline: [23343333](https://pubmed.ncbi.nlm.nih.gov/23343333/)]
88. Nicholson WK, Beckham AJ, Hatley K, Diamond M, Johnson LS, Green SL, et al. The Gestational Diabetes Management System (GoodMomS): development, feasibility and lessons learned from a patient-informed, web-based pregnancy and postpartum lifestyle intervention. *BMC Pregnancy Childbirth* 2016 Sep 21;16(1):277 [FREE Full text] [doi: [10.1186/s12884-016-1064-z](https://doi.org/10.1186/s12884-016-1064-z)] [Medline: [27654119](https://pubmed.ncbi.nlm.nih.gov/27654119/)]
89. Ganapathy R, Grewal A, Castleman JS. Remote monitoring of blood pressure to reduce the risk of preeclampsia related complications with an innovative use of mobile technology. *Pregnancy Hypertens* 2016 Oct;6(4):263-265. [doi: [10.1016/j.preghy.2016.04.005](https://doi.org/10.1016/j.preghy.2016.04.005)] [Medline: [27939464](https://pubmed.ncbi.nlm.nih.gov/27939464/)]
90. Harrison TN, Sacks DA, Parry C, Macias M, Ling Grant DS, Lawrence JM. Acceptability of virtual prenatal visits for women with gestational diabetes. *Womens Health Issues* 2017;27(3):351-355. [Medline: [28153743](https://pubmed.ncbi.nlm.nih.gov/28153743/)]
91. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of Mummoodbooster. *J Med Internet Res* 2016 Mar 07;18(3):e54 [FREE Full text] [doi: [10.2196/jmir.4993](https://doi.org/10.2196/jmir.4993)] [Medline: [26952645](https://pubmed.ncbi.nlm.nih.gov/26952645/)]
92. Ngai FW, Wong PW, Leung KY, Chau PH, Chung KF. The effect of telephone-based cognitive-behavioral therapy on postnatal depression: a randomized controlled trial. *Psychother Psychosom* 2015;84(5):294-303. [doi: [10.1159/000430449](https://doi.org/10.1159/000430449)] [Medline: [26278623](https://pubmed.ncbi.nlm.nih.gov/26278623/)]
93. Shamshiri Milani H, Azargashb E, Beyraghi N, Defaie S, Asbaghi T. Effect of telephone-based support on postpartum depression: a randomized controlled trial. *Int J Fertil Steril* 2015;9(2):247-253 [FREE Full text] [Medline: [26246884](https://pubmed.ncbi.nlm.nih.gov/26246884/)]
94. Fontein-Kuipers YJ, Nieuwenhuijze MJ, Ausems M, Budé L, de Vries R. Antenatal interventions to reduce maternal distress: a systematic review and meta-analysis of randomised trials. *BJOG* 2014 Mar;121(4):389-397 [FREE Full text] [doi: [10.1111/1471-0528.12500](https://doi.org/10.1111/1471-0528.12500)] [Medline: [24397691](https://pubmed.ncbi.nlm.nih.gov/24397691/)]
95. 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](https://doi.org/10.2196/mental.4074)] [Medline: [26543910](https://pubmed.ncbi.nlm.nih.gov/26543910/)]
96. Figueiredo FP, Parada AP, Cardoso VC, Batista RF, Silva AA, Barbieri MA, et al. Postpartum depression screening by telephone: a good alternative for public health and research. *Arch Womens Ment Health* 2015 Jun;18(3):547-553. [doi: [10.1007/s00737-014-0480-1](https://doi.org/10.1007/s00737-014-0480-1)] [Medline: [25380783](https://pubmed.ncbi.nlm.nih.gov/25380783/)]
97. Pugh NE, Hadjistavropoulos HD, Fuchs CM. Internet therapy for postpartum depression: a case illustration of emailed therapeutic assistance. *Arch Womens Ment Health* 2014 Aug;17(4):327-337. [doi: [10.1007/s00737-014-0439-2](https://doi.org/10.1007/s00737-014-0439-2)] [Medline: [24965602](https://pubmed.ncbi.nlm.nih.gov/24965602/)]
98. Pineros-Leano M, Tabb K, Sears H, Meline B, Huang H. Clinic staff attitudes towards the use of mHealth technology to conduct perinatal depression screenings: a qualitative study. *Fam Pract* 2015 Apr;32(2):211-215. [doi: [10.1093/fampra/cmu083](https://doi.org/10.1093/fampra/cmu083)] [Medline: [25535280](https://pubmed.ncbi.nlm.nih.gov/25535280/)]

99. Ivey TL, Hughes D, Dajani NK, Magann EF. Antenatal management of at-risk pregnancies from a distance. *Aust N Z J Obstet Gynaecol* 2015 Feb;55(1):87-89. [doi: [10.1111/ajo.12254](https://doi.org/10.1111/ajo.12254)] [Medline: [25355283](https://pubmed.ncbi.nlm.nih.gov/25355283/)]
100. Cuneo BF, Moon-Grady AJ, Sonesson SE, Levasseur S, Hornberger L, Donofrio MT, et al. Heart sounds at home: feasibility of an ambulatory fetal heart rhythm surveillance program for anti-SSA-positive pregnancies. *J Perinatol* 2017 Mar;37(3):226-230. [doi: [10.1038/jp.2016.220](https://doi.org/10.1038/jp.2016.220)] [Medline: [27977016](https://pubmed.ncbi.nlm.nih.gov/27977016/)]
101. Krishnamurti T, Davis AL, Wong-Parodi G, Fischhoff B, Sadovsky Y, Simhan HN. Development and testing of the Myhealthypregnancy app: a behavioral decision research-based tool for assessing and communicating pregnancy risk. *JMIR Mhealth Uhealth* 2017 Apr 10;5(4):e42 [FREE Full text] [doi: [10.2196/mhealth.7036](https://doi.org/10.2196/mhealth.7036)] [Medline: [28396302](https://pubmed.ncbi.nlm.nih.gov/28396302/)]
102. Rhoads SJ, Serrano CI, Lynch CE, Ounpraseuth ST, Gauss CH, Payakachat N, et al. Exploring implementation of m-health monitoring in postpartum women with hypertension. *Telemed J E Health* 2017 Oct;23(10):833-841. [doi: [10.1089/tmj.2016.0272](https://doi.org/10.1089/tmj.2016.0272)] [Medline: [28475431](https://pubmed.ncbi.nlm.nih.gov/28475431/)]
103. Marko KI, Krapf JM, Meltzer AC, Oh J, Ganju N, Martinez AG, et al. Testing the feasibility of remote patient monitoring in prenatal care using a mobile app and connected devices: a prospective observational trial. *JMIR Res Protoc* 2016 Nov 18;5(4):e200 [FREE Full text] [doi: [10.2196/resprot.6167](https://doi.org/10.2196/resprot.6167)] [Medline: [27864167](https://pubmed.ncbi.nlm.nih.gov/27864167/)]
104. Lanssens D, Vandenberk T, Smeets CJ, De Cannière H, Molenberghs G, Van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. *JMIR Mhealth Uhealth* 2017 Mar 09;5(3):e25 [FREE Full text] [doi: [10.2196/mhealth.6552](https://doi.org/10.2196/mhealth.6552)] [Medline: [28279948](https://pubmed.ncbi.nlm.nih.gov/28279948/)]
105. Pflugeisen BM, Mou J. Patient satisfaction with virtual obstetric care. *Matern Child Health J* 2017 Jul;21(7):1544-1551. [doi: [10.1007/s10995-017-2284-1](https://doi.org/10.1007/s10995-017-2284-1)] [Medline: [28176034](https://pubmed.ncbi.nlm.nih.gov/28176034/)]
106. Hall JL, McGraw D. For telehealth to succeed, privacy and security risks must be identified and addressed. *Health Aff (Millwood)* 2014 Feb;33(2):216-221. [doi: [10.1377/hlthaff.2013.0997](https://doi.org/10.1377/hlthaff.2013.0997)] [Medline: [24493763](https://pubmed.ncbi.nlm.nih.gov/24493763/)]
107. Dorsey ER, Topol EJ. State of Telehealth. *N Engl J Med* 2016 Jul 14;375(2):154-161. [doi: [10.1056/NEJMr1601705](https://doi.org/10.1056/NEJMr1601705)] [Medline: [27410924](https://pubmed.ncbi.nlm.nih.gov/27410924/)]
108. Raposo VL. Telemedicine: the legal framework (or the lack of it) in Europe. *GMS Health Technol Assess* 2016;12:Doc03 [FREE Full text] [doi: [10.3205/hta000126](https://doi.org/10.3205/hta000126)] [Medline: [27579146](https://pubmed.ncbi.nlm.nih.gov/27579146/)]
109. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. *PLoS One* 2017;12(2):e0170581 [FREE Full text] [doi: [10.1371/journal.pone.0170581](https://doi.org/10.1371/journal.pone.0170581)] [Medline: [28152012](https://pubmed.ncbi.nlm.nih.gov/28152012/)]
110. Kurzweil R. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking; 2005.
111. Wang T, King E, Perman M, Tecco H. Rockhealth. Digital health funding: 2015 year in review URL: <https://rockhealth.com/reports/digital-health-funding-2015-year-in-review/> [accessed 2018-05-03] [WebCite Cache ID 6z8UGkXhI]
112. Gladwell M. *The Tipping Point: How Little Things Can Make A Big Difference*. US: Little, Brown; 2000.

Abbreviations

Anti-SSA: Anti-Sjögren's-syndrome-related antigen A
eHealth: electronic health
e-mental health: electronic mental health
EU: European Union
f-ECG: fetal electrocardiography
GDM: gestational diabetes mellitus
RCT: randomized controlled trial
SMS: short message service

Edited by G Eysenbach; submitted 02.11.17; peer-reviewed by M Ashford, T Rasekaba, S Wallwiener, S Aggarwal; comments to author 14.12.17; revised version received 19.01.18; accepted 10.03.18; published 05.06.18.

Please cite as:

van den Heuvel JFM, Groenhof TK, Veerbeek JHW, van Solinge WW, Lely AT, Franx A, Bekker MN
 eHealth as the Next-Generation Perinatal Care: An Overview of the Literature
J Med Internet Res 2018;20(6):e202
 URL: <http://www.jmir.org/2018/6/e202/>
 doi:[10.2196/jmir.9262](https://doi.org/10.2196/jmir.9262)
 PMID:

©Josephus FM van den Heuvel, T Katrien Groenhof, Jan HW Veerbeek, Wouter W van Solinge, A Titia Lely, Arie Franx, Mireille N Bekker. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.06.2018. 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.

Original Paper

Relevance of a Mobile Internet Platform for Capturing Inter- and Intrasubject Variabilities in Circadian Coordination During Daily Routine: Pilot Study

Sandra Komarzynski^{1,2*}, Dipl Ing, MSc; Qi Huang^{1,3*}, PhD; Pasquale F Innominato^{1,2,4}, MD, PhD; Monique Maurice^{1,2}, PhD; Alexandre Arbaud², Dipl Ing; Jacques Beau², PhD; Mohamed Bouchahda^{2,5}, MD; Ayhan Ulusakarya^{2,5}, MD; Nicolas Beaumatin⁶, MSc; Gabrièle Breda⁶, PhD; Bärbel Finkenstädt^{2,3}, PhD; Francis Lévi^{1,2,5}, MD, PhD

¹Cancer Chronotherapy Team, School of Medicine, University of Warwick, Coventry, United Kingdom

²European Associated Laboratory of the Unité Mixte de Recherche Scientifique 935, Institut National de la Santé et de la Recherche Médicale, Villejuif, France

³Cancer Chronotherapy Team, Department of Statistics, University of Warwick, Coventry, United Kingdom

⁴Department of Oncology, North Wales Cancer Treatment Centre, Bodelwyddan, United Kingdom

⁵Department of Oncology, Paul Brousse Hospital, Assistance Publique-Hôpitaux de Paris, Villejuif, France

⁶Altran Research, Vélizy-Villacoublay, France

*these authors contributed equally

Corresponding Author:

Francis Lévi, MD, PhD
Cancer Chronotherapy Team
School of Medicine
University of Warwick
Gibbet Hill Road
Coventry, CV4 7AL
United Kingdom
Phone: 44 2476575132
Email: F.Levi@warwick.ac.uk

Abstract

Background: Experimental and epidemiologic studies have shown that circadian clocks' disruption can play an important role in the development of cancer and metabolic diseases. The cellular clocks outside the brain are effectively coordinated by the body temperature rhythm. We hypothesized that concurrent measurements of body temperature and rest-activity rhythms would assess circadian clocks coordination in individual patients, thus enabling the integration of biological rhythms into precision medicine.

Objective: The objective was to evaluate the circadian clocks' coordination in healthy subjects and patients through simultaneous measurements of rest-activity and body temperature rhythms.

Methods: Noninvasive real-time measurements of rest-activity and chest temperature rhythms were recorded during the subject's daily life, using a dedicated new mobile electronic health platform (PiCADO). It involved a chest sensor that jointly measured accelerations, 3D orientation, and skin surface temperature every 1-5 min and relayed them out to a mobile gateway via Bluetooth Low Energy. The gateway tele-transmitted all stored data to a server via General Packet Radio Service every 24 hours. The technical capabilities of PiCADO were validated in 55 healthy subjects and 12 cancer patients, whose rhythms were e-monitored during their daily routine for 3-30 days. Spectral analyses enabled to compute rhythm parameters values, with their 90% confidence limits, and their dynamics in each subject.

Results: All the individuals displayed a dominant circadian rhythm in activity with maxima occurring from 12:09 to 20:25. This was not the case for the dominant temperature period, which clustered around 24 hours for 51 out of 67 subjects (76%), and around 12 hours for 13 others (19%). Statistically significant sex- and age-related differences in circadian coordination were identified in the noncancerous subjects, based upon the range of variations in temperature rhythm amplitudes, maxima (acrophases), and phase relations with rest-activity. The circadian acrophase of chest temperature was located at night for the majority of people, but it occurred at daytime for 26% (14/55) of the noncancerous people and 33% (4/12) of the cancer patients, thus supporting

important intersubject differences in circadian coordination. Sex, age, and cancer significantly impacted the circadian coordination of both rhythms, based on their phase relationships.

Conclusions: Complementing rest-activity with chest temperature circadian e-monitoring revealed striking intersubject differences regarding human circadian clocks' coordination and timing during daily routine. To further delineate the clinical importance of such finding, the PiCADO platform is currently applied for both the assessment of health effects resulting from atypical work schedules and the identification of the key determinants of circadian disruption in cancer patients.

(*J Med Internet Res* 2018;20(6):e204) doi:[10.2196/jmir.9779](https://doi.org/10.2196/jmir.9779)

KEYWORDS

circadian clock; eHealth; temperature rhythm; rest-activity rhythm; time series analyses; domomedicine; biomarkers

Introduction

Background

Circadian (about-24-hour) rhythms regulate mammalian physiology, as well as cell metabolism, proliferation, and survival over the 24 hours. These rhythms play an important role in disease processes and treatment effects, which has been largely overlooked in medicine [1-3]. They are generated at single-cell level by molecular clocks, consisting of interwoven feedback loops involving transcription-translation of 15 known specific *clock* genes including *Bmal1*, *Clock*, *Per2*, and *Rev-erb α* [4]. The molecular clocks are coordinated at the whole organism level by the suprachiasmatic nuclei, a hypothalamic pacemaker, which also helps circadian rhythms adjust to light-dark and other environmental 24-hour cycles through the rhythmic control of rest-activity, body temperature, feeding, as well as cortisol and melatonin secretions [1-4]. Thus, both glucocorticoids and body temperature rhythms reset molecular clocks and cellular circadian rhythms in vitro and in vivo [5-8].

Rhythm studies in humans have assumed similar circadian synchronization among subjects, thus inferring the reliability of transverse sampling of different subjects at different time points over 24 hours, and using average values for describing circadian patterns in the group or the population [9]. Treatment effects could also differ according to circadian timing or chronomodulated scheduling of medications in a consistent fashion across individual subjects with a similar circadian entrainment [1,10-12]. Such standardized approaches to chronotherapy proved valid in experimental rodents of same sex, strain, and age, which were synchronized with the same alternation of 12-hours of light and darkness, especially for anticancer drugs [1,13]. This was also true for healthy subjects maintained in human physiology laboratories under controlled environmental conditions [14]. However, little is known regarding circadian rhythms in individual healthy humans or patients during their daily routine.

Intersubject variability in circadian phase has been suggested, based on chronotype questionnaires administered to large populations of presumably healthy subjects [15]. Intersubject differences in both daily timing and 24-hour pattern have also been shown in individual patients collecting up to 5 daily samples of salivary cortisol and/or melatonin determinations for up to 2 days [16-18]. The limitations resulting from such low sampling frequency were overcome through rest-activity monitoring, using a wrist watch accelerometer for a few days

to a few weeks [19-21]. The rest-activity time series led to identify the dichotomy index I<O, the relative amount of activity in bed that was below the median activity out of bed, as an independent predictor of progression-free survival and overall survival among 436 patients with metastatic colorectal cancer [22]. Most importantly, the patients whose I<O was below the median value of 97.5% had a median survival of 11.9 months, as compared with 21.6 months for those with an I<O index over 97.5% [22]. The I<O was also negatively associated with fatigue and appetite loss, and positively with health-related quality of life as assessed by both the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30 EORTC QLQ-C30 and the MD Anderson Symptom Inventory questionnaires in cancer patients with locally advanced or metastatic disease [23,24]. Furthermore, circadian rest-activity disruption, as measured with an I<O of 97.5% or less, in patients receiving cancer chemotherapy could indeed represent an early warning signal of deterioration and emergency hospitalization [25,26]. However, I<O values did not correlate with sex, both in healthy subjects or in cancer patients, or with efficacy of a standardized chronomodulated chemotherapy protocol in cancer patients [19,20,27], whereas the latter profoundly differed between men and women [28]. These clinical data stressed the need for the combination of rest-activity with circadian temperature biomarker to gather more reliable estimates of the circadian phase and to personalize the timing of chronotherapy.

Indeed, despite their consistent and reproducible clinical relevance, the rest-activity time series provide imprecise estimates regarding circadian phase, due to both its square-wave 24-hour pattern and the strong masking effect of the societal routine on the endogenous activity rhythm. Predominant low values of activity suggest prolonged periods of rest during nighttime, whereas frequent high intensity activity occurs during the day, with substantial within-day and day-to-day variability [29]. Nevertheless, a first fixed electronic health (eHealth) internet platform was developed within the inCASA European project—combining telemetric activity monitoring with self-rated symptoms and self-measured body weight. Testing of the platform in 31 cancer patients on chronotherapy at home demonstrated a per protocol compliance rate of ~85% over 1 month, and enabled prediction of emergency hospitalization due to treatment toxicity 3 days in advance [26].

Precise information regarding circadian phase and circadian coordination is also critical for the appropriate timing of treatment delivery to reduce adverse events and/or enhance efficacy [1,10-13,30]. Moreover, both diseases such as cancer

and treatments can disrupt the circadian timing system (CTS) and result in associated symptoms and reduced survival, especially in cancer patients [31]. To address these issues, we have designed an upper chest e-sensor that records and teletransmits activity, temperature and tri-axial orientation. This sensor is integrated into a novel eHealth platform (PiCADO). The circadian rhythms in core and skin surface temperature of men are usually 8-12 hours out of phase, with respective maxima occurring near 16:00 at day time, and near 2:00 at night [32]. The early night drop in core body temperature results from the vasodilatation of the skin vessels and associated rise in skin surface temperature [33]. Such temperature changes are critical for triggering the onset of sleep [34]. The site of temperature measurements for achieving continuous and noninvasive, yet reliable, assessment of human body temperature rhythms in real life has been a challenge over the past decades. The use of a rectal probe has been discouraged as a result of the risk of rectal perforation [35]. Axillary and wrist skin surface temperature records were shown to be largely contaminated by changes in environmental temperature [36-39]. The recent availability of an oral temperature pill has enabled the continuous recording of internal body temperature, yet only for durations that match the gastro-intestinal transit time, ie, ~24-48 hours [40,41]. Previous work by others highlighted the reduced influence of environmental temperature changes on skin surface temperature measurements taken at the upper-anterior chest wall [42,43]. We confirmed these findings through combining infrared technology with continuous recording of patched temperature sensors [44,45]. We further developed dedicated statistical methods to compute dynamic changes in rhythm parameters by combining the inference methods for obtaining interval estimates based on spectral bootstrap with time-varying spectral estimation [46,47].

Objectives

Here, we assessed—for the first time—the performance and relevance of the PiCADO platform for capturing inter- and intrasubject variabilities in the CTS both during daily routine and in real time. We hypothesized that the combination of rest-activity and temperature monitoring would identify large interindividual differences in circadian coordination. The latter would notably support the personalized adaptation of the optimal timing of medications to jointly minimize treatment morbidities and enhance efficacy.

Methods

Study Design

The main objective was to determine whether any inter- and intrasubject differences in human circadian coordination could be captured in real time through remote and noninvasive real-time monitoring during the subjects' usual routine. Such a goal represented a critical step toward the personalization of treatment timing according to individual circadian rhythms, especially for cancer therapies. A new mobile eHealth platform (PiCADO) was designed on purpose. The PiCADO specifications were defined within several multidisciplinary and multiuser focus groups involving nurses, medical oncologists, general practitioners, biomedical and informatics engineers,

socio-anthropologists, and chronobiologists, and through analyzing elderly people's responses in living labs. Three parameters—activity, temperature, and position—are measured using a single CE-marked chest sensor emitter (Movisens, Karlsruhe, Germany) and a pocket-sized CE-marked gateway (Eeleo, Montrouge, France), which could gather further information from other Bluetooth (BT)- and Bluetooth Low Energy (BLE)-connected devices, and send them to a server via the general packet radio service (GPRS) at the required frequency, which may be tuned down as low as every hour, in case of measurements of preset emergency values (Figure 1). This latter function was not activated here. Thus, the PiCADO platform consisted of a chest sensor that measured skin surface temperature every 5 min, the number of accelerations, and the orientation in 3 dimensions every min. All data were teletransmitted via BLE to a pocket-sized gateway, which also could gather data measured by other connected Bluetooth and BLE devices, such as a weight scale. The gateway sent all data to a server every 24 hours. Three cohorts of people were involved, each with different specifications regarding observation span (4 days vs 7-30 days), sensor-carrying method (patch vs dedicated vest or bra), and health condition (healthy vs cancer). Subjects in cohorts 1 and 2 had to be 21 years or older, display no active disease, and not work at night. The study was planned without any intervention. Subjects were advised to remove the sensor for around 20 min once per day to avoid contact with water during showering. The study was conducted according to the Helsinki Declaration [48]. All the subjects enthusiastically volunteered and provided informed consent for carrying and testing the platform system.

Data Management

All anonymous electronic data were transferred from the gateway at home to the server according to the GPRS communication protocol. Once on the server, the data were stored based on HL7 standards (international standards for transfer of clinical and administrative data). Data were downloaded from the server to the computer of our biomedical engineer, who was the only person having access to the server. Data were saved on a secure storage server according to the national Data Protection and Freedom of Information Acts guidance. Data transmission was inspected at least twice per week during the monitoring sessions to insure the proper functioning. Data were retrieved and processed at the end of the monitoring.

Statistical Methods

Treatment of Missing Data and Data Preprocessing

Times at which subjects removed the sensor were identified retrospectively by noting a string of zero Position and Activity counts, jointly with corresponding temperature measurements decreasing toward room temperature values. These intervals were marked as missing values. To perform spectral analysis, the raw activity and temperature data were aggregated over hourly intervals and corresponding median values were computed. The hourly interval length provided a good resolution of the periodogram at frequencies of interest, especially those corresponding to the circadian range. The endpoints of the hourly aggregated data were connected using linear interpolation

in case of recording gaps ≤ 7 hours. For gaps > 7 hours, as it did occur for 7 subjects, the recordings of the whole corresponding 24 hours were ignored for purposes of the analyses.

Estimation of Circadian Parameters

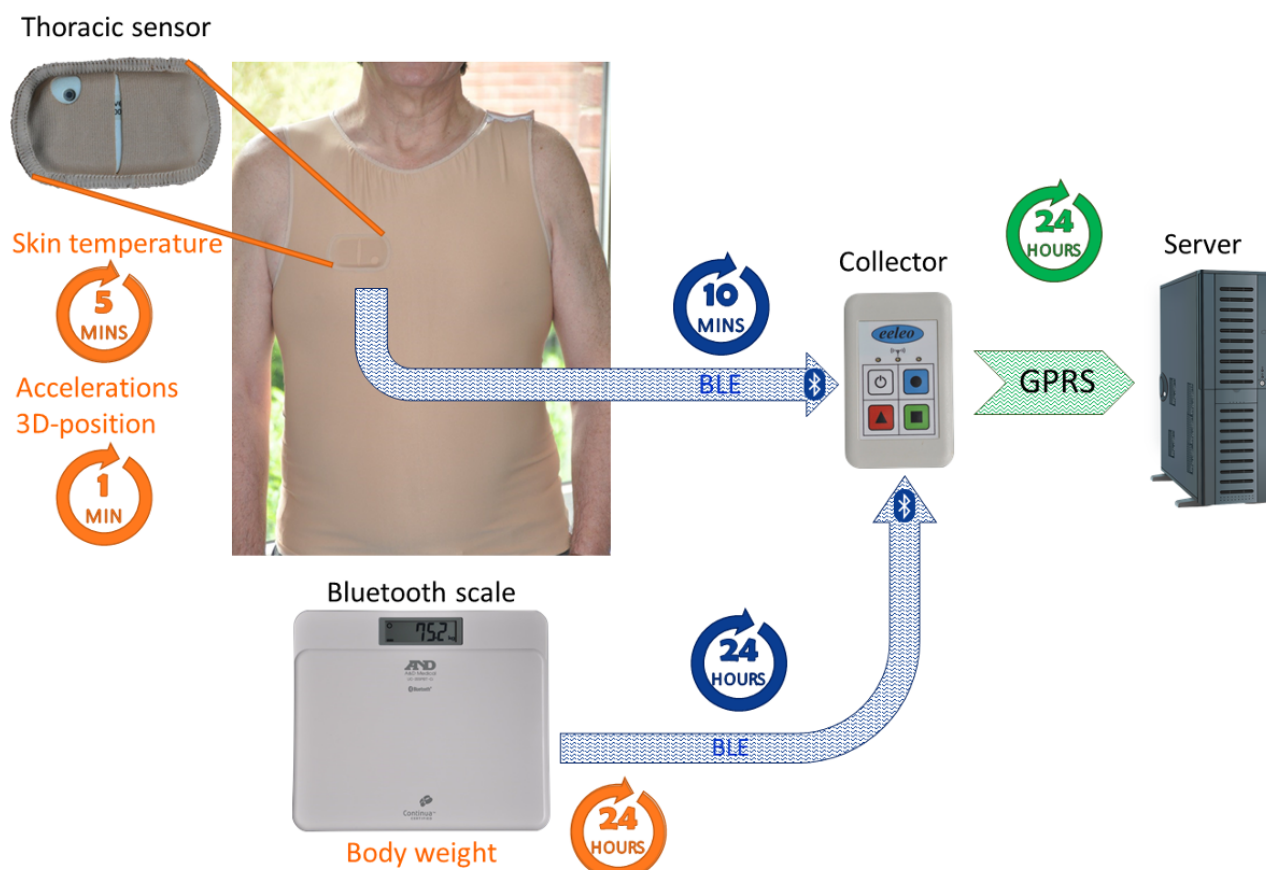
The Spectrum-Resampling (SR) algorithm [46] was applied to estimate the circadian parameters of interest, namely, period, amplitude, and phase of the first and second largest peaks in the spectra for both rest-activity and temperature time series. This method first identified the most important frequencies and their corresponding periods from the estimated spectrum, then fitted a Fourier-type regression model to the data, to obtain the corresponding amplitudes and phases. The SR method provided a bootstrap framework, where all circadian parameters were estimated as the median of the bootstrap samples, and their 90% central confidence intervals were approximated by the corresponding percentiles of the bootstrap samples. To analyze the intrasubject variability over time, the same methodology of spectral analysis was performed over moving windows of 3 days each, with a 1-hour shift per spectral estimate [49]. We note that the window length should be at least 2-3 days, as at least 2 full cycles are needed to estimate the period length.

Regression Analysis

Multivariate regression analysis was applied to test for the effects of covariates such as sex, age, weight, and cohort. As

response variable Y , we considered a selection of estimated circadian parameters that summarized the behavior of the biomarkers for each subject, including (1) the amplitude of the main period of temperature, which was obtained from spectral estimation; (2) the amplitude of activity, during prolonged (usually daily) activity spans, as approximated by the interquartile range (IQR) or amplitude of 50% central values of observed values; (3) the spectral gravity center (or mean period) of temperature; and (4) the spectral gravity center (or mean period) of rest-activity. Possible explanatory variables were sex, age, weight, and amplitude of daily activity as given by the IQR of daily activity counts, cohort, and the following interaction terms: sex*age, sex*weight, and sex*IQR of daily activity. Model selection was performed stepwise (as implemented in R function and based on Akaike's information criterion) [50]. Significance of explanatory variables was tested by t tests of the corresponding coefficients—where a significant effect was concluded for P values smaller than .05. As 0 and 1, respectively, encoded for females and males, a sex-specific effect was computed in a straightforward way in that a sex-specific influence of a covariate was concluded, if the coefficient corresponding to the interaction term sex*covariate was significant in the regression.

Figure 1. eHealth domomedicine platform technology. The chest sensor embedded into the vest shirt with an open area for infrared temperature measurements is shown in the upper left corner. The epoch length of the data points as well as length of time intervals between teletransmission events are indicated within each circle for each variable. Teletransmissions involve Bluetooth Low Energy (BLE) from sensor to gateway and General Packet Radio Service (GPRS) from gateway to server, from which data can be retrieved continuously.



Results

Subject Characteristics and Study Conduct

The PiCADO Domomedicine platform was tested by 69 people. Assessable time series of the 3 variables were obtained for 67 of 69 subjects (97%) over a median duration of 6 days (1st-3rd quartiles, IQ, 4.0 to 12.1), ranging from 3 to 29.7 days. A total of 30 males and 37 females, aged 21 to 83 years, participated in 1 of 3 cohorts of subjects (Figure 2; Table 1).

All subjects were asked to keep their usual daily routines, besides carrying the sensor day and night for the whole monitoring duration, and keeping the gateway within a distance of 2 m. The sensor was initially patched onto the upper left anterior thorax of 28 healthy subjects in Cohort 1, using Tegaderm (10x12 cm, 1626W, 3M^T, Diegem, Belgium) for a median duration of 4 days. To avoid the use of the potentially irritating patches for longer durations, a dedicated vest and bra were designed, which could properly lodge the sensor (Thuasne Medical, Saint Etienne, France). This system was tested by 27 healthy subjects in 2 countries for durations ranging from 3.6 to 28.3 days (Cohort 2). The platform was used by 18 subjects in the United Kingdom for a median duration of 7 days (Cohort 2.1), and by 9 subjects in France, jointly with a BLE weight scale, for a median duration of 19 days (Cohort 2.2; Table 1). To probe the platform for a prolonged use in patients with cancer, the sensor, cloth, and gateway system was further

assessed in 12 patients with advanced or metastatic cancer for a median duration of 18.5 days (Cohort 3, see characteristics in Multimedia Appendix 1).

Inter- and Intrasubject Differences Captured by Time Series Analyses

Time series were preprocessed to account for missing data (see Material and Methods). Using the spectrum resampling method, period, amplitude, and acrophase corresponding to the largest (fundamental) and, if significant, second-largest peak in the spectrum were estimated, along with their respective 90% Confidence Limits, for each subject over the whole time span [46]. Time-varying features of the spectrum were also estimated, through the application of the same method to 3-day windows, which were moved along each time series with 1-hour shifts. The clinically relevant Dichotomy Index I<O was further computed, over consecutive 3-day spans, with 24-hour shifts [26,51]. Strikingly different circadian patterns, rhythm parameters, and I<O values were identified among the 55 healthy subjects (Figures 3 and 4). Although the rest-activity pattern remained consistent from 1 day to the next in the 55 healthy subjects, the temperature pattern varied from day to day in 16 of them, as revealed by changes over time in the dominant periods and the amplitude-acrophase vectors (Figure 4, subjects 3 and 4). Thus, our methodology revealed and quantified day-to-day changes in circadian parameters, thereby enabling the determination of circadian variability both within and between subjects during real-life conditions.

Figure 2. Consort diagram. The 69 subjects were enrolled in 1 of 3 cohorts that differed according to the method of sensor attachment and record duration (first row), health status (second row), and number of subject per group, including sex, age (median and range), and record duration (median and range).

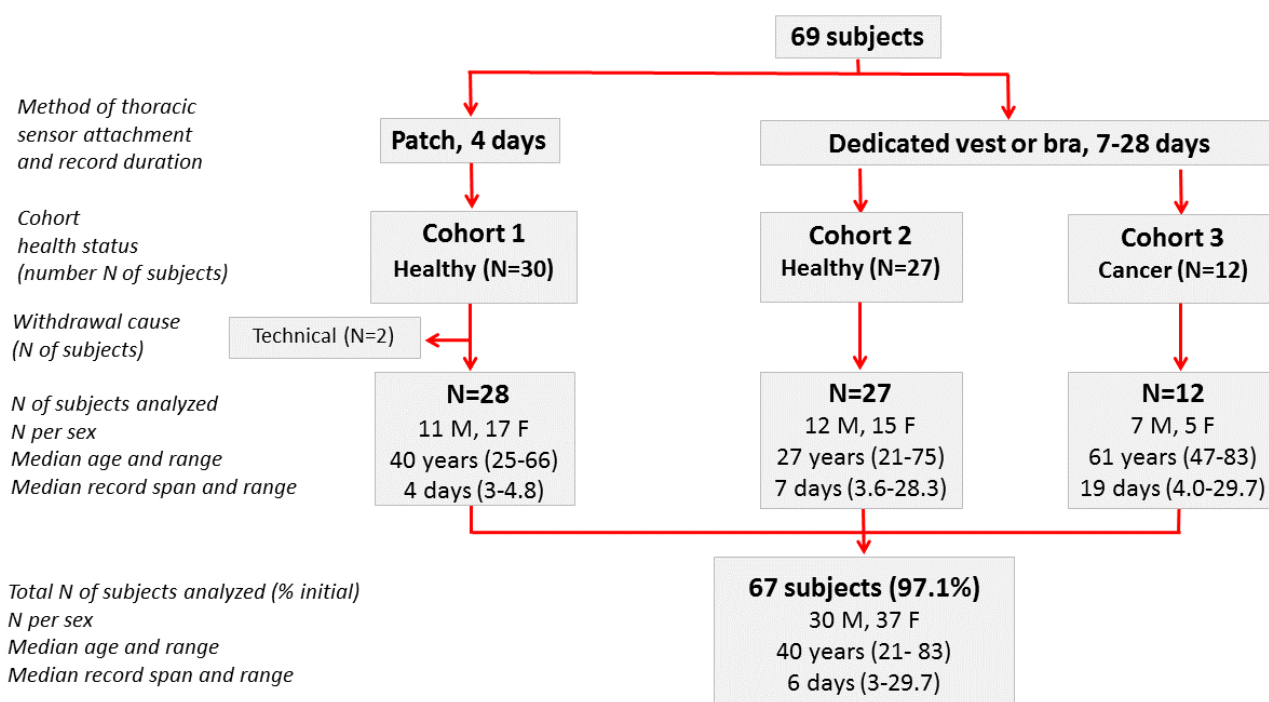


Table 1. Subjects and records characteristics. Median and distribution of data for all quantitative variables. M/F refers to male/female.

Cohort	Total N (M/F)	Age, years (range) [25-75% IQ ^a]	Weight, kg (range) [25-75% IQ]	Height, cm (range) [25-75% IQ]	BMI, kg/m ² (range) [25-75% IQ]	# of subjects with partial record (cause)	Valid time series duration, days (range) [25-75% IQ]	Missing data, days (%) (range) [25-75% IQ]
1	28 (11/17)	40 (25-66) [32.7-49.5]	73 (44-93) [61-77.5]	—	—	0	4.0 (3.0-4.8) [3.9-4.0]	0 (0.8) (0-0.3) [0-0.1]
2.1	18 (8/10)	26 (21-75) [24-32.3]	73 (46-93) [66.7-78]	172 (156-183) [168.5-175.5]	24 (19-28) [22.2-26.1]	1 (charger dysfunction)	7.0 (3.6-12.4) [7.0-7.4]	0.1 (1.5) (0-0.6) [0.05-0.15]
2.2	9 (4/5)	34 (25-57) [27-38]	71 (54-83) [65-72]	169 (155-195) [162-171]	24 (18-31) [23.3-26]	3 (subject-related) ^b 1 (unsticking of electronic circuit)	19.0 (4.9-28.3) [17.0-21.7]	2.3 (7.7) (0.3-7.7) [1.1-2.6]
2	27 (12/15)	27 (21-75) [24.5-37.5]	70 (46-93) [65-77]	171 (155-195) [165-174]	24.3 (17.9-31.2) [22.3-26.2]	3 subject-related 2 technical failures	7.4 (3.6-28.3) [7.0-15.3]	0.1 (2.0) (0-7.7) [0.1-1.0]
3	12 (7/5)	61 (47-83) [54-66.5]	66 (45-80) [60-72]	170 (152-185) [164-176]	22.8 (18.5-26.1) [21.4-23.9]	5 (subject-related) ^c	20.3 (4.7-29.7) [16.5-27.2]	1.2 (5.0) (0-20.4) [0.7-2.7]

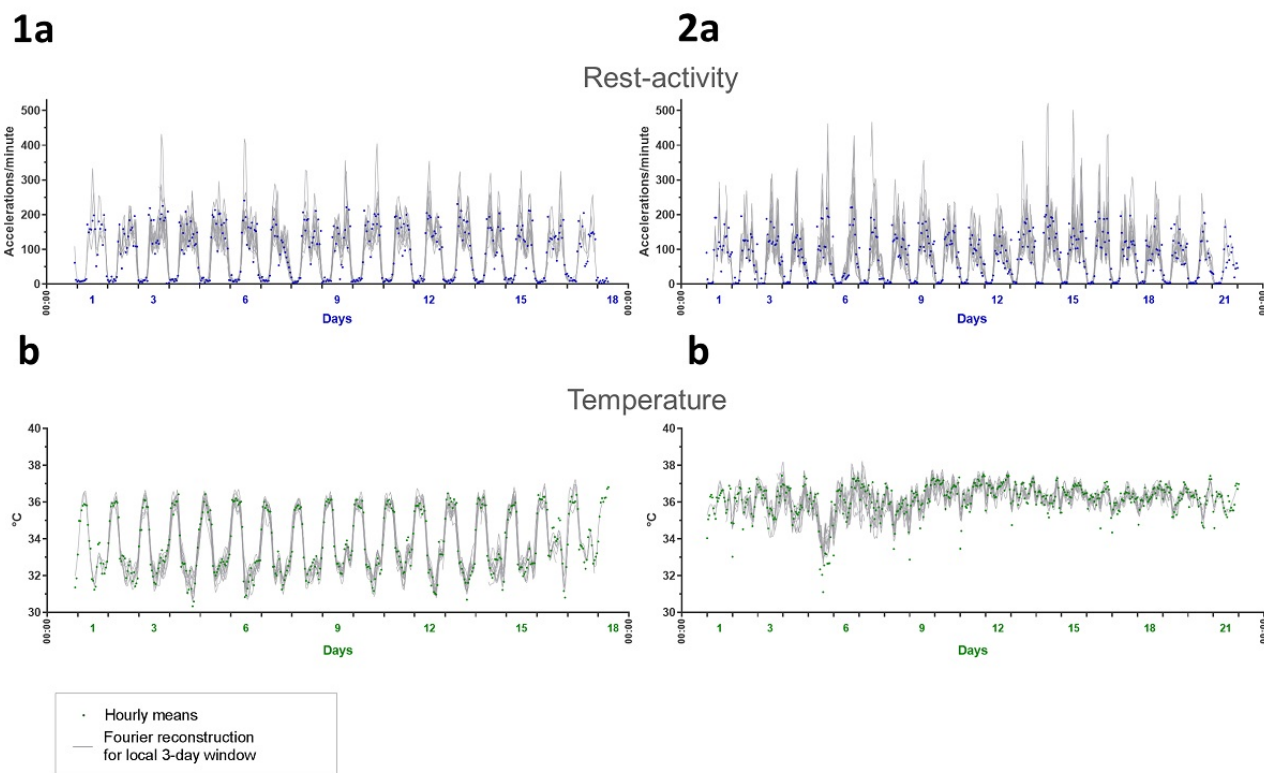
^aIQ: interquartile.^bForgetfulness after charging (N=1), travel abroad starting before end of recording span (N=1), wrong charging procedure applied (N=1).^cWrong charging procedure applied (N=2); poor tolerability of adjusted sensor-dedicated cloth due to no current use of bra (N=1), or treatment-related itching (N=1); need for more feedback and support (N=1).**Figure 3.** Inter- and intrasubject variability in circadian patterns illustrated by chronograms of rest-activity (a) and thoracic skin surface temperature (b) of 2 healthy subjects. Panel 1 (left): time series from a 57-year-old female researcher, with usual respective times of awakening and retiring at 8:30 and 22:30; mean rest-activity I<O of 99.7%, with intrasubject coefficient of variation of 0.2%. Panel 2 (right): time series from a 27-year-old female student, usually awakening at 8:40 and retiring at 01:00; mean I<O of 99.2%, with a coefficient of variation of 1.2%.

Figure 4. Inter- and intrasubject variability in circadian acrophases and amplitudes of rest-activity (left panels), temperature (middle panels), and circadian coordination (right panels). Illustrative examples through polar plots in 4 healthy subjects, whose 3-day time series shifted by 6 hours (subjects 1 and 4) or 1 hour (subjects 2 and 3), have been analyzed using the sampling-resampling spectrum analysis. The length of each vector represents the amplitude of the dominant period and its direction points toward the timing of the corresponding acrophase.

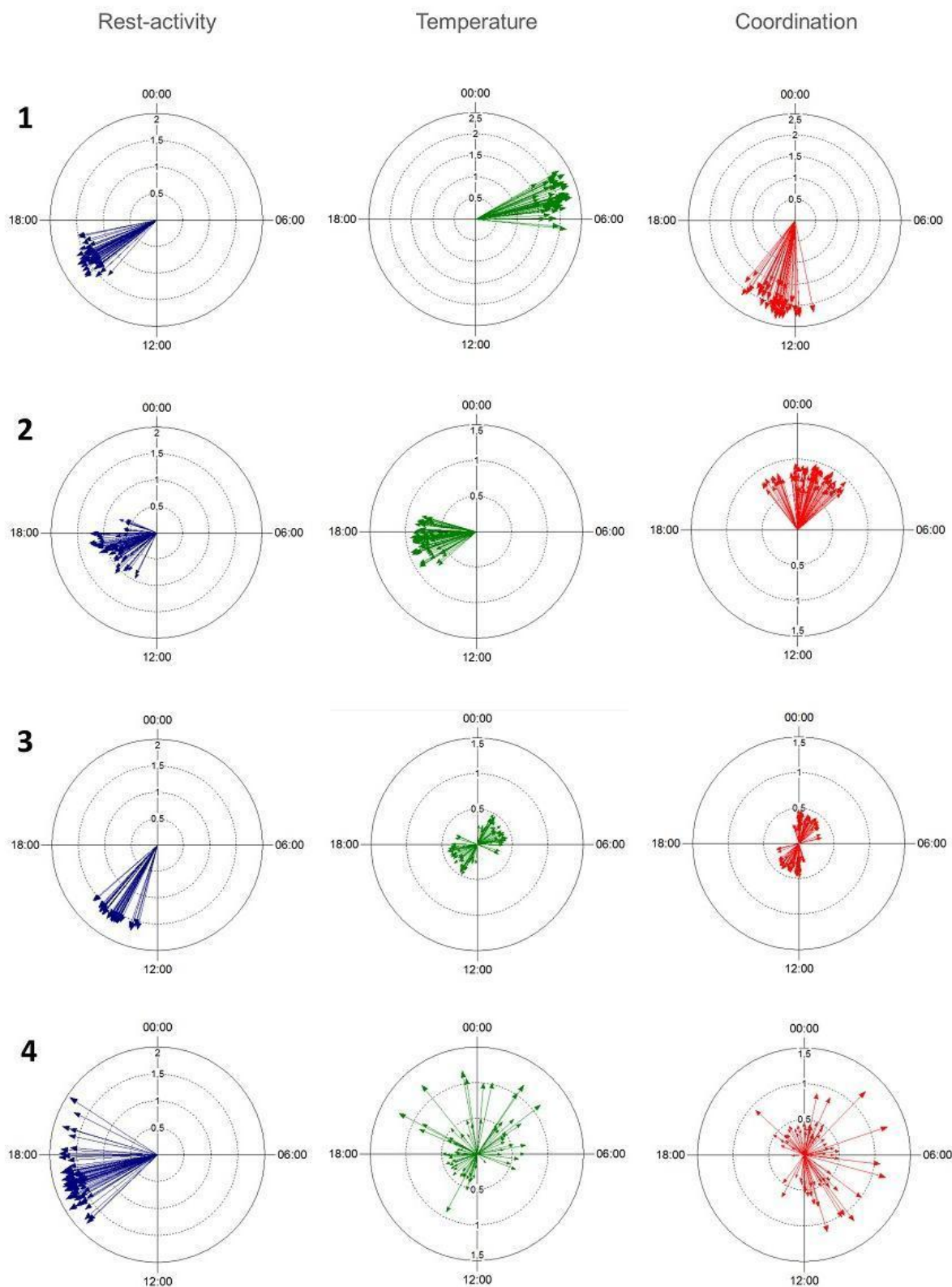
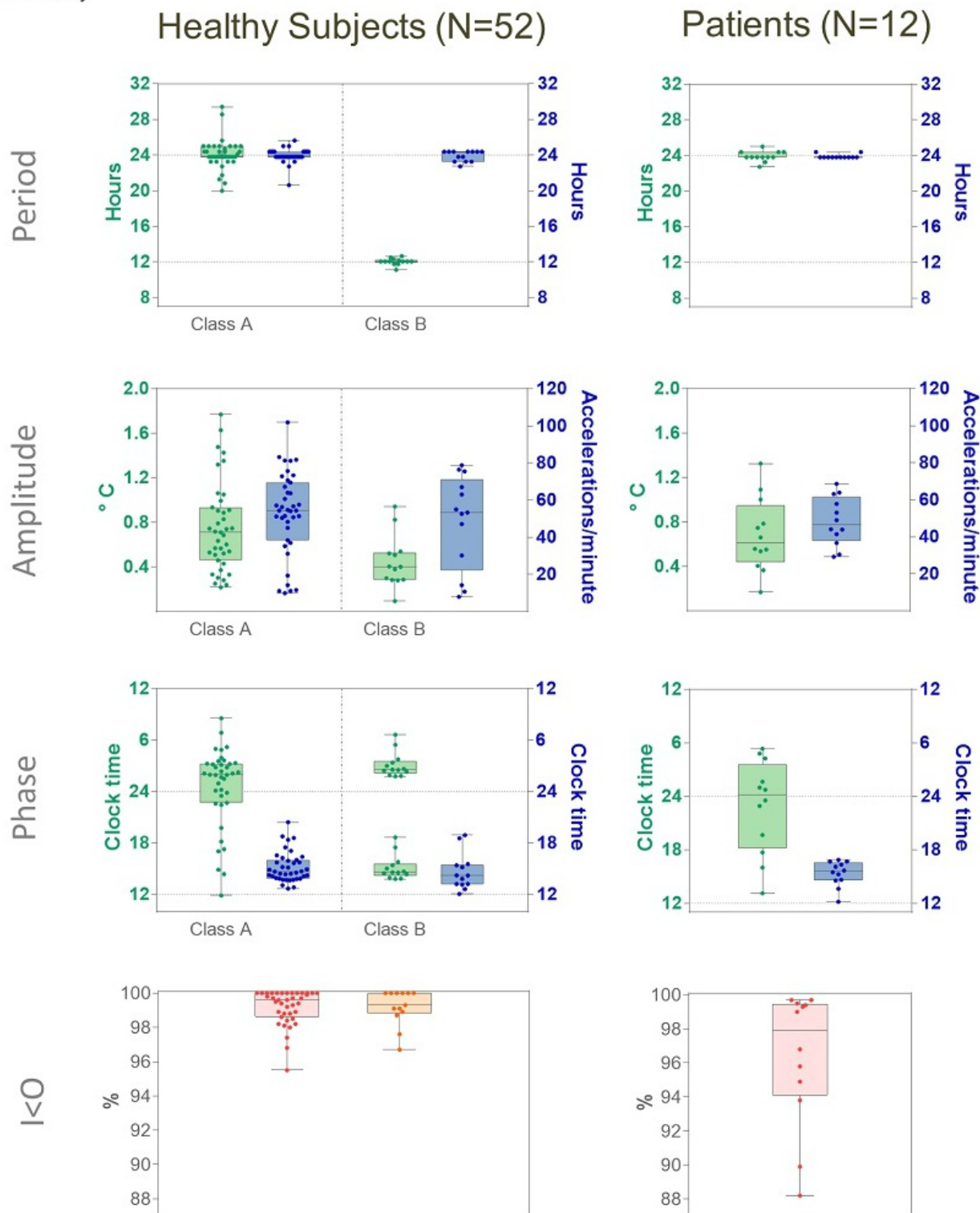


Figure 5. Intersubject variabilities in main rhythm parameters of healthy subjects (left columns) and cancer patients (right column). Median, interquartiles, range, and individual values of dominant periods and corresponding amplitudes and acrophases of temperature (green) and rest-activity (blue), based on spectral analysis of time series over the whole time span. The bottom row depicts the distribution of the dichotomy index $I < 0$ of the rest-activity pattern in Class A or B healthy subjects, as well as in cancer patients.

Temperature

Activity



Subjects Classification According to Temperature Periods and Circadian Coordination

Results of the spectral analyses over the complete time series of the 28 subjects in Cohort 1 were very similar to those of the 27 subjects in Cohort 2 (see [Multimedia Appendices 2 and 3](#)). For instance, a dominant circadian rhythm in rest-activity was identified for all subjects, yet with a dominant circadian rhythm in skin surface temperature for 68% (19/28) of the subjects in Cohort 1 and 74% (20/27) of those in Cohort 2; the dominant temperature period was about 12 hour (circa-hemidian) for 25% (7/28) and 22% (6/27) of the subjects in Cohorts 1 and 2, respectively. Atypical patterns were found for 2 and 1 subjects, respectively.

Both cohorts of healthy subjects were pooled, so as to further examine the relations between summary statistics of the temperature and activity variables and their spectral properties, as well as the relevance of available covariates, such as sex, age, and weight in the 55 healthy subjects. The differences between the acrophases of both rhythms on each tested timespan were taken as estimates of circadian coordination.

Large intersubject differences characterized the median values of both activity, whose median was 39.8 movements per min

[IQ, 19.3 to 54.7], and skin surface temperature, whose median was 35.2°C [34.7 to 35.8]. The rest-activity rhythm displayed a dominant 24-hour period for all 55 subjects. In contrast, the skin surface temperature displayed a dominant period in the circadian range for 39/55 subjects (71%), and in the circahemidian range for 13/55 subjects (27%). Three out of 55 subjects (5%) displayed atypical patterns, one for both variables, due to trans meridian travel; one with a dominant temperature period of 6 hours; and one with an unstable temperature pattern. We categorized the subjects as belonging to *Class A* (circadian rhythms in both variables), *Class B* (circadian activity and circahemidian temperature) or *Class C* (atypical rest-activity and/or temperature patterns, not shown; [Figure 5](#)). For *Class A*, the median circadian acrophases were located at 14:40 for activity (with individual acrophases ranging over 7 hours 44 min, from 12:41 to 20:25), and at 3:33 at night for skin surface temperature, yet with intersubject differences spread over 24 hours. Thus, the skin surface temperature acrophase occurred at night (22:01 to 7:00) for 31 subjects and during day-time (7:01 to 22:00) for 8 subjects in *Class A* ($P<.001$ from exact Fisher test). As a result, the median time interval between both rhythm acrophases was 11 hours 47 min, yet it had a wide range from 7 min to 12 hours among the 39 *Class A* subjects.

Figure 6. Sex and sex-age dependencies of circadian amplitude (upper row) and spectrum gravity center of temperature time series (lower row) in 55 healthy subjects. (a) Box plot of temperature amplitudes of the estimated main harmonic for females (left, N=32) and males (right, N=23); (b) Sex-specific effect of age on temperature amplitude shown by estimated regression line with 95% confidence bands. The vertical dashes along the horizontal axes show corresponding age of each subject. (c) Box plot of the estimated gravity center of temperature spectra for females (left, N=32) and males (right, N=23). (d) Sex-specific effect of age on the gravity center of temperature spectra shown by estimated regression line with 95% confidence bands.

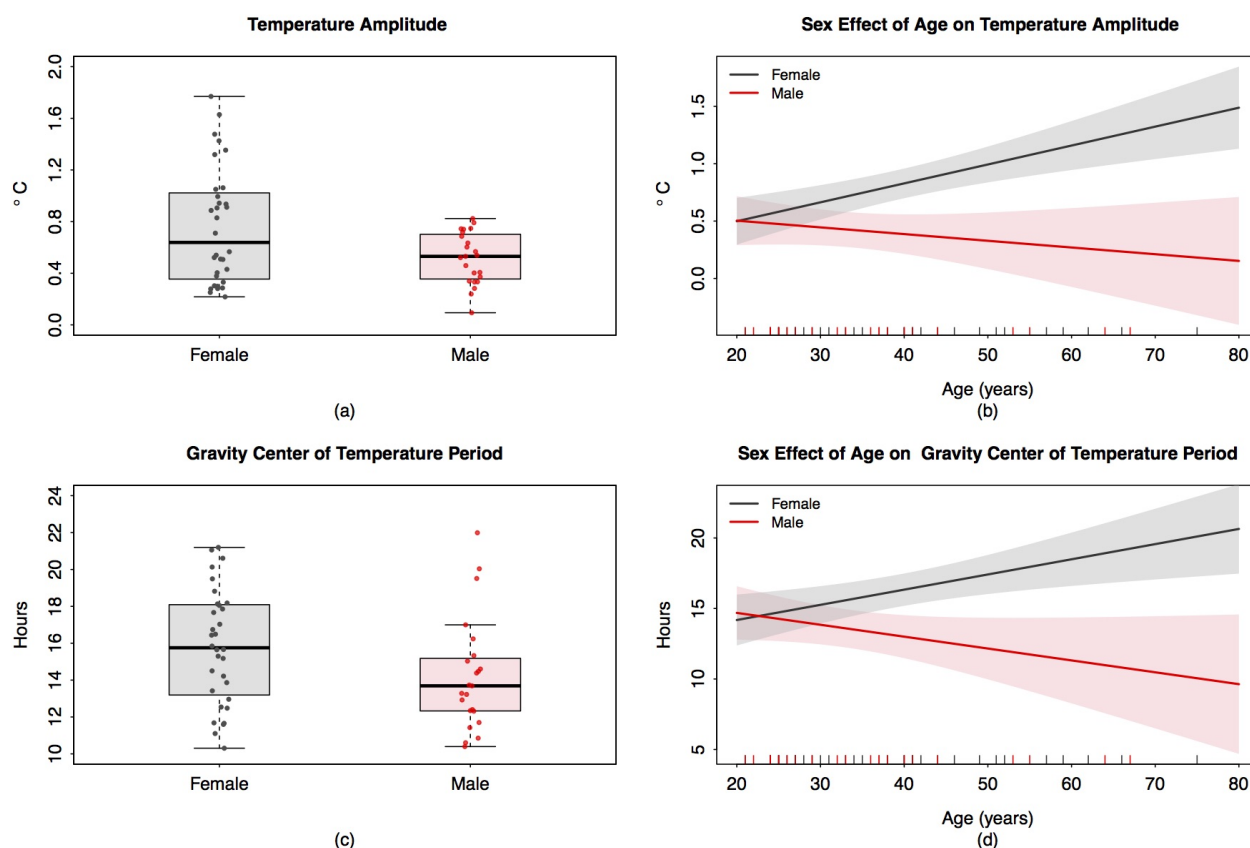
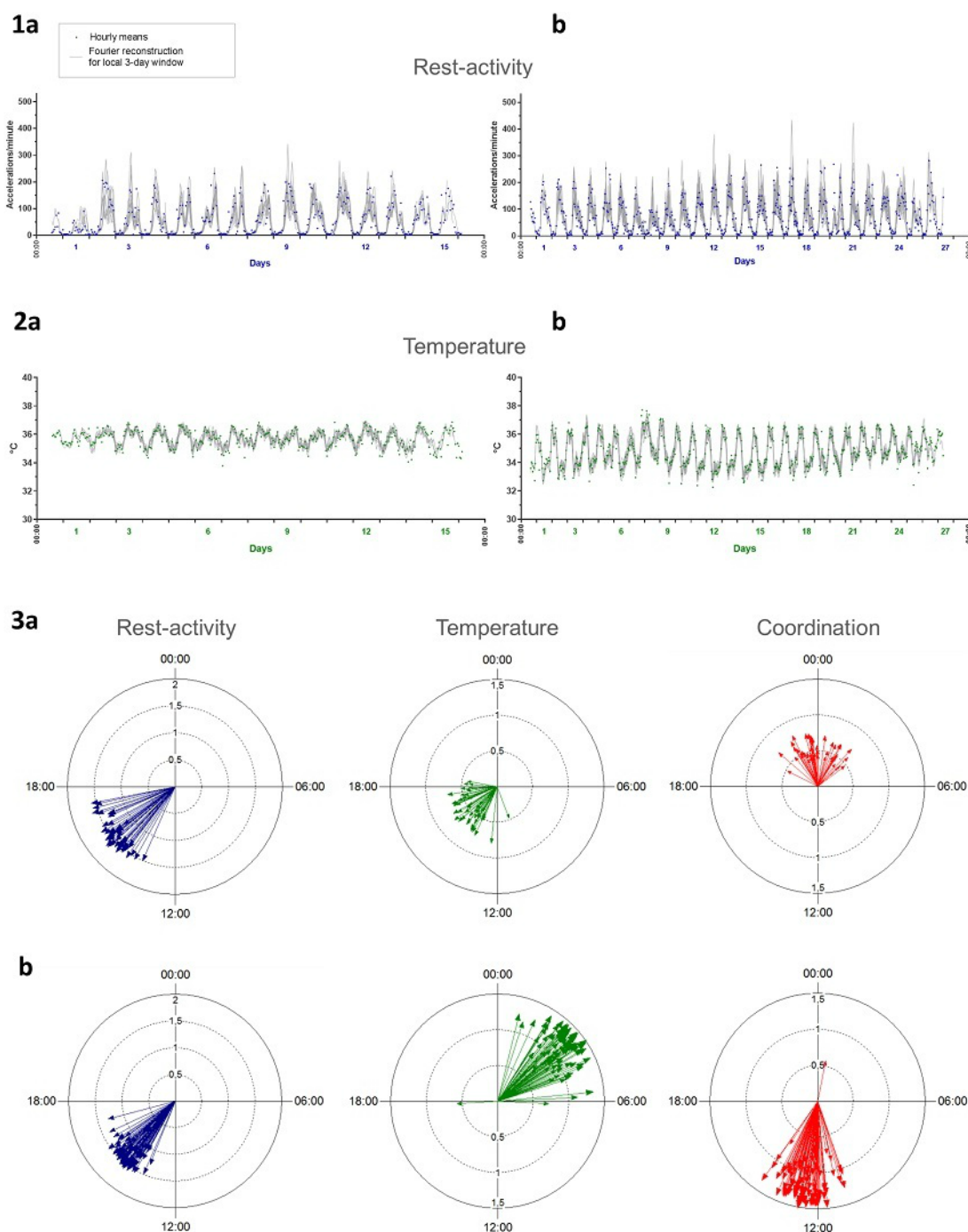


Figure 7. Inter- and intrasubject differences in circadian patterns in rest-activity and chest surface temperature as illustrated in 2 cancer patients (a and b), with chronograms (panels 1 and 2) and polar plot representations of amplitude-acrophase vectors (panel 3).



Sex and Age Moderation of Temperature but Not Activity Rhythms

Statistically significant pairwise Spearman correlations with $|r| \geq 0.4$ were identified for median values and their corresponding dominant rhythm amplitudes for both activity ($r=.77$) and temperature ($r=.49$). Sex jointly with age had a selective influence on the rhythmic organization of temperature, but not that of rest-activity in the 55 healthy subjects. This was statistically validated by multivariate regression analysis of amplitude and spectral gravity center (or “mean period”) of temperature, as response variables, with age and sex as

covariates (Figure 6). Females displayed larger temperature amplitude than males (two-sample t test, $P=.005$; Figure 6). Females also had larger values than males in the estimated gravity center of their temperature spectrum (two-sample t test, $P=.03$). Overall, the results indicated that females tended to mostly display dominant 24-hour rhythm periods with large amplitudes, whereas 12-hour rhythms tended to predominate in males (Figure 6). The interaction between sex and both temperature rhythm parameters was mostly apparent beyond 35 years. Regression analysis showed that sex, age, and the interaction term sex*age had a significant effect on temperature amplitude (with P values=.04, <.001 and .002, respectively)

and on the gravity center of the temperature spectra (with P values of .03, .004, and .003, respectively; [Figure 6](#)).

Circadian Coordination in Cancer Patients

A total of 7 male and 5 female patients tested the platform for a median of 19 days, ranging from 4 to 29 days. All the patients had previously received 1 or more chemotherapy protocols for metastatic gastrointestinal cancer, including colorectal ($N=5$), pancreatic ($N=3$), liver ($N=2$), stomach ($N=1$), or anal ($N=1$) cancer. Moreover, 5 patients had co-morbidities (see [Multimedia Appendix 1](#)). Six patients received chronomodulated infusions, including 2 courses at home, while being telemonitored. Circadian patterns in rest-activity and thoracic temperature were identified in all the patients' records, as shown by dominant 24-hour periods. [Figure 7](#) depicts examples of circadian patterns in 2 cancer patients, a 53-year-old male (patient 1) on conventional chemotherapy with oxaliplatin-5-fluorouracil-leucovorin delivered during hospitalization for liver metastases from stomach cancer, on days 1 and 2 of the recording, and a 78-year-old male (patient 2) with lung and lymph node metastases from rectal cancer, whose treatment involved intravenous panitumumab and chronomodulated chemotherapy with irinotecan, oxaliplatin, 5-fluorouracil, and leucovorin at home, using a 4-channel programmable-in-time pump, from day 6 to day 9 of the recording. Usual awaking and retiring times were 7:50 and 21:50, respectively, for patient 1 and 7:15 and 23:00, respectively, for patient 2. Despite all patients belonging to Class A, large interpatient variations were found regarding the circadian amplitudes in both rest-activity (range 29.3 to 68.6 accelerations per minute) and temperature (0.17°C to 1.33°C). Individual rest-activity acrophases varied within a 5-hour range, occurring between 12:09 and 16:52. In contrast, the circadian acrophases of temperature were located between 22:01 and 07:00 at night for 8 patients, and at daytime for another 4 patients. Intersubject variations were also revealed with dichotomy index $I<O$, whose individual median values ranged from 88.2% to 99.7%, with rather low estimated intrasubject coefficients of variations of 0.3% to 6.4%. Interestingly, the circadian acrophase of temperature was located at daytime for 3 of 6 patients with an $I<O$ under 97.5%, but at nighttime for 5 of 6 patients, whose $I<O$ was over 97.5% ($P<.001$ from exact Fisher test).

Relations Between $I<O$ and Chest Temperature Timing Among the Study Population

An $I<O$ of 97.5% or less was found in 9 of the 51 subjects, with a 24-hour temperature rhythm (18% of Class A), and in 1 of the 13 subjects, with a 12-hour temperature pattern (8% of Class B). We did not consider further the 3 Class C subjects (1 with a high $I<O$, and 2 with a low one). Interestingly, the average circadian acrophase of chest temperature was localized at daytime, ie, between 7:01 and 22:00 for 6 of the 54 subjects with a high $I<O$ value (11%), as compared with 6 of the 10 subjects whose $I<O$ was low (60%); $P<.001$ from exact Fisher test). Indeed, having a high $I<O$ enabled the accurate prediction of a chest temperature occurring at night, whereas no reliable prediction of temperature rhythm timing could be made in the subjects whose $I<O$ was low (see [Multimedia Appendix 4](#)).

Discussion

Effective Mobile eHealth Platform

A mobile eHealth platform (PiCADO) consisting of a BLE chest sensor and a pocket-sized BT/BLE/GPRS gateway effectively measured body accelerations, 3D orientation, and upper chest temperature every 1-5 min for prolonged timespans, teletransmitting them to a central server every 24 hours. This interval can be reduced to 1 hour. In 67 healthy people or cancer patients, this technology revealed large interindividual differences in circadian coordination that was captured in real time during their daily routine. The multidimensional platform and its capabilities to combine electronic patient-reported outcomes (PROs) with circadian rhythm monitoring and body weight measures meet the expectations of the American Society for Clinical Oncology, regarding the future of eHealth technologies in cancer medicine [52]. Thus, the weekly transmission of electronic PROs significantly improved overall survival in 2 randomized trials conducted in cancer patients, indicating this information might elicit initially unplanned interventions and/or modify patient's engagement with an oncologic benefit equivalent to that of the addition of an active drug [53,54]. The PiCADO platform provides a novel framework for the further integration of circadian rhythms and other parameters into proactive timely care interventions and the ready assessment of their efficacy. Such technology enables a novel systems approach for a coordinated medical and care logistics fit for the management of chronic disease and cancer patients, so called Domomedicine [26,55].

Technology and Compliance

The reliability and acceptance of the PiCADO platform is illustrated from the records obtained in the subjects from 2 countries, including 55 healthy ones and 12 cancer patients at home or during their usual activities, for prespecified durations of 4 up to 30 days. Indeed, valid time series were extracted from the server, and amenable to longitudinal statistical analyses. These results are in line with those obtained previously in 31 cancer patients through a fixed internet platform within the inCASA European project [26]. The current PiCADO platform was developed to bypass the limitations of the inCASA platform, through mobile technology, and multirhythm monitoring, including skin temperature. Its specifications aimed at the broad integration of circadian rhythms into medicine, as a potential new information for triggering progress in the proactive management of cancer and chronic diseases.

Further real-life tests (not shown here) indicated the reliability of combining such rhythm telemonitoring with other physiologic or patient-reported outcomes parameters, such as body weight or self-rated symptoms, through additional BLE-connected devices.

Circadian Coordination of Healthy People or Cancer Patients During Their Daily Routine

This study has highlighted the consistency of the 24-hour patterns in both rest-activity and thoracic skin surface temperature from one day to the next both in healthy and cancerous patients. However, marked intersubject differences

were found regarding the dominant period, the spectrum central gravity, and the amplitude and acrophase location of temperature. More specifically, nearly 1/4 of the healthy subjects had a prominent 12-hour rather than 24-hour periodic temperature rhythm. Sex and age, but not weight, were influential on the temperature rhythms, as discussed below. Among the 39 healthy subjects with a predominantly circadian chest temperature rhythm, the acrophase was located at night, as expected [33,56,57], for 79% (31/39), or at daytime for 20% (8/39). Indeed, it is known that skin surface temperature usually increases at night, thus resulting in heat dissipation and the core body temperature drop that has been associated to sleep triggering [57]. In contrast, intersubject variability was limited for rest-activity, with all the 67 subjects displaying a dominant circadian rhythm, and 97% (65/67) of them having an acrophase of activity located in the afternoon or early evening. Consistently, only 5 of 55 healthy subjects had a dichotomy index $I < O$ between 97.5% and 95.5%, a rate in good agreement with a prior study using wrist actigraphy in 182 young subjects [27]. Half of the 12 cancer patients had an $I < O$ ranging between 97.5% and 88.7%, ie, the same rate as that previously reported using wrist actigraphy in 436 cancer patients [22,27]. This suggests that there is consistency of both recording methods regarding $I < O$ estimation. There were statistically significant relations between temperature period and amplitude on the one hand, and sex, age, and $I < O$ on the other hand; however, no such correlations were found for the circadian rest-activity parameters derived from the spectrum. Moreover, the circadian acrophase of thoracic temperature occurred at nighttime, thus indicating physiologic circadian coordination, for 48 of the 54 subjects (89%) whose $I < O$ exceeded 97.5%, whereas it was located at daytime for 6 of the 10 subjects (60%) with a lower $I < O$. These findings suggest that low $I < O$ values could reflect nonphysiologic circadian coordination. Hence, temperature monitoring significantly increased the information already provided by rest-activity regarding CTS function, as suspected in pilot studies [44,45].

Sex and Age as Important Determinants of Circadian Coordination and Timing

Here, women aged >35 years tended to have robust 24-hour rhythms in their temperature, with larger amplitudes as compared with those in males. This apparent difference between circadian rhythms in temperature and activity has a neuroanatomic basis, which has been demonstrated in rodents [58]. Thus, the rest-activity and temperature rhythms were shown to be generated by the neurons located in the caudal [59] and subparaventricular zone of the suprachiasmatic nuclei [60,61], respectively. As body temperature is an important driver of both the resetting and the coordination of peripheral clocks [8], its rhythm could play a critical role for the circadian timing of medications [13]. The larger amplitude of the circadian rhythm in temperature shown here for women supports the hypothesis that circadian timing of medications is even more critical in women as compared with men. Moreover, larger circadian amplitudes have also been found in females, compared with males, for cortisol secretion, another key driver of peripheral clocks and clock-controlled metabolism and cell-cycle determinants of drug effects [5,62]. Two separate clinical trials

uncovered that the optimal timing of a multidrug chronomodulated chemotherapy protocol could lag 6 hours behind in women as compared with men with the same cancer type [63,64]. Such differences in circadian amplitudes and time lag between males and females were similar in cancer patients and in mouse models [65,66]. The sex-related differences in CTS that were discovered here could explain why the same multidrug cancer chronotherapy protocol improved response rate, progression-free survival, and overall survival in men, but not in women with metastatic colorectal cancer, independently of all other prognostic factors within a meta-analysis of 3 randomized international clinical trials involving 842 patients [28].

Study Limitations and Current Perspectives

The main limitations of our study beyond the technological possibilities of the eHealth platform was its exploratory nature, with unbalanced sex and age distribution, as well as a limited number of cancer patients. We did not assess chronotype via an established questionnaire [67], or working hours, although the temperature and activity records provided an objective and quantitative phase assessment [68,69]. Other indicators of circadian function, such as cortisol or melatonin rhythms or Dim Light Melatonin Onset [70], were not determined, nor did we investigate clock genes polymorphisms [71]. Indeed, genetic variants of human clock genes have been shown to be associated with phenotypic differences, which could allegedly impact disease processes, including cancer [72]. Moreover, the various biomarker rhythms of circadian function can be differentially altered by disease processes [73].

Comparison With Prior Work

However, the technology supporting the platform here presented allows to fully integrate the new information brought about by circadian rhythms, jointly with repeated self-assessed symptoms and other measurable parameters. The PiCADO platform represents an answer to the limitations of the current hospital-centered care system, which was designed for responding to acute medical events, rather than for managing the long-term medical care required for cancer and other chronic diseases [74-76]. The latter illnesses not only represent the vast bulk of health care payload in Western countries, but also their management is suboptimal, given the current hospital logistics constraints [77,78]. This is especially true, as disease and treatment response dynamics vary from patient to patient, with most events occurring at home and remaining unnoticed, while impacting daily life, and eventually cumulating and leading to emergency hospitalizations. In the previous inCASA study, we showed the reliability of a fixed internet platform for the remote monitoring of self-rated symptoms, body weight, and rest-activity rhythm in cancer patients, and the rather good patient compliance. Moreover, such remote monitoring provided early warning signals of patient deterioration at home that resulted in emergency hospitalization 3 days later. Circadian disruption played a prominent role in the determination of such early warning signaling [26]. Here, we aimed at improving circadian timing disorders detection through combining rest-activity and temperature monitoring, while being also able to integrate self-rated symptoms and body weight measurements,

using a mobile, rather than fixed platform. We now believe that such PiCADO eHealth platform could shift the current hospital-centered system of care toward a patient-centered system promoting biomedical progress. This would involve the safe delivery of care and support treatments at home or during the patient's daily activities, through adjusting cancer therapies to circadian rhythms, ie, chronotherapy [26,79-81]. PiCADO further provides an ongoing monitoring system of the patient's well-being through developing forecasting analytical methodology integrating multiple sources (patient-reported outcomes, circadian rhythms in activity and temperature, physiological measures).

Conclusions

We have shown that such a mobile eHealth circadian platform allows automatic and noninvasive monitoring of precise circadian parameters in nonhospitalized healthy subjects and

cancer patients. Preliminary analyses point to large and unsuspected intersubject variabilities, which may be of great importance when administering treatments and preventing emergency situations. Hence, it deserves further testing as a tool for the real-time assessment of the CTS of humans of both sexes in various conditions, where 2 types of studies are currently in progress or about to start, respectively, in France and in the United Kingdom. The platform is being used for the determination of the impact of the occupational schedule on circadian function in the CIRCADIEM study, whereas the service rendered by this circadian eHealth platform is to be investigated further in cancer patients in the IDEAs study. Clearly, such system now needs assessment over months and in large patient cohorts within a prospective clinical trial. eHealth circadian monitoring may indeed be of preventive and curative interest in a number of situations involving chronic conditions [82].

Acknowledgments

The authors warmly thank all the subjects who enthusiastically participated in this study, both in France and in the United Kingdom, and provided such reliable information and data. The authors are grateful to Thierry Didi (Eeleo, Montrouge, France) who developed the PiCADO platform according to the authors' specifications, and ensured it was working well all along the tests; Matthieu Ducottet (Thuasne Médical), who designed the dedicated vests and bras for lodging the chest sensor; Aly Chkeir and Jacques Duchene (UTT); and Alexandre Duclos (MADOPA, Expert Center in Technologies and Services for Maintaining Elderly Persons Autonomy at Home, 255 rue de Vaugirard, 75015—Paris, France). The authors are indebted to David Spiegel (Stanford), Eva Schernhammer (Harvard and Vienna), and David Rand (Warwick) for their critical review and constructive comments on an earlier version of this manuscript.

This study was supported by the Conseil Régional d'Ile de France, the Conseil Régional de Champagne-Ardenne, Mairie de Paris and the Banque Publique d'Investissement (BPI France) through the Fonds Unique Interministériel 12 (PiCADO, Contract 11017951), the Institut de Recherche en Santé Publique from France (CLOCK-DOM1, grant 2014-BDCR-EC), the European Union through the Coordinating Action on Systems Medicine (Grant 305033, FP7), and the Medical Research Council, London, United Kingdom (Grant MR/M013170).

Authors' Contributions

SK, PFI, AA, MB, AU, FV, and FL designed the study, recruited the subjects, and followed data acquisition. SK, QH, and BF performed the statistical analyses and iconography. MM, JB, NB, GB, and FL conceived the platform, coordinated the technology development, and performed preliminary descriptive analyses. BF and FL coordinated the manuscript development and codirected the work. SK, QH, BF, and FL wrote the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

JB and FL have cofounded ETIC-Systems, a start-up aiming at further developing eHealth-connected devices for human health. NB and GB coordinated the PiCADO project on behalf of Altran Research, an international engineering company. The patent entitled "Procédé de détermination automatique de l'index de Dichotomie I<O d'un individu, Brevet français n° 1562557 du 16-déc-2015, N/Réf: UTT/16/F" involves shares of Université Technologique de Troyes, INSERM, and Altran.

Multimedia Appendix 1

Patient characteristics.

[PDF File (Adobe PDF File), 190KB - [jmir_v20i6e204_app1.pdf](#)]

Multimedia Appendix 2

Intersubject variabilities in main rhythm parameters of healthy subjects in Cohort 1.

[PDF File (Adobe PDF File), 330KB - [jmir_v20i6e204_app2.pdf](#)]

Multimedia Appendix 3

Intersubject variabilities in main rhythm parameters of healthy subjects in Cohort 2.

[[PDF File \(Adobe PDF File\), 321KB - jmir_v20i6e204_app3.pdf](#)]

Multimedia Appendix 4

Polar plots of mean temperature acrophases in study population according to dichotomy index I<O value.

[[PDF File \(Adobe PDF File\), 355KB - jmir_v20i6e204_app4.pdf](#)]

References

1. Levi F, Schibler U. Circadian rhythms: mechanisms and therapeutic implications. *Annu Rev Pharmacol Toxicol* 2007;47:593-628. [doi: [10.1146/annurev.pharmtox.47.120505.105208](#)] [Medline: [17209800](#)]
2. Bass J, Lazar MA. Circadian time signatures of fitness and disease. *Science* 2016 Nov 25;354(6315):994-999. [doi: [10.1126/science.aah4965](#)] [Medline: [27885004](#)]
3. Panda S. Circadian physiology of metabolism. *Science* 2016 Nov 25;354(6315):1008-1015. [doi: [10.1126/science.aah4967](#)] [Medline: [27885007](#)]
4. Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet* 2017 Dec;18(3):164-179 [FREE Full text] [doi: [10.1038/nrg.2016.150](#)] [Medline: [27990019](#)]
5. Balsalobre A, Damiola F, Schibler U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* 1998 Jun 12;93(6):929-937 [FREE Full text] [Medline: [9635423](#)]
6. Balsalobre A, Brown SA, Marcacci L, Tronche F, Kellendonk C, Reichardt HM, et al. Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science* 2000 Sep 29;289(5488):2344-2347 [FREE Full text] [Medline: [11009419](#)]
7. Brown SA, Zimbrun G, Fleury-Olela F, Preitner N, Schibler U. Rhythms of mammalian body temperature can sustain peripheral circadian clocks. *Curr Biol* 2002 Sep 17;12(18):1574-1583 [FREE Full text] [Medline: [12372249](#)]
8. Buhr ED, Yoo S, Takahashi JS. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science* 2010 Oct 15;330(6002):379-385 [FREE Full text] [doi: [10.1126/science.1195262](#)] [Medline: [20947768](#)]
9. Nelson W, Cornélissen G, Hinkley D, Bingham C, Halberg F. Construction of rhythm-specified reference intervals and regions, with emphasis on 'hybrid' data, illustrated for plasma cortisol. *Chronobiologia* 1983;10(2):179-193. [Medline: [6872677](#)]
10. Levi F, Le Louarn C, Reinberg A. Timing optimizes sustained-release indomethacin treatment of osteoarthritis. *Clin Pharmacol Ther* 1985 Jan;37(1):77-84. [Medline: [3880688](#)]
11. Lévi F, Zidani R, Misset JL. Randomised multicentre trial of chronotherapy with oxaliplatin, fluorouracil, and folinic acid in metastatic colorectal cancer. *International Organization for Cancer Chronotherapy. Lancet* 1997 Sep 06;350(9079):681-686. [Medline: [9291901](#)]
12. Buttgerit F, Doering G, Schaeffler A, Witte S, Sierakowski S, Gromnica-Ihle E, et al. Efficacy of modified-release versus standard prednisone to reduce duration of morning stiffness of the joints in rheumatoid arthritis (CAPRA-1): a double-blind, randomised controlled trial. *Lancet* 2008 Jan 19;371(9608):205-214. [doi: [10.1016/S0140-6736\(08\)60132-4](#)] [Medline: [18207016](#)]
13. Lévi F, Okyar A, Dulong S, Innominato PF, Clairambault J. Circadian timing in cancer treatments. *Annu Rev Pharmacol Toxicol* 2010;50:377-421. [doi: [10.1146/annurev.pharmtox.48.113006.094626](#)] [Medline: [20055686](#)]
14. Boivin DB, James FO, Wu A, Cho-Park PF, Xiong H, Sun ZS. Circadian clock genes oscillate in human peripheral blood mononuclear cells. *Blood* 2003 Dec 01;102(12):4143-4145 [FREE Full text] [doi: [10.1182/blood-2003-03-0779](#)] [Medline: [12893774](#)]
15. Fischer D, Lombardi DA, Marucci-Wellman H, Roenneberg T. Chronotypes in the US - influence of age and sex. *PLoS One* 2017;12(6):e0178782 [FREE Full text] [doi: [10.1371/journal.pone.0178782](#)] [Medline: [28636610](#)]
16. Segerstrom SC, Boggero IA, Smith GT, Sephton SE. Variability and reliability of diurnal cortisol in younger and older adults: implications for design decisions. *Psychoneuroendocrinology* 2014 Nov;49:299-309 [FREE Full text] [doi: [10.1016/j.psyneuen.2014.07.022](#)] [Medline: [25137484](#)]
17. Mormon MC, Bogdan A, Cormont S, Touitou Y, Lévi F. Cortisol diurnal variation in blood and saliva of patients with metastatic colorectal cancer: relevance for clinical outcome. *Anticancer Res* 2002;22(2B):1243-1249. [Medline: [12168933](#)]
18. Sephton SE, Sapolsky RM, Kraemer HC, Spiegel D. Diurnal cortisol rhythm as a predictor of breast cancer survival. *J Natl Cancer Inst* 2000 Jun 21;92(12):994-1000. [Medline: [10861311](#)]
19. Mormont MC, Waterhouse J, Bleuzen P, Giacchetti S, Jami A, Bogdan A, et al. Marked 24-h rest/activity rhythms are associated with better quality of life, better response, and longer survival in patients with metastatic colorectal cancer and good performance status. *Clin Cancer Res* 2000 Aug;6(8):3038-3045 [FREE Full text] [Medline: [10955782](#)]
20. Innominato PF, Focan C, Goria T, Moreau T, Garufi C, Waterhouse J, Chronotherapy Group of the European Organization for Research Treatment of Cancer. Circadian rhythm in rest and activity: a biological correlate of quality of life and a predictor

- of survival in patients with metastatic colorectal cancer. *Cancer Res* 2009 Jun 01;69(11):4700-4707 [FREE Full text] [doi: [10.1158/0008-5472.CAN-08-4747](https://doi.org/10.1158/0008-5472.CAN-08-4747)] [Medline: [19470769](https://pubmed.ncbi.nlm.nih.gov/19470769/)]
21. Ancoli-Israel S, Liu L, Rissling M, Natarajan L, Neikrug AB, Palmer BW, et al. Sleep, fatigue, depression, and circadian activity rhythms in women with breast cancer before and after treatment: a 1-year longitudinal study. *Support Care Cancer* 2014 Sep;22(9):2535-2545 [FREE Full text] [doi: [10.1007/s00520-014-2204-5](https://doi.org/10.1007/s00520-014-2204-5)] [Medline: [24733634](https://pubmed.ncbi.nlm.nih.gov/24733634/)]
 22. Lévi F, Dugué PA, Innominato P, Karaboué A, Dispersyn G, Parganiha A, ARTBC Chronotherapy Group. Wrist actimetry circadian rhythm as a robust predictor of colorectal cancer patients survival. *Chronobiol Int* 2014 Oct;31(8):891-900. [doi: [10.3109/07420528.2014.924523](https://doi.org/10.3109/07420528.2014.924523)] [Medline: [24927369](https://pubmed.ncbi.nlm.nih.gov/24927369/)]
 23. Mormont MC, Waterhouse J. Contribution of the rest-activity circadian rhythm to quality of life in cancer patients. *Chronobiol Int* 2002 Jan;19(1):313-323. [Medline: [11962684](https://pubmed.ncbi.nlm.nih.gov/11962684/)]
 24. Innominato PF, Palesh O, Bjarnason GA, Ulusakarya A, Spiegel D, Levi F. Symptoms associated with circadian rest-activity rhythm disruption in 237 patients with metastatic colorectal cancer. *JCO* 2015 Oct 10;33(29_suppl):1. [doi: [10.1200/jco.2015.33.29_suppl.1](https://doi.org/10.1200/jco.2015.33.29_suppl.1)]
 25. Ortiz-Tudela E, Iurisci I, Beau J, Karaboue A, Moreau T, Rol MA, et al. The circadian rest-activity rhythm, a potential safety pharmacology endpoint of cancer chemotherapy. *Int J Cancer* 2014 Jun 01;134(11):2717-2725 [FREE Full text] [doi: [10.1002/ijc.28587](https://doi.org/10.1002/ijc.28587)] [Medline: [24510611](https://pubmed.ncbi.nlm.nih.gov/24510611/)]
 26. Innominato PF, Komarzynski S, Mohammad-Djafari A, Arbaud A, Ulusakarya A, Bouchahda M, et al. Clinical relevance of the first domomedicine platform securing multidrug chronotherapy delivery in metastatic cancer patients at home: the inCASA European project. *J Med Internet Res* 2016 Nov 25;18(11):e305 [FREE Full text] [doi: [10.2196/jmir.6303](https://doi.org/10.2196/jmir.6303)] [Medline: [27888171](https://pubmed.ncbi.nlm.nih.gov/27888171/)]
 27. Natale V, Innominato PF, Boreggiani M, Tonetti L, Filardi M, Parganiha A, et al. The difference between in bed and out of bed activity as a behavioral marker of cancer patients: a comparative actigraphic study. *Chronobiol Int* 2015;32(7):925-933. [doi: [10.3109/07420528.2015.1053909](https://doi.org/10.3109/07420528.2015.1053909)] [Medline: [26158757](https://pubmed.ncbi.nlm.nih.gov/26158757/)]
 28. Giacchetti S, Dugué PA, Innominato PF, Bjarnason GA, Focan C, Garufi C, ARTBC International Chronotherapy Group. Sex moderates circadian chemotherapy effects on survival of patients with metastatic colorectal cancer: a meta-analysis. *Ann Oncol* 2012 Dec;23(12):3110-3116. [doi: [10.1093/annonc/mds148](https://doi.org/10.1093/annonc/mds148)] [Medline: [22745214](https://pubmed.ncbi.nlm.nih.gov/22745214/)]
 29. Xiao L, Huang L, Schrack JA, Ferrucci L, Zipunnikov V, Crainiceanu CM. Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics* 2015 Apr;16(2):352-367 [FREE Full text] [doi: [10.1093/biostatistics/kxu045](https://doi.org/10.1093/biostatistics/kxu045)] [Medline: [25361695](https://pubmed.ncbi.nlm.nih.gov/25361695/)]
 30. Dallmann R, Okyar A, Lévi F. Dosing-time makes the poison: circadian regulation and pharmacotherapy. *Trends Mol Med* 2016 May;22(5):430-445. [doi: [10.1016/j.molmed.2016.03.004](https://doi.org/10.1016/j.molmed.2016.03.004)] [Medline: [27066876](https://pubmed.ncbi.nlm.nih.gov/27066876/)]
 31. Innominato PF, Roche VP, Palesh OG, Ulusakarya A, Spiegel D, Lévi FA. The circadian timing system in clinical oncology. *Ann Med* 2014 Jun;46(4):191-207. [doi: [10.3109/07853890.2014.916990](https://doi.org/10.3109/07853890.2014.916990)] [Medline: [24915535](https://pubmed.ncbi.nlm.nih.gov/24915535/)]
 32. Kräuchi K, Wirz-Justice A. Circadian rhythm of heat production, heart rate, and skin and core temperature under unmasking conditions in men. *Am J Physiol* 1994 Sep;267(3):R819-R829. [doi: [10.1152/ajpregu.1994.267.3.R819](https://doi.org/10.1152/ajpregu.1994.267.3.R819)] [Medline: [8092328](https://pubmed.ncbi.nlm.nih.gov/8092328/)]
 33. Kräuchi K. How is the circadian rhythm of core body temperature regulated? *Clin Auton Res* 2002 Jun;12(3):147-149. [Medline: [12269545](https://pubmed.ncbi.nlm.nih.gov/12269545/)]
 34. Van Someren EJ. Mechanisms and functions of coupling between sleep and temperature rhythms. *Prog Brain Res* 2006;153:309-324. [Medline: [16876583](https://pubmed.ncbi.nlm.nih.gov/16876583/)]
 35. Reinberg A. Circadian changes in the temperature of human beings. *Bibl Radiol* 1975(6):128-139. [Medline: [1101884](https://pubmed.ncbi.nlm.nih.gov/1101884/)]
 36. Motohashi Y, Reinberg A, Levi F, Nougier J, Benoit O, Foret J, et al. Axillary temperature: a circadian marker rhythm for shift workers. *Ergonomics* 1987 Sep;30(9):1235-1247. [doi: [10.1080/00140138708966019](https://doi.org/10.1080/00140138708966019)] [Medline: [3428253](https://pubmed.ncbi.nlm.nih.gov/3428253/)]
 37. Ortiz-Tudela E, Martinez-Nicolas A, Campos M, Rol MA, Madrid JA. A new integrated variable based on thermometry, actimetry and body position (TAP) to evaluate circadian system status in humans. *PLoS Comput Biol* 2010 Nov 11;6(11):e1000996 [FREE Full text] [doi: [10.1371/journal.pcbi.1000996](https://doi.org/10.1371/journal.pcbi.1000996)] [Medline: [21085644](https://pubmed.ncbi.nlm.nih.gov/21085644/)]
 38. Martinez-Nicolas A, Ortiz-Tudela E, Rol MA, Madrid JA. Uncovering different masking factors on wrist skin temperature rhythm in free-living subjects. *PLoS One* 2013;8(4):e61142 [FREE Full text] [doi: [10.1371/journal.pone.0061142](https://doi.org/10.1371/journal.pone.0061142)] [Medline: [23577201](https://pubmed.ncbi.nlm.nih.gov/23577201/)]
 39. Longato E, Garrido M, Saccardo D, Montesinos Guevara C, Mani AR, Bolognesi M, et al. Expected accuracy of proximal and distal temperature estimated by wireless sensors, in relation to their number and position on the skin. *PLoS One* 2017;12(6):e0180315 [FREE Full text] [doi: [10.1371/journal.pone.0180315](https://doi.org/10.1371/journal.pone.0180315)] [Medline: [28666029](https://pubmed.ncbi.nlm.nih.gov/28666029/)]
 40. Monnard CR, Fares EJ, Calonne J, Miles-Chan JL, Montani JP, Durrer D, et al. Issues in continuous 24-h core body temperature monitoring in humans using an ingestible capsule telemetric sensor. *Front Endocrinol (Lausanne)* 2017 Jun 13;8:130 [FREE Full text] [doi: [10.3389/fendo.2017.00130](https://doi.org/10.3389/fendo.2017.00130)] [Medline: [28659868](https://pubmed.ncbi.nlm.nih.gov/28659868/)]
 41. Bongers CC, Daanen HA, Bogerd CP, Hopman MT, Eijsvogels TM. Validity, reliability, and inertia of four different temperature capsule systems. *Med Sci Sports Exerc* 2018 Jan;50(1):169-175. [doi: [10.1249/MSS.0000000000001403](https://doi.org/10.1249/MSS.0000000000001403)] [Medline: [28816921](https://pubmed.ncbi.nlm.nih.gov/28816921/)]
 42. Bogh M, Minors DS, Waterhouse JM. Can insulated skin temperature act as a substitute for rectal temperature when studying circadian rhythms? *Chronobiol Int* 1994 Oct;11(5):332-339. [Medline: [7828216](https://pubmed.ncbi.nlm.nih.gov/7828216/)]

43. Kim S, Lee JY. Skin sites to predict deep-body temperature while wearing firefighters' personal protective equipment during periodical changes in air temperature. *Ergonomics* 2016 Apr;59(4):496-503. [doi: [10.1080/00140139.2015.1075604](https://doi.org/10.1080/00140139.2015.1075604)] [Medline: [26214379](https://pubmed.ncbi.nlm.nih.gov/26214379/)]
44. Scully CG, Karaboué A, Liu WM, Meyer J, Innominato PF, Chon KH, et al. Skin surface temperature rhythms as potential circadian biomarkers for personalized chronotherapeutics in cancer patients. *Interface Focus* 2011 Feb 06;1:48-60 [FREE Full text] [doi: [10.1098/rsfs.2010.0012](https://doi.org/10.1098/rsfs.2010.0012)] [Medline: [21544221](https://pubmed.ncbi.nlm.nih.gov/21544221/)]
45. Roche VP, Mohamad-Djafari A, Innominato PF, Karaboué A, Gorbach A, Lévi FA. Thoracic surface temperature rhythms as circadian biomarkers for cancer chronotherapy. *Chronobiol Int* 2014 Apr;31(3):409-420 [FREE Full text] [doi: [10.3109/07420528.2013.864301](https://doi.org/10.3109/07420528.2013.864301)] [Medline: [24397341](https://pubmed.ncbi.nlm.nih.gov/24397341/)]
46. Costa MJ, Finkenstädt B, Roche V, Lévi F, Gould PD, Foreman J, et al. Inference on periodicity of circadian time series. *Biostatistics* 2013 Sep;14(4):792-806 [FREE Full text] [doi: [10.1093/biostatistics/kxt020](https://doi.org/10.1093/biostatistics/kxt020)] [Medline: [23743206](https://pubmed.ncbi.nlm.nih.gov/23743206/)]
47. Dahlhaus R. Local inference for locally stationary time series based on the empirical spectral measure. *J Econom* 2009 Aug;151(2):101-112. [doi: [10.1016/j.jeconom.2009.03.002](https://doi.org/10.1016/j.jeconom.2009.03.002)]
48. Carlson RV, Boyd KM, Webb DJ. The revision of the Declaration of Helsinki: past, present and future. *Br J Clin Pharmacol* 2004 Jun;57(6):695-713 [FREE Full text] [doi: [10.1111/j.1365-2125.2004.02103.x](https://doi.org/10.1111/j.1365-2125.2004.02103.x)] [Medline: [15151515](https://pubmed.ncbi.nlm.nih.gov/15151515/)]
49. Paraschakis K, Dahlhaus R. Frequency and phase estimation in time series with quasi periodic components. *J Time Ser Anal* 2012;33(1):13-31. [doi: [10.1111/j.1467-9892.2011.00736.x](https://doi.org/10.1111/j.1467-9892.2011.00736.x)]
50. Seber GA, Lee AJ. *Linear Regression Analysis*. US: John Wiley & Sons; Jan 23, 2012.
51. Minors D, Akerstedt T, Atkinson G, Dahlitz M, Folkard S, Levi F, et al. The difference between activity when in bed and out of bed. I. Healthy subjects and selected patients. *Chronobiol Int* 1996 May;13(1):27-34. [Medline: [8761934](https://pubmed.ncbi.nlm.nih.gov/8761934/)]
52. Fisch MJ, Chung AE, Accordino MK. Using technology to improve cancer care: social media, wearables, and electronic health records. *Am Soc Clin Oncol Educ Book* 2016;35:200-208 [FREE Full text] [doi: [10.14694/EDBK_156682](https://doi.org/10.14694/EDBK_156682)] [Medline: [27249700](https://pubmed.ncbi.nlm.nih.gov/27249700/)]
53. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *J Am Med Assoc* 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: [10.1001/jama.2017.7156](https://doi.org/10.1001/jama.2017.7156)] [Medline: [28586821](https://pubmed.ncbi.nlm.nih.gov/28586821/)]
54. Denis F, Yossi S, Septans AL, Charron A, Voog E, Dupuis O, et al. Improving survival in patients treated for a lung cancer using self-evaluated symptoms reported through a web application. *Am J Clin Oncol* 2017 Oct;40(5):464-469. [doi: [10.1097/COC.000000000000189](https://doi.org/10.1097/COC.000000000000189)] [Medline: [25811297](https://pubmed.ncbi.nlm.nih.gov/25811297/)]
55. Lévi F, Saguez C. *Le patient, Les Technologies et la Médecine Ambulatoire*. France: Académie des technologies; Apr 21, 2008.
56. Waterhouse J, Fukuda Y, Morita T. Daily rhythms of the sleep-wake cycle. *J Physiol Anthropol* 2012 Mar 13;31:5 [FREE Full text] [doi: [10.1186/1880-6805-31-5](https://doi.org/10.1186/1880-6805-31-5)] [Medline: [22738268](https://pubmed.ncbi.nlm.nih.gov/22738268/)]
57. Kräuchi K. The human sleep-wake cycle reconsidered from a thermoregulatory point of view. *Physiol Behav* 2007 Feb 28;90(2-3):236-245. [doi: [10.1016/j.physbeh.2006.09.005](https://doi.org/10.1016/j.physbeh.2006.09.005)] [Medline: [17049364](https://pubmed.ncbi.nlm.nih.gov/17049364/)]
58. Moore RY. The suprachiasmatic nucleus and the circadian timing system. *Prog Mol Biol Transl Sci* 2013;119:1-28. [doi: [10.1016/B978-0-12-396971-2.00001-4](https://doi.org/10.1016/B978-0-12-396971-2.00001-4)] [Medline: [23899592](https://pubmed.ncbi.nlm.nih.gov/23899592/)]
59. Abrahamson EE, Moore RY. Lesions of suprachiasmatic nucleus efferents selectively affect rest-activity rhythm. *Mol Cell Endocrinol* 2006 Jun 27;252(1-2):46-56. [doi: [10.1016/j.mce.2006.03.036](https://doi.org/10.1016/j.mce.2006.03.036)] [Medline: [16793198](https://pubmed.ncbi.nlm.nih.gov/16793198/)]
60. Moore RY, Danchenko RL. Paraventricular-subparaventricular hypothalamic lesions selectively affect circadian function. *Chronobiol Int* 2002 Mar;19(2):345-360. [Medline: [12025929](https://pubmed.ncbi.nlm.nih.gov/12025929/)]
61. Saper CB, Lu J, Chou TC, Gooley J. The hypothalamic integrator for circadian rhythms. *Trends Neurosci* 2005 Mar;28(3):152-157. [doi: [10.1016/j.tins.2004.12.009](https://doi.org/10.1016/j.tins.2004.12.009)] [Medline: [15749169](https://pubmed.ncbi.nlm.nih.gov/15749169/)]
62. Gunn PJ, Middleton B, Davies SK, Revell VL, Skene DJ. Sex differences in the circadian profiles of melatonin and cortisol in plasma and urine matrices under constant routine conditions. *Chronobiol Int* 2016;33(1):39-50 [FREE Full text] [doi: [10.3109/07420528.2015.1112396](https://doi.org/10.3109/07420528.2015.1112396)] [Medline: [26731571](https://pubmed.ncbi.nlm.nih.gov/26731571/)]
63. Lévi F, Focan C, Karaboué A, de la Valette V, Focan-Henrard D, Baron B, et al. Implications of circadian clocks for the rhythmic delivery of cancer therapeutics. *Adv Drug Deliv Rev* 2007 Aug 31;59(9-10):1015-1035. [doi: [10.1016/j.addr.2006.11.001](https://doi.org/10.1016/j.addr.2006.11.001)] [Medline: [17692427](https://pubmed.ncbi.nlm.nih.gov/17692427/)]
64. Levi F, Garufi C, Karaboué A, Focan C, Chollet P, Li XM, et al. Sex-related differences in circadian-dependent tolerance of Irinotecan (I) added to chronomodulated (chrono) 5-Fluorouracil (F), Leucovorin (L) and Oxaliplatin (O): Final results from international randomised time-finding study in patients with metastatic colorectal cancer (MCC). *Ann Oncol* 2017 Sep 18;28(suppl_5):522P. [doi: [10.1093/annonc/mdx393.048](https://doi.org/10.1093/annonc/mdx393.048)]
65. Ahowesso C, Li XM, Zampera S, Peteri-Brunbäck B, Dulong S, Beau J, et al. Sex and dosing-time dependencies in irinotecan-induced circadian disruption. *Chronobiol Int* 2011 May;28(5):458-470. [doi: [10.3109/07420528.2011.569043](https://doi.org/10.3109/07420528.2011.569043)] [Medline: [21721861](https://pubmed.ncbi.nlm.nih.gov/21721861/)]
66. Li XM, Mohammad-Djafari A, Dumitru M, Dulong S, Filipinski E, Siffroi-Fernandez S, et al. A circadian clock transcription model for the personalization of cancer chronotherapy. *Cancer Res* 2013 Dec 15;73(24):7176-7188 [FREE Full text] [doi: [10.1158/0008-5472.CAN-13-1528](https://doi.org/10.1158/0008-5472.CAN-13-1528)] [Medline: [24154875](https://pubmed.ncbi.nlm.nih.gov/24154875/)]

67. Roenneberg T. Having trouble typing? What on earth is chronotype? *J Biol Rhythms* 2015 Dec;30(6):487-491. [doi: [10.1177/0748730415603835](https://doi.org/10.1177/0748730415603835)] [Medline: [26446872](https://pubmed.ncbi.nlm.nih.gov/26446872/)]
68. Lack L, Bailey M, Lovato N, Wright H. Chronotype differences in circadian rhythms of temperature, melatonin, and sleepiness as measured in a modified constant routine protocol. *Nat Sci Sleep* 2009;1:1-8 [FREE Full text] [Medline: [23616692](https://pubmed.ncbi.nlm.nih.gov/23616692/)]
69. Adan A, Archer SN, Hidalgo MP, Di Milia L, Natale V, Randler C. Circadian typology: a comprehensive review. *Chronobiol Int* 2012 Nov;29(9):1153-1175. [doi: [10.3109/07420528.2012.719971](https://doi.org/10.3109/07420528.2012.719971)] [Medline: [23004349](https://pubmed.ncbi.nlm.nih.gov/23004349/)]
70. Klerman EB, Gershengorn HB, Duffy JF, Kronauer RE. Comparisons of the variability of three markers of the human circadian pacemaker. *J Biol Rhythms* 2002 Apr;17(2):181-193. [doi: [10.1177/0748730002129002474](https://doi.org/10.1177/0748730002129002474)] [Medline: [12002165](https://pubmed.ncbi.nlm.nih.gov/12002165/)]
71. Jagannath A, Taylor L, Wakaf Z, Vasudevan SR, Foster RG. The genetics of circadian rhythms, sleep and health. *Hum Mol Genet* 2017 Oct 01;26(R2):R128-R138. [doi: [10.1093/hmg/ddx240](https://doi.org/10.1093/hmg/ddx240)] [Medline: [28977444](https://pubmed.ncbi.nlm.nih.gov/28977444/)]
72. Benna C, Helfrich-Förster C, Rajendran S, Monticelli H, Pilati P, Nitti D, et al. Genetic variation of clock genes and cancer risk: a field synopsis and meta-analysis. *Oncotarget* 2017 Apr 04;8(14):23978-23995 [FREE Full text] [doi: [10.18632/oncotarget.15074](https://doi.org/10.18632/oncotarget.15074)] [Medline: [28177907](https://pubmed.ncbi.nlm.nih.gov/28177907/)]
73. Mormont MC, Langouët AM, Claustrat B, Bogdan A, Marion S, Waterhouse J, et al. Marker rhythms of circadian system function: a study of patients with metastatic colorectal cancer and good performance status. *Chronobiol Int* 2002 Jan;19(1):141-155. [Medline: [11962672](https://pubmed.ncbi.nlm.nih.gov/11962672/)]
74. Mansour D, Simcock R, Gilbert DC. Acute oncology service: assessing the need and its implications. *Clin Oncol (R Coll Radiol)* 2011 Apr;23(3):168-173. [Medline: [21465697](https://pubmed.ncbi.nlm.nih.gov/21465697/)]
75. McNiff KK, Jacobson JO. Aiming for ideal care: a proposed framework for cancer quality improvement. *J Oncol Pract* 2014 Nov;10(6):339-344. [doi: [10.1200/JOP.2014.001305](https://doi.org/10.1200/JOP.2014.001305)] [Medline: [25398953](https://pubmed.ncbi.nlm.nih.gov/25398953/)]
76. Aprile G, Pisa FE, Follador A, Foltran L, De Pauli F, Mazzer M, et al. Unplanned presentations of cancer outpatients: a retrospective cohort study. *Support Care Cancer* 2013 Feb;21(2):397-404. [doi: [10.1007/s00520-012-1524-6](https://doi.org/10.1007/s00520-012-1524-6)] [Medline: [22722887](https://pubmed.ncbi.nlm.nih.gov/22722887/)]
77. Bosanquet N, Sikora K. The economics of cancer care in the UK. *Lancet Oncol* 2004 Sep;5(9):568-574. [doi: [10.1016/S1470-2045\(04\)01569-4](https://doi.org/10.1016/S1470-2045(04)01569-4)]
78. Sambamoorthi U, Tan X, Deb A. Multiple chronic conditions and healthcare costs among adults. *Expert Rev Pharmacoecon Outcomes Res* 2015;15(5):823-832 [FREE Full text] [doi: [10.1586/14737167.2015.1091730](https://doi.org/10.1586/14737167.2015.1091730)] [Medline: [26400220](https://pubmed.ncbi.nlm.nih.gov/26400220/)]
79. Innominato P, Komarzynski S, Karaboué A, Ulusakarya A, Bouchahda M, Haydar M, et al. Home-based e-health platform for multidimensional telemonitoring of symptoms, body weight, sleep, and circadian activity: relevance for chronomodulated administration of irinotecan, fluorouracil-leucovorin, and oxaliplatin at home—results from a pilot study. *JCO Clin Cancer Inform* 2018 Feb(2):1-15. [doi: [10.1200/CCCL.17.00125](https://doi.org/10.1200/CCCL.17.00125)]
80. Ballesta A, Innominato PF, Dallmann R, Rand DA, Lévi FA. Systems chronotherapeutics. *Pharmacol Rev* 2017 Apr;69(2):161-199 [FREE Full text] [doi: [10.1124/pr.116.013441](https://doi.org/10.1124/pr.116.013441)] [Medline: [28351863](https://pubmed.ncbi.nlm.nih.gov/28351863/)]
81. CASYM Europe. 2014 Jun. URL: https://www.casym.eu/lw_resource/datapool/items/item_328/roadmap_1.0.pdf [accessed 2018-05-09] [WebCite Cache ID 6zHnluHaa]
82. Gumz ML, editor. *Circadian Clocks: Role in Health and Disease*. New York: Springer; 2016.

Abbreviations

BLE: Bluetooth Low Energy
BT: Bluetooth
CTS: circadian timing system
eHealth: electronic health
GPRS: general packet radio service
IQ: interquartiles
IQR: interquartile range
PRO: patient-reported outcome

Edited by G Eysenbach; submitted 05.01.18; peer-reviewed by A Goldbeter, J Harsh; comments to author 28.02.18; revised version received 06.03.18; accepted 07.03.18; published 11.06.18.

Please cite as:

Komarzynski S, Huang Q, Innominato PF, Maurice M, Arbaud A, Beau J, Bouchahda M, Ulusakarya A, Beaumatin N, Breda G, Finkenstädt B, Lévi F

Relevance of a Mobile Internet Platform for Capturing Inter- and Intrasubject Variabilities in Circadian Coordination During Daily Routine: Pilot Study

J Med Internet Res 2018;20(6):e204

URL: <http://www.jmir.org/2018/6/e204/>

doi: [10.2196/jmir.9779](https://doi.org/10.2196/jmir.9779)

PMID: [29704408](https://pubmed.ncbi.nlm.nih.gov/29704408/)

©Sandra Komarzynski, Qi Huang, Pasquale F Innominato, Monique Maurice, Alexandre Arbaud, Jacques Beau, Mohamed Bouchahda, Ayhan Ulusakarya, Nicolas Beaumatin, Gabrièle Breda, Bärbel Finkenstädt, Francis Lévi. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.06.2018. 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.

Review

Telemedicine Use Among Caregivers of Cancer Patients: Systematic Review

Chiara Marzorati^{1,2}, MSc; Chiara Renzi¹, PhD; Samuel William Russell-Edu³, BSc, Dip Lib; Gabriella Pravettoni^{1,2}, PhD

¹Applied Research Division for Cognitive and Psychological Science, European Institute of Oncology, Milan, Italy

²Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy

³Library, European Institute of Oncology, Milan, Italy

Corresponding Author:

Chiara Marzorati, MSc

Applied Research Division for Cognitive and Psychological Science

European Institute of Oncology

via Ripamonti 435

Milan,

Italy

Phone: 39 0257489207

Email: chiara.marzorati@ieo.it

Abstract

Background: The number of published studies and systematic reviews examining different telehealth interventions targeting patients and their effects on patients' well-being and quality of life have grown in recent decades. However, the use of telemedicine tools aimed at the family members and caregivers of adult cancer patients is less defined.

Objective: We aimed to conduct a systematic review to provide a more complete picture regarding telemedicine tools for informal caregivers (usually family members or close friends) implemented in all phases of cancer care. More specifically, the review aimed to better describe the study samples' characteristics, to analyze measured outcomes and the specific questionnaires used to assess them, and to describe in depth the implemented interventions and their formats. Finally, we examined the role of telehealth, and usability and feasibility trends in supporting patients' caregivers.

Methods: We systematically searched the literature in the following databases: Web of Science, Cochrane Library, PubMed, Scopus, CINAHL, MEDLINE, EMBASE, Google Scholar, and PsycINFO. Inclusion criteria were being written in English, published in peer-reviewed journals, describing a telehealth-implemented intervention, and focusing on caregivers of adult cancer patients at any stage of the disease. We selected studies published up to November 2017. We critically appraised included articles using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses and graded the quality of evidence by outcome using the Centre for Evidence-Based Medicine framework.

Results: We included 24 studies in the final selection. In 21 of the 24 studies, the patient-caregiver dyad was analyzed, and the study population dealt with different types of cancer at different stages. Included studies considered the caregiver's condition from both an individual and a relational point of view. Along with psychosocial variables, some studies monitored engagement and user satisfaction regarding Web-based platforms or telehealth interventions. All studies reported significant improvements in some of the investigated areas, but they often showed small effect sizes. Two types of telehealth intervention formats were used: Web-based platforms and telephone calls. Some of the included studies referred to the same project, but on study samples with different cancer diagnoses or with new versions of previously developed interventions.

Conclusions: Reported outcomes seem to suggest that we are in an exploratory phase. More detailed and targeted research hypotheses are still needed. Clarifying caregivers' needs related to telehealth tools and better defining outcome measures may yield more significant results.

(*J Med Internet Res* 2018;20(6):e223) doi:[10.2196/jmir.9812](https://doi.org/10.2196/jmir.9812)

KEYWORDS

telemedicine; family; caregivers; neoplasms; systematic review

Introduction

Telemedicine Overview

Information and communication technology (ICT) has in recent decades become essential in supporting information provision, sharing data, overcoming face-to-face boundaries, and meeting people's needs [1]. In the medical field, transferring information through telecommunication networks easily provides an opportunity for innovation, and helps in managing resources and increasing health care quality.

The use of medical information exchanged from one site to another via electronic communications to improve a patient's clinical health status is defined as telemedicine [2]. It includes a variety of applications that allow the transfer of eHealth data. eHealth systems allow many different hospital facilities to cooperate to improve health care services, patient engagement, monitoring, and management, and to provide prompt access to expert advice and patient information, regardless of where patients are or where information is collected. From this perspective, ICT may support the global drive to achieve consistent, integrated, sustainable, high-quality, and cost-efficient health care [3,4]. Web-based interventions present innovative methods for using and improving public health services with easily accessible, up-to-date, and tailored information, education programs, self-management training and monitoring, and family-physician communication [5-11].

To date, eHealth interventions have mostly been implemented to support the self-management of "the big five" diseases identified by the World Health Organization, namely diabetes mellitus, cardiovascular and chronic respiratory diseases, cancer, and stroke [12]. A scoping review showed how people affected by chronic conditions used ICT especially for self-management, thus enhancing patient engagement. The broadest category where ICT interventions were implemented was cancer care, with specific focus on shared management activities among patients and their providers [13].

Telemedicine for Caregivers of Cancer Patients

Alongside the development of telehealth interventions aimed at cancer patients, attention is increasingly being directed toward telemedicine tools aimed at satisfying the needs of caregivers.

Caregivers are usually family members or close friends whose efforts to care for their loved ones have a considerable physical and psychological impact on them. Family members are often considered fundamental in the process of care, especially for those diseases that require continuous or extended treatments. Demographic and health trends among the European population are increasing the need for reorganizing and delivering better and more cost-effective health services [14], not only for patients but also for caregivers. Caregiver care has thus become a core topic of contemporary scientific research because it can be related to prevention: if more attention and assistance is given to caregivers, they will experience fewer physical and psychological impairments, thereby having less of an impact on the health care system from an economic or a social perspective [15,16]. Literature reviews and meta-analyses confirm the association between greater mental burden and

poorer physical and mental well-being: responsibilities and stressful experiences related to the caregiving role can lead to depression, anxiety, worry, and loneliness [17-22]. Similarly, the greater mental burden and emotional distress caregivers experience can result in fatigue, sleep impairment, and unhealthy behaviors [23-26]. Several studies have demonstrated highly distressing conditions among caregivers, affecting them not only psychologically but also physiologically. Depression, anxiety, or poor sleep quality can cause a decline in immunocompetence and can be associated with the onset of cardiovascular disease or earlier death [27-29]. The physical and psychological impairments of caregivers are well documented in the field of oncology; poorer physical health of cancer patients is significantly associated with a deterioration of physical health among family members [30], as well as with symptoms of depression or anxiety [31,32]. Providing cancer care for years or resuming care before the patient's death can also be related to the emergence of arthritis, heart diseases, and chronic back pain in the caregiver [33]. These studies have shown how cancer caregiving is highly demanding and emotionally burdensome, leading to the need for information to manage patients' symptoms or improve knowledge in medical procedures. Longacre [34] classified caregivers' information needs into personal psychosocial care, the provision of direct care, and care management. She pointed out that meeting those needs positively interacts with caregivers' perception of managing emotional and physical stress. Another systematic review [35] showed how needs were unmet mostly in terms of diagnosis- and prognosis-related information, information about the impact on the family or partner, information on practical issues, coping information, and medical information. Caregivers also asked for support for their psychological condition and their fears concerning the patient's disease progression or recurrence [36].

The possibility of creating new direct and interactive interventions—directed not only at patients, but also at caregivers—places greater attention on eHealth tools in the context of long-term diseases.

Caregivers, as well as patients, are increasingly using apps and Web-based interventions to cope with their uncertainty and need for information. Caregivers need to be informed about and prepared for patient symptoms or side effects, and they want better knowledge to counter their fears of inadequacy, for example [37,38].

Even though much has been done concerning patient empowerment, more attention needs to be paid to the effects and support of telemedicine on family caregivers and on how promising eHealth programs are in responding to their needs [39,40]. Despite caregivers' requests for provision of support and information competence, a recent meta-review on the effects of eHealth for cancer patients and caregivers concluded that there is indeed a paucity of systematic reviews on this topic and that Web-based interventions focused on family members are still an unexplored area [41].

The number of published studies and systematic reviews examining many types and effects of Web-based interventions targeting patients have increased in recent decades [42-44].

However, less is known about telehealth interventions aimed at the cancer patient's family members. Scoping reviews have been conducted only on Web-based interventions or on the effects of eHealth tools for cancer patients and their informal caregivers [41,45,46], while others had a broader focus on all implemented telehealth tools for family caregivers, but not specifically involved in cancer care [39]. Therefore, the need for obtaining a more complete picture of implemented telemedicine tools for caregivers in all phases of cancer care is emerging.

Objectives

This systematic review aimed to describe the main characteristics of previously developed telehealth tools for family members of cancer patients. More specifically, the objectives of the study were to better describe the samples' characteristics, to specify the measured outcomes and the specific questionnaires used to assess them, and to describe in depth the implemented interventions and their formats. Alongside the implementation of telemedicine systems for caregivers, we hoped to identify the main considered outcomes, to analyze the role of eHealth technology, and to discuss the usability and feasibility trends in supporting patient caregivers.

Methods

We conducted a systematic review of studies on telehealth-based intervention for caregivers of cancer patients at any stage of the disease.

Search Strategy

We systematically searched the following databases: Web of Science, Cochrane Library, PubMed, Scopus, CINAHL, MEDLINE, EMBASE, Google Scholar, and PsycINFO.

We used various combinations of database-specific controlled vocabularies (subject headings), supplemented by keywords, and title and abstract terms for the concepts and synonyms relating to telemedicine, telehealth, Web-based intervention, eHealth, mHealth, carers, caregivers, family, and cancer. We examined bibliographies and reference lists of relevant articles and identified citing articles using Web of Science. No time restrictions were applied. English language restriction was applied. [Multimedia Appendix 1](#) reports the full search strategies we used.

Selection Strategy

One of the authors (SWRE), a qualified medical librarian, conducted the systematic literature search. Two other authors (CR and CM) selected articles for full review based on the

inclusion and exclusion criteria and assessed their eligibility. Agreement was reached on the final selection of included studies.

For study inclusion in this systematic review, we applied the following selection criteria: (1) written in English, (2) published in peer-reviewed journals, (3) including a telehealth-implemented intervention (4) involving human participants, and (5) focusing on caregivers of adult patients at any cancer stage. We excluded studies that did not involve human participants or did not have an experimental study design (eg, commentary, review, or expert opinions). We selected studies published up to November 2017.

Review Strategy and Data Extraction

The initial search resulted in 655 articles. We also searched the reference lists of relevant articles to identify other articles. We excluded 413 articles based on a review of titles and keywords. Subsequently, we excluded 170 articles based on their abstracts because they did not meet the inclusion criteria. After eliminating 48 duplicates, we included 24 studies in the final selection.

We applied the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. [Figure 1](#) shows the PRISMA flowchart.

To evaluate the strength of the studies' findings, we also scored each article for the level of evidence according to the Centre for Evidence-Based Medicine framework: 1a: meta-analyses; 1b: individual randomized controlled trials (RCTs); 1c: non-RCTs; 2a: systematic reviews of cohort studies; 2b: individual cohort studies; 2c: outcomes research; 3a: systematic reviews of case-control studies; 3b: individual case-control studies; 4: case series; and 5: expert opinions without explicit critical appraisal [47].

We used a standardized form for data extracted from the included articles, outlining the year of publication, authors, study country, aim, sample characteristics, study design, type of intervention, measured outcomes, assessment, and principal results.

We grouped the included studies into 3 subcategories according to which kind of intervention was implemented: eHealth intervention, telephone sessions, or both. We divided the measured outcomes into clinical and usability subgroups, then split clinical outcomes into psychosocial (in turn divided into the caregiver's individual and dyadic dimensions) and behavioral factors.

[Figure 2](#) shows the categorization of the measured outcomes.

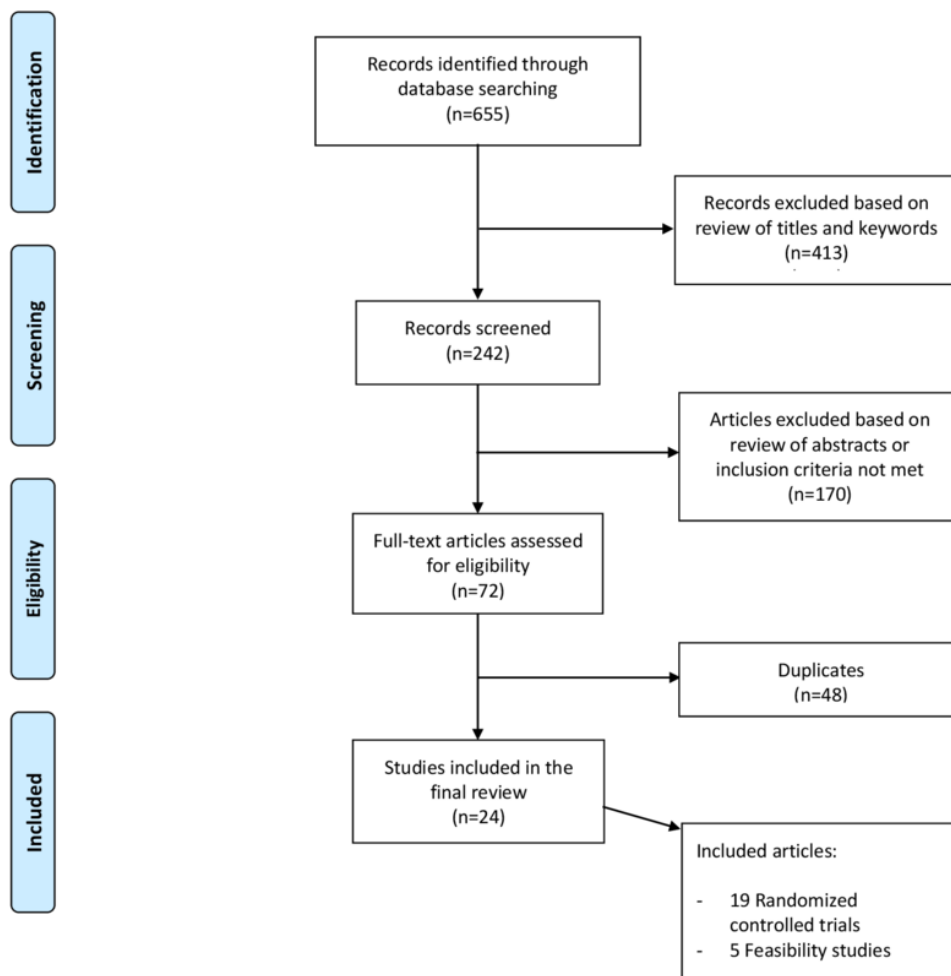
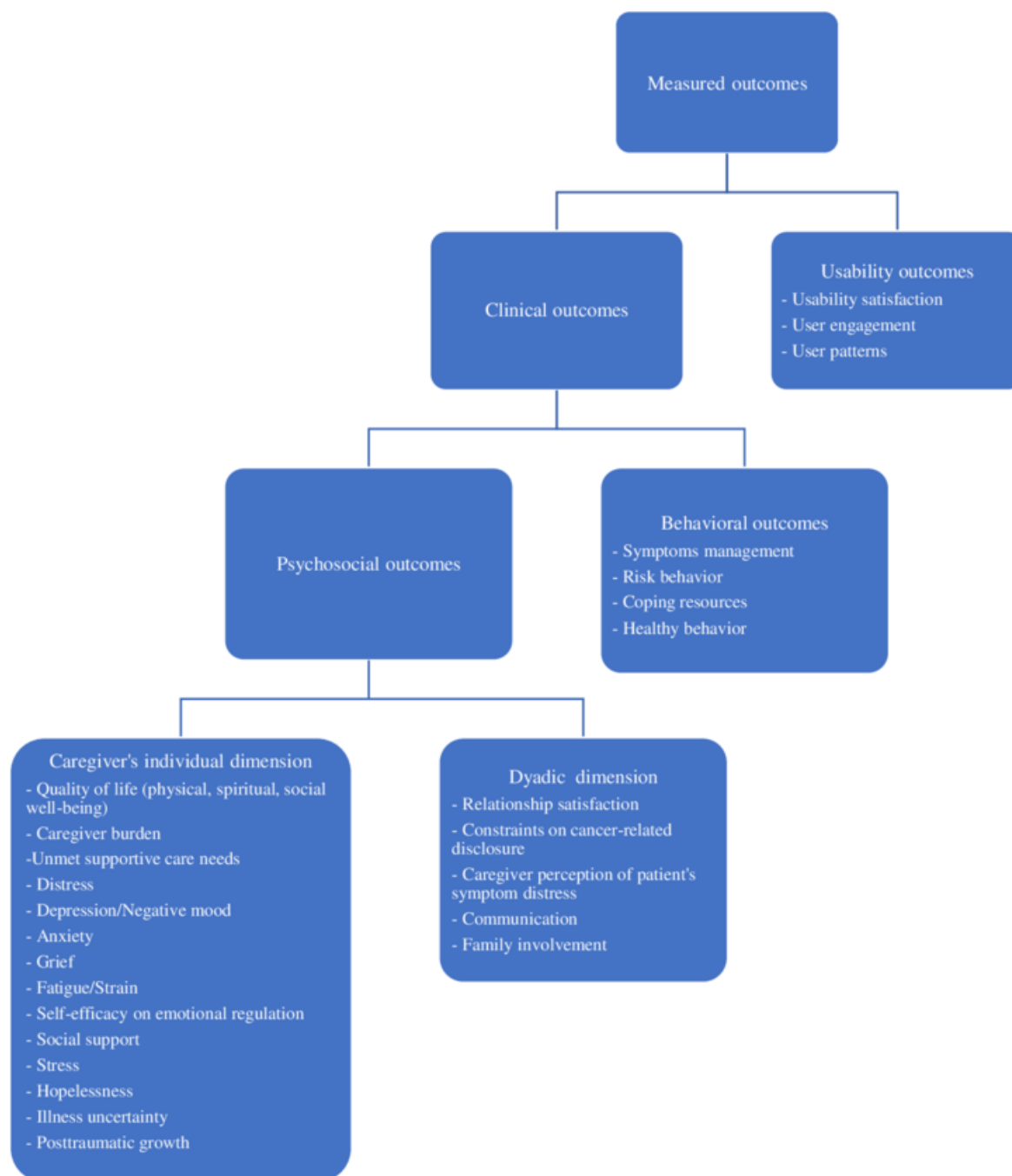
Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart.

Figure 2. Categorization of measured outcomes in the 24 reviewed articles.

Results

Study Designs

Study designs comprised 19 RCTs (1b) and 5 feasibility studies (1c). Of the RCTs, 16 had 2 arms, whereas the remaining 3 had 3 arms; 7 did not include a no-treatment control group. All RCTs provided follow-up assessment at 3 and 6 months at least.

Caregiver Characteristics

A total of 3301 caregivers of cancer patients were enrolled, with the number of family caregivers in each study ranging from 6 to 481 (see [Multimedia Appendix 2](#) [48-71]). Of the 24 studies, only 1 did not provide sufficient data on their sample's sex and age statistics (N=12) [48]. Of the 3289 remaining family

members, 70.75% (2327/3289) were female and 56 years old on average. In 21 of the 24 studies, the patient-caregiver dyad was analyzed; interventions were delivered to both patients and caregivers, and outcomes were measured in both populations. In 3 studies, caregivers were the only target sample and patients variables were not considered [49-51]. A total of 3 studies focused the intervention only on women, 2 studies examined partners [52,53], and 1 included caregivers with different relationships to patients [54]. In many studies, patients identified not only their partners as their main caregivers (2590/3301, 78.46%), but also adult children, siblings, parents, or friends. Dyads dealt with different types of cancer at different stages; among the included studies, lung and gastrointestinal cancers were the most considered, followed by genitourinary and breast tumors, and hematological neoplasms (see [Figure 3](#)). Of the 24

studies, 8 focused on advanced cancers [48,51,54-59] and 2 on early-stage tumors [60,61]; the other trials included patients with all stages of the disease.

Study attrition rate varied from 3% to 64%. The main reasons for withdrawal were patient death, lack of interest, and medical condition (eg, progression of the disease).

Most of the studies were conducted in the United States (see [Multimedia Appendix 2](#)), except for 3 studies carried out in Australia [50,62], 1 in Sweden [63], and 1 in Canada [48].

Measured Outcomes

Clinical and usability outcomes were measured in 7 studies [48,50,58,63-66]; in the remaining studies, only the first cluster was included. The caregiver's well-being was assessed considering various psychosocial and behavioral variables, except for 2 studies that used a single outcome measure. Kinney et al [49] studied only change in colonoscopy prevention behavior, and Clark et al [67] assessed quality of life among caregivers of patients with advanced cancer undergoing radiotherapy treatment (see [Multimedia Appendix 2](#)). Other studies evaluated several outcomes, ranging from 2 [55,57] to 12 [59].

The 2 most-assessed behavioral outcomes were coping resources and symptom management. Coping resources were assessed either by the Brief COPE questionnaire [54,58] or by combining multiple questionnaires: the Lewis Mutuality and Interpersonal Sensitivity Scale, the Brief version of the Social Support Scale, and the Lewis Cancer Self-efficacy Scale [52,59,68]. Symptom management, on the other hand, was assessed only by qualitative analysis on audio-recorded interviews [65,69].

We created 2 categories of psychosocial outcome measures: one cluster included the caregiver's individual dimensions, and the other cluster was related more to dyadic interactions (eg, relationship with partner, perception of patient's health condition). In the first category, the most examined constructs

were quality of life [50,52,54,56,59,62,66-68] and distress [50,52,53,59,61,62,64,68,70,71]. Quality of life was assessed with the Functional Assessment of Cancer Therapy-General (FACT-G; n=4), the Short Form Survey 36-item and 12-item versions (n=1), the Caregiver Quality of Life Scale-Cancer (n=2), the European Organisation for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30; n=1), and the Medical Outcomes Study 12-item Short Form Survey (n=1). Questionnaires used to measure the distress or stress condition were the Profile of Mood States (n=2), the General Symptom Distress Scale (n=2), the 77-item Omega Screening Questionnaire (n=2), the Distress Thermometer (n=1), the Brief Symptom Inventory 18 (n=1), the Perceived Stress Scale (PSS; n=1), and the Posttraumatic Stress Disorder Symptom Scale to measure cancer-specific distress (n=1). In addition, the caregiver's depression [50,56,57,60,64,70], social support [58,59,64,66,68,70], and self-efficacy [52,59,61,68,69] were taken into account in almost one-third of the trials. Anxiety [69,70], hopelessness [52,54,59], fatigue [61,64,70], cancer knowledge [52,54,59,68], spiritual well-being [60,64,70], and uncertainty [52,54,59,65] were also measured in several studies (see [Multimedia Appendix 2](#)).

The second cluster, concerning the caregiver's relationship with the patient, evaluated perceived support such as difficulties encountered in communicating with the patient about the disease and the caregiver's perception of the patient's symptom management [51,55,69]. The Edmonton Symptom Assessment Scale was the main tool used to assess these variables.

Along with clinical variables, some studies monitored engagement and user satisfaction regarding the Web-based platforms or telehealth interventions. User satisfaction and device usability were explored through open-ended questions [60,65,66], single-item questions such as "How comfortable are you using the internet?" [58], or semistructured interviews [48,50,63] that were audio recorded, transcribed verbatim, and coded with latent content analysis.

Figure 3. Cancer diagnoses considered in the 24 reviewed articles.

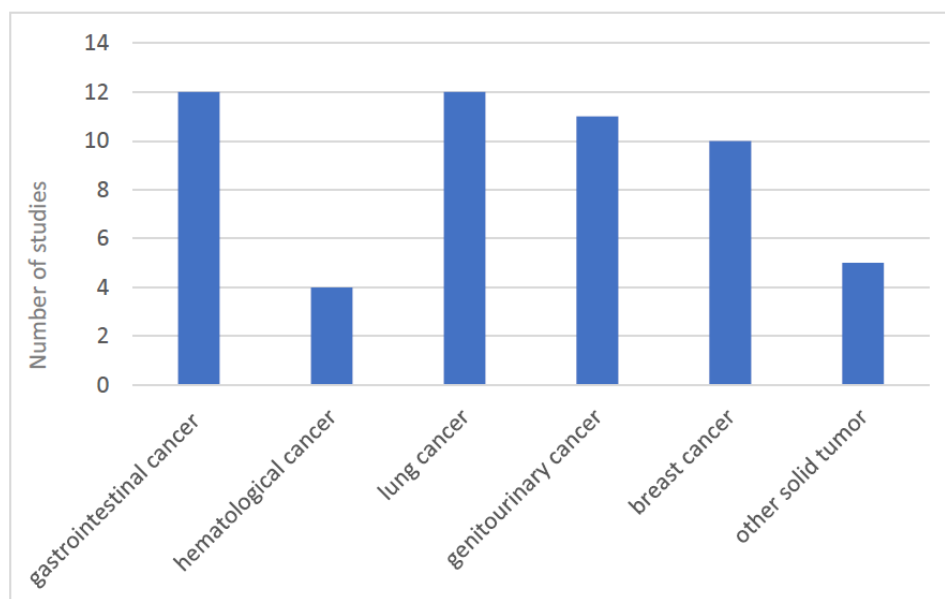
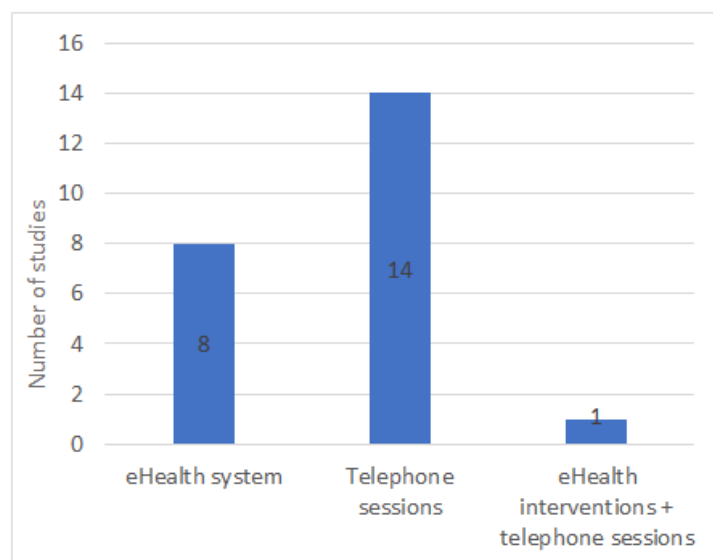


Figure 4. Intervention formats.

All studies reported significant improvements in some of the investigated areas, but they often showed small effect sizes. Even though statistically significant outcomes differed among the included studies, some of them were significant in more than 1 study measuring that specific outcome. These improved outcomes were caregiver self-efficacy, quality of life, distress, depression, appraisal of caregiving, and perceived social support. Caregiver self-efficacy, both in managing one's own emotions and in helping patients to control symptoms, was measured in 5 studies [52,59,61,68,69] and was always statistically significant. A total of 5 studies reported significant differences in quality of life: 3 showed significant effects in all dimensions of quality of life [52,64,68]; 2 studies showed significant effects in spiritual and social well-being [59,64]; and 1 study [60] showed significant outcomes in social well-being. Emotional distress or stress symptoms were significantly different between pre- and posttreatment assessments in 5 of the considered studies [50,52,64,68,71], while both depressive symptoms [55,57,64] and perceived social support [62,64,66] improved in 3 trials. Furthermore, the included articles reported significant, though smaller, effects in other measured outcomes: after completing the interventions, family members also experienced less anxiety, less sense of disruptiveness, hopelessness, and uncertainty, less burden, and negative appraisal of caregiving.

Technology Use and Intervention Format

Different telehealth interventions were conducted in the analyzed studies. [Multimedia Appendix 3](#) provides a full description of the interventions delivered in each study.

In 12 studies, an intervention was developed for a sole cancer type [49,51-54,58,60,61,64,66,69,70]; the other studies implemented telemedicine tools for a multifaceted sample including patients with different cancer diagnoses [48,50,55-57,59,62,63,65,67,68].

A total of 2 studies, 2 conducted with a qualitative methodology [65] and 2 with a mixed-methods methodology [48], were scheduled for a one-time-only data collection, unlike all other trials, in which at least two follow-ups were planned after study completion.

The included studies used various telehealth intervention formats: some studies involved the development of a Web-based platform, while others used scheduled telephone calls to improve the dyads' psychosocial condition. More precisely, 8 studies [50,51,55,58,63,65,66,68] implemented eHealth interventions aimed at exploring caregiver coping strategies, emotional well-being, and patient symptom management. Schover et al [53] combined the eHealth system with telephone sessions. All other studies used only telephone calls with supplemented written material to support patients and caregivers in their process of care (see [Figure 4](#)).

Of the 20 studies that provided for the presence of a practitioner to conduct part of the described interventions, 8 were conducted by trained nurses [48,52,54,56,59,61,62,65], 4 were carried out by social workers [60,64,69,70], and 3 were conducted at a distance by caregivers using Web-based platforms [51,55,58]. In 3 RCTs the intervention was delivered by clinical psychologists [53,62,71], and in another trial a genetic counsellor conducted the telehealth risk communication to promote colonoscopy screening [49]. Another 2 studies provided psychosocial support via a multidisciplinary team including a chaplain, counsellor, dietician, physiotherapist, and physician [63], or a clinical psychologist, psychiatrist, advanced practice nurse, hospital chaplain, clinical social worker, psychiatrist, and physical therapist [67]. Master's-level nurses, clinical social workers, and psychologists were trained, and sometimes allocated to multidisciplinary teams, to help caregivers in their process of patient care. All participants in both the experimental and the control groups received the telehealth intervention, except in 7 studies, where the control group received only usual care [49,52,54,59,62,66,67].

Some of the included studies referred to the same project but referred to population samples with a different cancer diagnosis or with new versions of previously developed interventions. Dionne-Odom et al [56,57] published 2 articles on patients with advanced cancer and their caregivers under the third Educate, Nurture, Advise Before Life Ends project; and 3 studies [51,55,58] were conducted within the Comprehensive Health

Enhancement Support System program. The family involvement, optimistic attitude, coping effectiveness, uncertainty reduction, and symptom management (FOCUS) Program comprised 5 studies [52,54,59,65,68], and Badger et al conducted 3 RCT studies on breast [60,70] and prostate [64] cancer patients and their partners to examine the effectiveness of 2 different telephone interventions.

Discussion

Principal Findings

We systematically reviewed 24 studies implementing telehealth tools for caregivers of cancer patients. Interventions used telephone calls or eHealth systems aimed at improving the physical and mental well-being of the study populations to satisfy different user needs.

In this systematic review, only one-third of the studies used eHealth systems to investigate psychosocial outcomes. All the studies using Web platforms were published after 2011, except for 1 article published in 2002 [66], reflecting the rapid development of eHealth technology over the last few years. Most of the included studies instead implemented a supportive or educational intervention based on scheduled telephone calls and the distribution of written material. Badger et al [60] investigated users' preferences for telephone, videophone, or face-to-face methods: dyads agreed that the telephone intervention (69% of patients and 73% of caregivers) was the most reliable and easy-to-use system, compared with a videophone intervention, which in turn was preferred to face-to-face interaction. Only 1 study [53] used both an eHealth system and telephone sessions, even though they were not compared, but formed part of the same intervention. The proportion of participants favoring telephone-delivered interventions over eHealth interventions suggests that the implementation of Web-based platforms in health care systems is still in development and that further research is needed.

Web-based interventions facilitate participant enrollment and data collection from patients, reduce the risk of missing items, overcome geographic and mobility problems, and are more cost effective, but researchers have less control over participants [72-75].

Telephone-based interventions, on the other hand, include personalized therapist guidance. This can positively influence patient outcomes and proactively support potential crises by virtue of a more individualized and tailored intervention than that delivered by telehealth programs [76]. These telephone sessions may also contribute to enhancing the sense of independence and autonomy for patients and caregivers [77]. However, published studies directly comparing internet-based versus telephone intervention are lacking. The use of Web-based platforms may thus reflect the shift in social and cultural trends related to the use of ICT rather than an actual evidence-based advantage of Web-based over telephone interventions [78].

In addition to the variability in telehealth interventions, there was also a consistent variability in the considered outcome measures used in the studies. In 24 articles, 30 different caregiver outcomes were measured to evaluate the interventions,

with most being psychosocial variables. These psychosocial constructs were measured using a variety of questionnaires. For instance, studies considering quality of life as an outcome used either the 12-item Short Form Survey, the FACT-G, or the EORTC QLQ-C30 as a measurement tool. This finding is consistent with other studies assessing the effects of supportive telehealth interventions on psychosocial variables: Agboola et al [79] found a nonuniformity in measured outcomes and questionnaires assessing quality of life, depression, and pain management in patients with cancer.

Considering this, it could be interesting to identify the main tools and variables on which to focus. This would enable us to better define the theoretical framework within which the study programs are developed and to analyze results, and thus to disentangle explanatory relations between different variables. According to a literature review on telehealth interventions [80], a precise structure linking all aspects of the intervention or of the outcomes is rarely used, and most of the projects lack a theoretical framework. Nevertheless, even in the general literature (beyond that relating to telehealth interventions), there are unfortunately only a few studies that assessed correlations or mediated effects between different constructs (eg, quality of life, depression, and self-efficacy) that are related to caregivers' needs (eg, information provision, social support, and self-management education) [81,82].

In most of the projects, telehealth tools were considered as a given: most of the studies focused on the efficacy of the tool to promote caregivers' well-being, overshadowing the usability and feasibility of the eHealth programs. The accessibility and usability of the technology have not often been assessed, and only 6 of the 24 studies assessed user satisfaction or Web-based program usage patterns [48,50,58,60,63,65,66]. The lack of investigation in this area may prevent a correct evaluation of and improvement in the effectiveness of the implemented telehealth tools. In fact, without assessing all the included aspects of the effectiveness concept (user satisfaction, usefulness, interaction quality, and ease of use), it would be more difficult to understand whether it is the specific tool that does not function or whether it could even be telehealth interventions in general [83,84].

Different aspects of each intervention may benefit to a greater or lesser extent from using an eHealth delivery. For instance, using eHealth may enable screening for aspects of caregivers' well-being, which may otherwise remain unconsidered due to lack of resources or due to inefficiency in standard care flows.

In accordance with other literature reviews and meta-analyses [41,85], our review found that family members who used telehealth tools reported a perception of increased social support [58,60,64,68] and a less negative appraisal of illness and caregiving [52,54,68], even though the overall effect sizes were small. These findings meet Kent and colleagues' research recommendations to improve the assessment of the prevalence and burden of informal cancer caregiving [86], emphasizing the need to direct attention toward the most vulnerable caregivers of cancer patients, such as those socially isolated, living in rural areas, or with low socioeconomic status. Social isolation and low appraisal of caregiving, along with depression, financial

stress, and lack of choice in being a caregiver, are important risk factors of caregiver strain, affecting their perception of burden [87,88]. Therefore, understanding the impact of caregiving and developing tailored interventions to provide assistance to caregivers can satisfy important unmet needs and reduce caregivers' psychological and emotional burden [89].

In some cases, different studies of the same project reported conflicting results on psychological outcomes. For instance, Badger et al [60] found significant improvements in perceived social well-being in the telephone and videophone interpersonal counselling group, but not in the health education group. This contrasts with 2 years previously, when they reported greater improvements in the health education group than in the videophone counselling group [64] in the same variable. This variability may be related to the enrollment of individuals in different disease stages, undergoing different treatments, or having different psychological and social characteristics, with these differences applying to both the patient and the caregiver. It is known that different aspects of the disease or of the treatment may imply different caregiving burdens [41], as well as different psychological or relational issues [90-94]. Individuals may be more or less able to manage the demands related to caregiving, depending on, for example, their socioeconomic status, literacy level, personality traits, resilience, and contingent factors [95,96]. It follows that it may be necessary to compare the same intervention across different caregiver populations and, further, to personalize the intervention depending on aspects that turn out to be significant in determining the outcome(s).

While usability testing and psychological variables have been sufficiently, though not equally, considered, studies are lacking that assess the specific dimension related to changes in the caregivers' perception of their role. More precisely, studies did not include specific measures to detect differences in the caregivers' appraisal of patient management after using telehealth interventions. For example, it is difficult to

disentangle whether and to what extent changes in caregivers' self-efficacy were directly related to use of the tool, since most of the studies had no control group. So far, the design of the studies has not allowed for evaluation of whether the use of the tools (dose, frequency, or satisfaction) was a mediator of the outcome (eg, self-efficacy).

It is clear that telehealth implies not only the mere use of electronic services to store medical data, but also a more complex framework. This includes practitioners' education, patients' and caregivers' empowerment, efficiency, equity, quality of service provision, and promotion of shared decision processes at local, national, and global health care levels [97].

It would be interesting to reconsider future directions: reported outcomes seem to suggest that we are in an exploratory phase. There is still a need to construct more detailed and targeted research hypotheses. The lack of theoretical frameworks leads to the implementation of broad yet possibly weak interventions, targeting many different constructs or aspects, and thus may lead to nonsignificant results or to small effect sizes. Clarifying caregivers' needs related to telehealth tools and better defining outcome measures may allow us to obtain more significant results.

Conclusion

It is crucial to identify unmet family needs or priority clusters and to take into account the relation to cost-effectiveness trends. There is a paucity of studies assessing the economic value of psychosocial interventions with standardized methods [98,99]. Future studies can provide further cost-related information to support decision-making processes and the planning of new large-scale care services. To achieve value-based health care, it is important to devise cost-effective study designs and to implement the most appropriate data collection methods and procedures in order for the results to be generalizable across different populations and contexts [100].

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 643529.

This systematic review reflects the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[PDF File (Adobe PDF File), 155KB - [jmir_v20i6e223_app1.pdf](#)]

Multimedia Appendix 2

Details of the included studies.

[PDF File (Adobe PDF File), 325KB - [jmir_v20i6e223_app2.pdf](#)]

Multimedia Appendix 3

Formats of interventions delivered in each study.

[[PDF File \(Adobe PDF File\), 157KB - jmir_v20i6e223_app3.pdf](#)]

References

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
2. Telemedicine glossary. Arlington, VA: American Telemedicine Association; 2018. URL: <http://thesource.americantelemed.org/resources/telemedicine-glossary> [accessed 2018-06-07] [WebCite Cache ID 700EQNaV9]
3. Stroetmann KA. Achieving the integrated and smart health and wellbeing paradigm: a call for policy research and action on governance and business models. *Int J Med Inform* 2013 Apr;82(4):e29-e37. [doi: [10.1016/j.ijmedinf.2012.05.008](https://doi.org/10.1016/j.ijmedinf.2012.05.008)] [Medline: [22727880](#)]
4. Farhadi M, Ismail R, Fooladi M. Information and communication technology use and economic growth. *PLoS One* 2012;7(11):e48903 [FREE Full text] [doi: [10.1371/journal.pone.0048903](https://doi.org/10.1371/journal.pone.0048903)] [Medline: [23152817](#)]
5. 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](https://doi.org/10.2196/jmir.2281)] [Medline: [23425685](#)]
6. Gorini A, Mazzocco K, Kondylakis H, McVie G, Pravettoni G. A web-based interactive tool to improve breast cancer patient centredness. *Ecancermedicalscience* 2016;10:659 [FREE Full text] [doi: [10.3332/ecancer.2016.659](https://doi.org/10.3332/ecancer.2016.659)] [Medline: [27563354](#)]
7. Kondylakis H, Kazantzaki E, Koumakis L, Genitsaridi I, Marias K, Gorini A, et al. Development of interactive empowerment services in support of personalised medicine. *Ecancermedicalscience* 2014;8:400 [FREE Full text] [doi: [10.3332/ecancer.2014.400](https://doi.org/10.3332/ecancer.2014.400)] [Medline: [24567757](#)]
8. Kondylakis H, Koumakis L, Tsiknakis M, Marias K, Genitsaridi E, Pravettoni G, et al. Smart recommendation services in support of patient empowerment and personalized medicine. In: Tsihrintzis G, Virvou M, Jain L, editors. *Multimedia Services in Intelligent Environments. Smart Innovation, Systems and Technologies*. Heidelberg: Springer; 2013:39-61.
9. Kondylakis H, Koumakis L, Kazantzaki E, Chatzimina M, Psaraki M, Marias K, et al. Patient empowerment through personal medical recommendations. *Stud Health Technol Inform* 2015;216:1117. [Medline: [26262416](#)]
10. Kondylakis H, Koumakis L, Genitsaridi E, Tsiknakis M, Marias K, Pravettoni G, et al. IEmS: a collaborative environment for patient empowerment. 2012 Presented at: 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering; Nov 11-13, 2012; Larnaca, Cyprus p. 535-540. [doi: [10.1109/BIBE.2012.6399770](https://doi.org/10.1109/BIBE.2012.6399770)]
11. Gilardi S, Guglielmetti C, Pravettoni G. Interprofessional team dynamics and information flow management in emergency departments. *J Adv Nurs* 2014 Jun;70(6):1299-1309. [doi: [10.1111/jan.12284](https://doi.org/10.1111/jan.12284)] [Medline: [24138152](#)]
12. Alwan A. Global status report on noncommunicable diseases 2010. Geneva, Switzerland: World Health Organization; 2011. URL: http://www.who.int/nmh/publications/ncd_report_full_en.pdf [accessed 2018-06-06] [WebCite Cache ID 6zyXtgQLT]
13. Wildevuur SE, Simonse LWL. Information and communication technology-enabled person-centered care for the “big five” chronic conditions: scoping review. *J Med Internet Res* 2015;17(3):e77 [FREE Full text] [doi: [10.2196/jmir.3687](https://doi.org/10.2196/jmir.3687)] [Medline: [25831199](#)]
14. Tsiknakis M, Kouroubali A. Organizational factors affecting successful adoption of innovative eHealth services: a case study employing the FITT framework. *Int J Med Inform* 2009 Jan;78(1):39-52. [doi: [10.1016/j.ijmedinf.2008.07.001](https://doi.org/10.1016/j.ijmedinf.2008.07.001)] [Medline: [18723389](#)]
15. Nichols LO, Chang C, Lummus A, Burns R, Martindale-Adams J, Graney MJ, Resources for Enhancing Alzheimer's Caregivers Health II Investigators. The cost-effectiveness of a behavior intervention with caregivers of patients with Alzheimer's disease. *J Am Geriatr Soc* 2008 Mar;56(3):413-420 [FREE Full text] [doi: [10.1111/j.1532-5415.2007.01569.x](https://doi.org/10.1111/j.1532-5415.2007.01569.x)] [Medline: [18179480](#)]
16. Kim Y, Carver CS. Recognizing the value and needs of the caregiver in oncology. *Curr Opin Support Palliat Care* 2012 Jun;6(2):280-288. [doi: [10.1097/SPC.0b013e3283526999](https://doi.org/10.1097/SPC.0b013e3283526999)] [Medline: [22436321](#)]
17. Pinquart M, Sorensen S. Correlates of physical health of informal caregivers: a meta-analysis. *J Gerontol Ser B Psychol Sci Soc Sci* 2007 Mar 01;62(2):P126-P137. [doi: [10.1093/geronb/62.2.P126](https://doi.org/10.1093/geronb/62.2.P126)]
18. Etters L, Goodall D, Harrison BE. Caregiver burden among dementia patient caregivers: a review of the literature. *J Am Acad Nurse Pract* 2008 Aug;20(8):423-428. [doi: [10.1111/j.1745-7599.2008.00342.x](https://doi.org/10.1111/j.1745-7599.2008.00342.x)] [Medline: [18786017](#)]
19. Buyck J, Bonnaud S, Boumendil A, Andrieu S, Bonenfant S, Goldberg M, et al. Informal caregiving and self-reported mental and physical health: results from the Gazel Cohort Study. *Am J Public Health* 2011 Oct;101(10):1971-1979. [doi: [10.2105/AJPH.2010.300044](https://doi.org/10.2105/AJPH.2010.300044)] [Medline: [21493948](#)]
20. Faller H, Schuler M, Richard M, Heckl U, Weis J, Küffner R. Effects of psycho-oncologic interventions on emotional distress and quality of life in adult patients with cancer: systematic review and meta-analysis. *J Clin Oncol* 2013 Feb 20;31(6):782-793. [doi: [10.1200/JCO.2011.40.8922](https://doi.org/10.1200/JCO.2011.40.8922)] [Medline: [23319686](#)]

21. Koyanagi A, DeVylder JE, Stubbs B, Carvalho AF, Veronese N, Haro JM, et al. Depression, sleep problems, and perceived stress among informal caregivers in 58 low-, middle-, and high-income countries: a cross-sectional analysis of community-based surveys. *J Psychiatr Res* 2018 Jan;96:115-123. [doi: [10.1016/j.jpsychires.2017.10.001](https://doi.org/10.1016/j.jpsychires.2017.10.001)] [Medline: [29031131](https://pubmed.ncbi.nlm.nih.gov/29031131/)]
22. Schulz R, Sherwood PR. Physical and mental health effects of family caregiving. *Am J Nurs* 2008 Sep;108(9 Suppl):23-7; quiz 27. [doi: [10.1097/01.NAJ.0000336406.45248.4c](https://doi.org/10.1097/01.NAJ.0000336406.45248.4c)] [Medline: [18797217](https://pubmed.ncbi.nlm.nih.gov/18797217/)]
23. Swore FBA, Dodd MJ, Schumacher KL, Miaskowski C. Symptom experience of family caregivers of patients with cancer. *Oncol Nurs Forum* 2008 Mar;35(2):E23-E44 [FREE Full text] [doi: [10.1188/08.ONF.E23-E44](https://doi.org/10.1188/08.ONF.E23-E44)] [Medline: [19405245](https://pubmed.ncbi.nlm.nih.gov/19405245/)]
24. Stenberg U, Ruland CM, Miaskowski C. Review of the literature on the effects of caring for a patient with cancer. *Psychooncology* 2010 Oct;19(10):1013-1025. [doi: [10.1002/pon.1670](https://doi.org/10.1002/pon.1670)] [Medline: [20014159](https://pubmed.ncbi.nlm.nih.gov/20014159/)]
25. Bevens M, Sternberg EM. Caregiving burden, stress, and health effects among family caregivers of adult cancer patients. *JAMA* 2012 Jan 25;307(4):398-403 [FREE Full text] [doi: [10.1001/jama.2012.29](https://doi.org/10.1001/jama.2012.29)] [Medline: [22274687](https://pubmed.ncbi.nlm.nih.gov/22274687/)]
26. de Oliveira GR, Neto JF, de Camargo SM, Lucchetti ALG, Espinha DCM, Lucchetti G. Caregiving across the lifespan: comparing caregiver burden, mental health, and quality of life. *Psychogeriatrics* 2014 Dec 17;15(2):123-132. [doi: [10.1111/psyg.12087](https://doi.org/10.1111/psyg.12087)] [Medline: [25521215](https://pubmed.ncbi.nlm.nih.gov/25521215/)]
27. Schulz R, Beach SR. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. *JAMA* 1999 Dec 15;282(23):2215-2219. [Medline: [10605972](https://pubmed.ncbi.nlm.nih.gov/10605972/)]
28. Bennett J, Fagundes C, Kiecolt-Glaser J. The chronic stress of caregiving accelerates the natural aging of the immune system. In: Bosch JA, Phillips AC, Lord JM, editors. *Immunosenescence*. Berlin, Germany: Springer; 2013:35-46.
29. Buyck J, Ankri J, Dugravot A, Bonnaud S, Nabi H, Kivimäki M, et al. Informal caregiving and the risk for coronary heart disease: the Whitehall II study. *J Gerontol A Biol Sci Med Sci* 2013 Oct;68(10):1316-1323 [FREE Full text] [doi: [10.1093/gerona/glt025](https://doi.org/10.1093/gerona/glt025)] [Medline: [23525476](https://pubmed.ncbi.nlm.nih.gov/23525476/)]
30. Shaffer KM, Kim Y, Carver CS. Physical and mental health trajectories of cancer patients and caregivers across the year post-diagnosis: a dyadic investigation. *Psychol Health* 2016 Feb 09;31(6):655-674. [doi: [10.1080/08870446.2015.1131826](https://doi.org/10.1080/08870446.2015.1131826)] [Medline: [26680247](https://pubmed.ncbi.nlm.nih.gov/26680247/)]
31. Segrin C, Badger TA, Meek P, Lopez AM, Bonham E, Sieger A. Dyadic interdependence on affect and quality-of-life trajectories among women with breast cancer and their partners. *J Soc Pers Relat* 2016 Jun 30;22(5):673-689. [doi: [10.1177/0265407505056443](https://doi.org/10.1177/0265407505056443)] [Medline: [22973420](https://pubmed.ncbi.nlm.nih.gov/22973420/)]
32. Segrin C, Badger T, Dorros SM, Meek P, Lopez AM. Interdependent anxiety and psychological distress in women with breast cancer and their partners. *Psycho-Oncology* 2007 Jul;16(7):634-643. [doi: [10.1002/pon.1111](https://doi.org/10.1002/pon.1111)] [Medline: [17094160](https://pubmed.ncbi.nlm.nih.gov/17094160/)]
33. Kim Y, Carver CS, Shaffer KM, Gansler T, Cannady RS. Cancer caregiving predicts physical impairments: roles of earlier caregiving stress and being a spousal caregiver. *Cancer* 2015 Jan 15;121(2):302-310. [doi: [10.1002/cncr.29040](https://doi.org/10.1002/cncr.29040)] [Medline: [25209592](https://pubmed.ncbi.nlm.nih.gov/25209592/)]
34. 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](https://doi.org/10.1007/s13187-013-0472-2)] [Medline: [23553000](https://pubmed.ncbi.nlm.nih.gov/23553000/)]
35. Adams E, Boulton M, Watson E. The information needs of partners and family members of cancer patients: a systematic literature review. *Patient Educ Couns* 2009 Nov;77(2):179-186. [doi: [10.1016/j.pec.2009.03.027](https://doi.org/10.1016/j.pec.2009.03.027)] [Medline: [19406609](https://pubmed.ncbi.nlm.nih.gov/19406609/)]
36. 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](https://doi.org/10.1002/cncr.29223)] [Medline: [25677095](https://pubmed.ncbi.nlm.nih.gov/25677095/)]
37. James N, Daniels H, Rahman R, McConkey C, Derry J, Young A. A study of information seeking by cancer patients and their carers. *Clin Oncol (R Coll Radiol)* 2007 Jun;19(5):356-362. [doi: [10.1016/j.clon.2007.02.005](https://doi.org/10.1016/j.clon.2007.02.005)] [Medline: [17399963](https://pubmed.ncbi.nlm.nih.gov/17399963/)]
38. Kernisan LP, Sudore RL, Knight SJ. Information-seeking at a caregiving website: a qualitative analysis. *J Med Internet Res* 2010;12(3):e31 [FREE Full text] [doi: [10.2196/jmir.1548](https://doi.org/10.2196/jmir.1548)] [Medline: [20675292](https://pubmed.ncbi.nlm.nih.gov/20675292/)]
39. Chi N, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. *J Telemed Telecare* 2015 Jan;21(1):37-44 [FREE Full text] [doi: [10.1177/1357633X14562734](https://doi.org/10.1177/1357633X14562734)] [Medline: [25475220](https://pubmed.ncbi.nlm.nih.gov/25475220/)]
40. Applebaum AJ, Breitbart W. Care for the cancer caregiver: a systematic review. *Palliat Support Care* 2013 Jun;11(3):231-252. [doi: [10.1017/S1478951512000594](https://doi.org/10.1017/S1478951512000594)] [Medline: [23046977](https://pubmed.ncbi.nlm.nih.gov/23046977/)]
41. Slev VN, Mistiaen P, Pasma HRW, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. *Int J Med Inform* 2016 Mar;87:54-67. [doi: [10.1016/j.jimedinf.2015.12.013](https://doi.org/10.1016/j.jimedinf.2015.12.013)] [Medline: [26806712](https://pubmed.ncbi.nlm.nih.gov/26806712/)]
42. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. *J Med Internet Res* 2013;15(12):e287 [FREE Full text] [doi: [10.2196/jmir.2661](https://doi.org/10.2196/jmir.2661)] [Medline: [24366061](https://pubmed.ncbi.nlm.nih.gov/24366061/)]
43. Brouard B, Bardo P, Bonnet C, Mounier N, Vignot M, Vignot S. Mobile applications in oncology: is it possible for patients and healthcare professionals to easily identify relevant tools? *Ann Med* 2016 Jun 27:1-7. [doi: [10.1080/07853890.2016.1195010](https://doi.org/10.1080/07853890.2016.1195010)] [Medline: [27348761](https://pubmed.ncbi.nlm.nih.gov/27348761/)]
44. Raingruber B. The effectiveness of psychosocial interventions with cancer patients: an integrative review of the literature (2006-2011). *ISRN Nurs* 2011;2011:638218 [FREE Full text] [doi: [10.5402/2011/638218](https://doi.org/10.5402/2011/638218)] [Medline: [22191052](https://pubmed.ncbi.nlm.nih.gov/22191052/)]

45. Kaltenbaugh DJ, Klem ML, Hu L, Turi E, Haines AJ, Hagerty LJ. Using Web-based interventions to support caregivers of patients with cancer: a systematic review. *Oncol Nurs Forum* 2015 Mar;42(2):156-164. [doi: [10.1188/15.ONF.156-164](https://doi.org/10.1188/15.ONF.156-164)] [Medline: [25806882](#)]
46. Wasilewski MB, Stinson JN, Cameron JJ. Web-based health interventions for family caregivers of elderly individuals: a scoping review. *Int J Med Inform* 2017 Jul;103:109-138. [doi: [10.1016/j.ijmedinf.2017.04.009](https://doi.org/10.1016/j.ijmedinf.2017.04.009)] [Medline: [28550996](#)]
47. Centre for Evidence-based Medicine. Levels of evidence (March 2009). Oxford, UK: CEBM; 2018. URL: <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/> [accessed 2018-06-06] [WebCite Cache ID 6zyZJvN4V]
48. Stern A, Valaitis R, Weir R, Jadad AR. Use of home telehealth in palliative cancer care: a case study. *J Telemed Telecare* 2012 Jul;18(5):297-300. [doi: [10.1258/jtt.2012.111201](https://doi.org/10.1258/jtt.2012.111201)] [Medline: [22790013](#)]
49. Kinney AY, Boonyasiriwat W, Walters ST, Pappas LM, Stroup AM, Schwartz MD, et al. Telehealth personalized cancer risk communication to motivate colonoscopy in relatives of patients with colorectal cancer: the family CARE randomized controlled trial. *J Clin Oncol* 2014 Mar 01;32(7):654-662 [FREE Full text] [doi: [10.1200/JCO.2013.51.6765](https://doi.org/10.1200/JCO.2013.51.6765)] [Medline: [24449229](#)]
50. Scott K, Beatty L. Feasibility study of a self-guided cognitive behaviour therapy internet intervention for cancer carers. *Aust J Prim Health* 2013;19(4):270-274. [doi: [10.1071/PY13025](https://doi.org/10.1071/PY13025)] [Medline: [23796231](#)]
51. DuBenske LL, Gustafson DH, Namkoong K, Hawkins RP, Atwood AK, Brown RL, et al. CHESS improves cancer caregivers' burden and mood: results of an eHealth RCT. *Health Psychol* 2014 Oct;33(10):1261-1272 [FREE Full text] [doi: [10.1037/a0034216](https://doi.org/10.1037/a0034216)] [Medline: [24245838](#)]
52. Northouse LL, Mood DW, Schafenacker A, Montie JE, Sandler HM, Forman JD, et al. Randomized clinical trial of a family intervention for prostate cancer patients and their spouses. *Cancer* 2007 Dec 15;110(12):2809-2818 [FREE Full text] [doi: [10.1002/cncr.23114](https://doi.org/10.1002/cncr.23114)] [Medline: [17999405](#)]
53. Schover LR, Canada AL, Yuan Y, Sui D, Neese L, Jenkins R, et al. A randomized trial of internet-based versus traditional sexual counseling for couples after localized prostate cancer treatment. *Cancer* 2012 Jan 15;118(2):500-509 [FREE Full text] [doi: [10.1002/cncr.26308](https://doi.org/10.1002/cncr.26308)] [Medline: [21953578](#)]
54. Northouse L, Kershaw T, Mood D, Schafenacker A. Effects of a family intervention on the quality of life of women with recurrent breast cancer and their family caregivers. *Psychooncology* 2005 Jun;14(6):478-491. [doi: [10.1002/pon.871](https://doi.org/10.1002/pon.871)] [Medline: [15599947](#)]
55. Chih M, DuBenske LL, Hawkins RP, Brown RL, Dinuer SK, Cleary JF, et al. Communicating advanced cancer patients' symptoms via the Internet: a pooled analysis of two randomized trials examining caregiver preparedness, physical burden, and negative mood. *Palliat Med* 2013 Jun;27(6):533-543 [FREE Full text] [doi: [10.1177/0269216312457213](https://doi.org/10.1177/0269216312457213)] [Medline: [22988042](#)]
56. Dionne-Odom JN, Azuero A, Lyons KD, Hull JG, Tosteson T, Li Z, et al. Benefits of early versus delayed palliative care to informal family caregivers of patients with advanced cancer: outcomes from the ENABLE III randomized controlled trial. *J Clin Oncol* 2015 May 01;33(13):1446-1452 [FREE Full text] [doi: [10.1200/JCO.2014.58.7824](https://doi.org/10.1200/JCO.2014.58.7824)] [Medline: [25800762](#)]
57. Dionne-Odom JN, Azuero A, Lyons KD, Hull JG, Prescott AT, Tosteson T, et al. Family caregiver depressive symptom and grief outcomes from the ENABLE III randomized controlled trial. *J Pain Symptom Manage* 2016 Sep;52(3):378-385 [FREE Full text] [doi: [10.1016/j.jpainsymman.2016.03.014](https://doi.org/10.1016/j.jpainsymman.2016.03.014)] [Medline: [27265814](#)]
58. Namkoong K, DuBenske LL, Shaw BR, Gustafson DH, Hawkins RP, Shah DV, et al. Creating a bond between caregivers online: effect on caregivers' coping strategies. *J Health Commun* 2012;17(2):125-140 [FREE Full text] [doi: [10.1080/10810730.2011.585687](https://doi.org/10.1080/10810730.2011.585687)] [Medline: [22004055](#)]
59. Northouse LL, Mood DW, Schafenacker A, Kalemkerian G, Zalupski M, LoRusso P, et al. Randomized clinical trial of a brief and extensive dyadic intervention for advanced cancer patients and their family caregivers. *Psychooncology* 2013 Mar;22(3):555-563 [FREE Full text] [doi: [10.1002/pon.3036](https://doi.org/10.1002/pon.3036)] [Medline: [22290823](#)]
60. Badger T, Segrin C, Pasvogel A, Lopez AM. The effect of psychosocial interventions delivered by telephone and videophone on quality of life in early-stage breast cancer survivors and their supportive partners. *J Telemed Telecare* 2013 Jul;19(5):260-265. [doi: [10.1177/1357633X13492289](https://doi.org/10.1177/1357633X13492289)] [Medline: [24163235](#)]
61. Porter LS, Keefe FJ, Garst J, Baucom DH, McBride CM, McKee DC, et al. Caregiver-assisted coping skills training for lung cancer: results of a randomized clinical trial. *J Pain Symptom Manage* 2011 Jan;41(1):1-13 [FREE Full text] [doi: [10.1016/j.jpainsymman.2010.04.014](https://doi.org/10.1016/j.jpainsymman.2010.04.014)] [Medline: [20832982](#)]
62. Shaw JM, Young JM, Butow PN, Badgery-Parker T, Durcinoska I, Harrison JD, et al. Improving psychosocial outcomes for caregivers of people with poor prognosis gastrointestinal cancers: a randomized controlled trial (Family Connect). *Support Care Cancer* 2016 Feb;24(2):585-595. [doi: [10.1007/s00520-015-2817-3](https://doi.org/10.1007/s00520-015-2817-3)] [Medline: [26111955](#)]
63. Högberg K, Sandman L, Nyström M, Stockelberg D, Broström A. Prerequisites required for the provision and use of web-based communication for psychosocial support in haematologic care. *Eur J Oncol Nurs* 2013 Oct;17(5):596-602. [doi: [10.1016/j.ejon.2013.01.005](https://doi.org/10.1016/j.ejon.2013.01.005)] [Medline: [23453568](#)]
64. Badger TA, Segrin C, Figueredo AJ, Harrington J, Sheppard K, Passalacqua S, et al. Psychosocial interventions to improve quality of life in prostate cancer survivors and their intimate or family partners. *Qual Life Res* 2011 Aug;20(6):833-844 [FREE Full text] [doi: [10.1007/s11136-010-9822-2](https://doi.org/10.1007/s11136-010-9822-2)] [Medline: [21170682](#)]

65. Zulman DM, Schafenacker A, Barr KLC, Moore IT, Fisher J, McCurdy K, et al. Adapting an in-person patient-caregiver communication intervention to a tailored web-based format. *Psychooncology* 2012 Mar;21(3):336-341 [[FREE Full text](#)] [doi: [10.1002/pon.1900](#)] [Medline: [21830255](#)]
66. Farnham S, Cheng L, Stone L, Zaner-Godsey M, Hibbeln C, Syrjala K, et al. HutchWorld: clinical study of computer-mediated social support for cancer patients and their caregivers. 2002 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 20-25, 2002; Minneapolis, MN, USA p. 375-382. [doi: [10.1145/503376.503444](#)]
67. Clark MM, Rummans TA, Atherton PJ, Cheville AL, Johnson ME, Frost MH, et al. Randomized controlled trial of maintaining quality of life during radiotherapy for advanced cancer. *Cancer* 2013 Feb 15;119(4):880-887 [[FREE Full text](#)] [doi: [10.1002/cncr.27776](#)] [Medline: [22930253](#)]
68. Northouse L, Schafenacker A, Barr KLC, Katapodi M, Yoon H, Brittain K, et al. A tailored Web-based psychoeducational intervention for cancer patients and their family caregivers. *Cancer Nurs* 2014;37(5):321-330 [[FREE Full text](#)] [doi: [10.1097/NCC.0000000000000159](#)] [Medline: [24945270](#)]
69. Mosher CE, Winger JG, Hanna N, Jalal SI, Einhorn LH, Birdas TJ, et al. Randomized pilot trial of a telephone symptom management intervention for symptomatic lung cancer patients and their family caregivers. *J Pain Symptom Manage* 2016 Oct;52(4):469-482 [[FREE Full text](#)] [doi: [10.1016/j.jpainsymman.2016.04.006](#)] [Medline: [27401514](#)]
70. Badger TA, Segrin C, Hepworth JT, Pasvogel A, Weihs K, Lopez AM. Telephone-delivered health education and interpersonal counseling improve quality of life for Latinas with breast cancer and their supportive partners. *Psychooncology* 2013 May;22(5):1035-1042. [doi: [10.1002/pon.3101](#)] [Medline: [22573418](#)]
71. Chambers SK, Girgis A, Occhipinti S, Hutchison S, Turner J, McDowell M, et al. A randomized trial comparing two low-intensity psychological interventions for distressed patients with cancer and their caregivers. *Oncol Nurs Forum* 2014 Jul 01;41(4):E256-E266. [doi: [10.1188/14.ONF.E256-E266](#)] [Medline: [24969260](#)]
72. 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](#)]
73. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. *J Med Internet Res* 2004 Nov 10;6(4):e40 [[FREE Full text](#)] [doi: [10.2196/jmir.6.4.e40](#)] [Medline: [15631964](#)]
74. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. *World Psychiatry* 2014 Feb;13(1):4-11 [[FREE Full text](#)] [doi: [10.1002/wps.20083](#)] [Medline: [24497236](#)]
75. 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](#)]
76. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? *Expert Rev Pharmacoecon Outcomes Res* 2007 Jun;7(3):291-297. [doi: [10.1586/14737167.7.3.291](#)] [Medline: [20528315](#)]
77. Mozer E, Franklin B, Rose J. Psychotherapeutic intervention by telephone. *Clin Interv Aging* 2008;3(2):391-396 [[FREE Full text](#)] [Medline: [18686761](#)]
78. 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](#)]
79. Agboola SO, Ju W, Elfiky A, Kvedar JC, Jethwani K. The effect of technology-based interventions on pain, depression, and quality of life in patients with cancer: a systematic review of randomized controlled trials. *J Med Internet Res* 2015;17(3):e65 [[FREE Full text](#)] [doi: [10.2196/jmir.4009](#)] [Medline: [25793945](#)]
80. Ventura F, Ohlén J, Koinberg I. An integrative review of supportive e-health programs in cancer care. *Eur J Oncol Nurs* 2013 Aug;17(4):498-507. [doi: [10.1016/j.ejon.2012.10.007](#)] [Medline: [23158437](#)]
81. Campbell LC, Keefe FJ, McKee DC, Edwards CL, Herman SH, Johnson LE, et al. Prostate cancer in African Americans: relationship of patient and partner self-efficacy to quality of life. *J Pain Symptom Manage* 2004 Nov;28(5):433-444. [doi: [10.1016/j.jpainsymman.2004.02.020](#)] [Medline: [15504620](#)]
82. Porter LS, Keefe FJ, Garst J, McBride CM, Baucom D. Self-efficacy for managing pain, symptoms, and function in patients with lung cancer and their informal caregivers: associations with symptoms and distress. *Pain* 2008;137(2):306-315. [doi: [10.1016/J.PAIN.2007.09.010](#)]
83. Weinert K, Koenig V, Brunner M, Martin R. Concept maps: a useful and usable tool for computer-based knowledge assessment? A literature review with a focus on usability. *Comput Educ* 2014 Sep;78:201-209. [doi: [10.1016/J.COMPEDU.2014.06.002](#)]
84. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. *J Am Med Inform Assoc* 2006;13(6):660-667 [[FREE Full text](#)] [doi: [10.1197/jamia.M2146](#)] [Medline: [16929036](#)]

85. Northouse LL, Katapodi MC, Song L, Zhang L, Mood DW. Interventions with family caregivers of cancer patients: meta-analysis of randomized trials. *CA Cancer J Clin* 2010;60(5):317-339 [[FREE Full text](#)] [doi: [10.3322/caac.20081](#)] [Medline: [20709946](#)]
86. Kent EE, Rowland JH, Northouse L, Litzelman K, Chou WS, Shelburne N, et al. Caring for caregivers and patients: research and clinical priorities for informal cancer caregiving. *Cancer* 2016 Dec 01;122(13):1987-1995 [[FREE Full text](#)] [doi: [10.1002/cncr.29939](#)] [Medline: [26991807](#)]
87. Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. *JAMA* 2014 Mar 12;311(10):1052-1060. [doi: [10.1001/jama.2014.304](#)] [Medline: [24618967](#)]
88. Pottie CG, Burch KA, Thomas LPM, Irwin SA. Informal caregiving of hospice patients. *J Palliat Med* 2014 Jul;17(7):845-856 [[FREE Full text](#)] [doi: [10.1089/jpm.2013.0196](#)] [Medline: [24992371](#)]
89. Halpern MT, Fiero MH, Bell ML. Impact of caregiver activities and social supports on multidimensional caregiver burden: analyses from nationally-representative surveys of cancer patients and their caregivers. *Qual Life Res* 2017 Feb 16;26(6):1587-1595. [doi: [10.1007/s11136-017-1505-9](#)] [Medline: [28210993](#)]
90. Wittmann D, Carolan M, Given B, Skolarus TA, An L, Palapattu G, et al. Exploring the role of the partner in couples' sexual recovery after surgery for prostate cancer. *Support Care Cancer* 2014 Sep;22(9):2509-2515. [doi: [10.1007/s00520-014-2244-x](#)] [Medline: [24728619](#)]
91. Bobridge A, Bond MJ, Marshall V, Paterson J. An investigation of the support needs of men and partners throughout the prostate cancer journey. *Psychooncology* 2015 Mar;24(3):341-347. [doi: [10.1002/pon.3655](#)] [Medline: [25116753](#)]
92. Harden JK, Sanda MG, Wei JT, Yarandi H, Hembroff L, Hardy J, PROSTQA Consortium Study Group. Partners' long-term appraisal of their caregiving experience, marital satisfaction, sexual satisfaction, and quality of life 2 years after prostate cancer treatment. *Cancer Nurs* 2013;36(2):104-113 [[FREE Full text](#)] [doi: [10.1097/NCC.0b013e3182567c03](#)] [Medline: [22728952](#)]
93. Cheng T, Jackman M, McQuestion M, Fitch M. 'Knowledge is power': perceived needs and preferred services of male partners of women newly diagnosed with breast cancer. *Support Care Cancer* 2014 Dec;22(12):3175-3183. [doi: [10.1007/s00520-014-2314-0](#)] [Medline: [24957004](#)]
94. Schubart JR, Kinzie MB, Farace E. Caring for the brain tumor patient: family caregiver burden and unmet needs. *Neuro Oncol* 2008 Feb;10(1):61-72 [[FREE Full text](#)] [doi: [10.1215/15228517-2007-040](#)] [Medline: [17993635](#)]
95. Lau DT, Berman R, Halpern L, Pickard AS, Schrauf R, Witt W. Exploring factors that influence informal caregiving in medication management for home hospice patients. *J Palliat Med* 2010 Sep;13(9):1085-1090 [[FREE Full text](#)] [doi: [10.1089/jpm.2010.0082](#)] [Medline: [20836633](#)]
96. Cutica I, Vie GM, Pravettoni G. Personalised medicine: the cognitive side of patients. *Eur J Intern Med* 2014 Oct;25(8):685-688. [doi: [10.1016/J.EJIM.2014.07.002](#)] [Medline: [25060407](#)]
97. Eysenbach G. What is e-health? *J Med Internet Res* 2001 Jun;3(2):E20. [doi: [10.2196/jmir.3.2.e20](#)] [Medline: [11720962](#)]
98. Gordon LG, Beesley VL, Scuffham PA. Evidence on the economic value of psychosocial interventions to alleviate anxiety and depression among cancer survivors: a systematic review. *Asia Pac J Clin Oncol* 2011 Jun;7(2):96-105. [doi: [10.1111/j.1743-7563.2011.01395.x](#)] [Medline: [21585688](#)]
99. Elbert NJ, van Os-Medendorp H, van Renselaar W, 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](#)]
100. Bergmo TS. How to measure costs and benefits of ehealth interventions: an overview of methods and frameworks. *J Med Internet Res* 2015;17(11):e254 [[FREE Full text](#)] [doi: [10.2196/jmir.4521](#)] [Medline: [26552360](#)]

Abbreviations

EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Core Questionnaire
FACT-G: Functional Assessment of Cancer Therapy-General
ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 10.01.18; peer-reviewed by M Falahee, J Kvedar, U Qamar; comments to author 23.03.18; revised version received 27.04.18; accepted 15.05.18; published 18.06.18.

Please cite as:

Marzorati C, Renzi C, Russell-Edu SW, Pravettoni G
Telemedicine Use Among Caregivers of Cancer Patients: Systematic Review
J Med Internet Res 2018;20(6):e223
URL: <http://www.jmir.org/2018/6/e223/>
doi: [10.2196/jmir.9812](https://doi.org/10.2196/jmir.9812)
PMID: [29914858](https://pubmed.ncbi.nlm.nih.gov/29914858/)

©Chiara Marzorati, Chiara Renzi, Samuel William Russell-Edu, Gabriella Pravettoni. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 18.06.2018. 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.

Original Paper

A Teledermatology Scale-Up Framework and Roadmap for Sustainable Scaling: Evidence-Based Development

Laticha Elizabeth Marolana Walters^{1,2}, BSc, MBL; Richard Ernest Scott^{1,3,4}, BSc, PhD; Maurice Mars¹, MBChB, MD

¹Department of TeleHealth, University of KwaZulu-Natal, Durban, South Africa

²Meraka Institute, Council for Scientific and Industrial Research, Pretoria, South Africa

³Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada

⁴NT Consulting – Global e-Health Inc, Calgary, AB, Canada

Corresponding Author:

Maurice Mars, MBChB, MD

Department of TeleHealth

University of KwaZulu-Natal

719 Umbilo Road, Congella

Durban, 4013

South Africa

Phone: 27 (31) 2604543

Email: mars@ukzn.ac.za

Abstract

Background: The objectives of South Africa's electronic health (eHealth) strategy recognize the value proposition that telemedicine practices hold for rural and urban referrals, but a lack of accepted and formalized scale-up has impeded realization of benefits. While both synchronous and asynchronous teledermatology exist, these remain localized and not scaled-up. Skin pathology is often the first sign of an HIV/AIDS infection, which remains a major cause of morbidity and mortality in South Africa. It is essential to replace the current inefficient dermatology referral process with a swift, organized, and efficacious one.

Objective: The objective of this study is to present an evidenced-based teledermatology scale-up framework (TDSF) and implementation roadmap (TDSF-IR).

Methods: A qualitative method with a design science research process model was used which consisted of 5 phases: (1) Awareness, which confirmed the need for an evidence-based TDSF and supporting TDSF-IR; (2) Suggestion, where a proposal was delivered on how to develop a TDSF and TDSF-IR; (3) Development, where we identified recommended design requirements and used these to identify and critique existing teledermatology or related scale-up frameworks; (4) Evaluation and validation, where we assessed outputs of the development phase against the design requirements and validated by confirming the veracity of the TDSF and TDSF-IR (validation involved 4 key senior teledermatology stakeholders using a questionnaire with a 5-point Likert scale); and (5) Conclusion, where validation results were used to finalize and communicate the TDSF and TDSF-IR to users.

Results: The study identified 5 TDSF components: eHealth building blocks, eHealth strategic objectives and budget, scale-up continuum periods, scale-up drivers, and scale-up phases. In addition, 36 subcomponents were identified. Each was further characterized and described to enable design of the final evidence-based TDSF. An implementation roadmap (TDSF-IR) was also prepared as a guide for an implementer with step-by-step instructions for application of the TDSF. For the validation study of the TDSF and supporting TDSF-IR, 4 purposively selected key senior teledermatology management stakeholders were asked if they found it useful as a guide to assist the South African public health system with teledermatology scale-up. The mean (SD) of Likert-scale rating was 4.0 (0.53) where 4=Agree and 33 of 36 responses were either agree or strongly agree.

Conclusions: This study developed a TDSF and supporting roadmap (TDSF-IR) that are evidence-based. The proposed approach and described tools could be adapted to assist with ensuring scale-up and sustainability for other eHealth practices in other locations.

(*J Med Internet Res* 2018;20(6):e224) doi:[10.2196/jmir.9940](https://doi.org/10.2196/jmir.9940)

KEYWORDS

tele dermatology; scale-up; Tele dermatology Scale-up Framework; TDSF; Tele dermatology Scale-up Framework Implementation Roadmap; TDSF-IR; design science research; KwaZulu-Natal; South Africa

Introduction

The objectives of the eHealth strategy in South Africa recognize the value proposition of telemedicine to address the shortage of specialists in rural hospitals and to improve access to health care [1]. Tele dermatology, due to its visual nature, is one of the most common uses of telemedicine [2] and has previously been found to be effective in enhancing access to dermatologists [3]. Similar to other developing countries, South Africa would benefit from sustainable scale-up of its existing tele dermatology activities and services, given the high prevalence of significant skin lesions in HIV/AIDS [4,5] and the shortage of dermatologists [6].

Scale-up and sustainability of telemedicine initiatives are long standing issues, and the recent Momentum document on implementing successful telemedicine programs describes these issues [7]. A recent review of tele dermatology activities in South Africa has documented both asynchronous and synchronous tele dermatology services, some of which have run for over 10 years, but have not yet been scaled up [8]. These circumstances are impeding the realization of the potential benefits of tele dermatology services [1,8,9], such as timely triage, diagnosis, and treatment initiation [10] of dermatological manifestations.

A previous study identified the minimum design requirements to inform a conceptual tele dermatology scale-up framework (TDSF) using key stakeholder interviews, literature review, program observations, and expert opinion [11]. The minimum requirements were grouped into 4 themes (framework organization, eHealth building blocks, eHealth planning, and eHealth action), which were further separated into 12 categories with 30 requirements [11]. Another study reported that no TDSF existed, and that no eHealth-related scale-up framework met all the design requirements [12].

Despite the recent launch of a tele dermatology toolkit [13] and adoption model [14], a gap remains for a conceptual framework supported with an implementation roadmap to assist public health systems with the process of sustainable scaling-up of successful pilots into routine health care. Therefore, there is a need to develop and validate a TDSF, and a roadmap (namely, the TDSF-IR) for its implementation, with measurable scale-up objectives which are aligned to public health system objectives. The framework and roadmap should be objectives realization management-driven, health sector aligned, holistic, and meet the defined TDSF design requirements.

For clarity, some terms need to be defined or described. In the context of this paper, a framework provides a network of interlinked concepts, assumptions, expectations, beliefs, and theories that, together, provide a comprehensive understanding

of a phenomenon or phenomena [15], and it lays out key factors, constructs, or variables and relationships among them [16]. Applied to scale-up, and in contrast to spontaneous adoption of innovations, a framework systematically guides the planning and implementing processes, leading to sustained practice [17]. The term scale-up has been defined by the World Health Organization as “deliberate efforts to increase the impact of successfully tested health innovations so as to benefit more people and to foster policy and program development on a lasting basis” [17].

A roadmap (in this study, the TDSF implementation roadmap or TDSF-IR) provides a structured method to guide use of the framework, with step-by-step instructions to ensure a logical flow of inputs and deliverables to achieve the framework’s objectives. Objectives realization management is a means to ensure that the intended top-down national eHealth goals and objectives are met whilst embracing bottom-up provincial strategic goals and objectives. An objectives realization management-driven approach ensures that the intended public eHealth sector goals and objectives are implemented and sustained by using an evidence-based TDSF and roadmap.

This paper is the culmination of studies looking at scale-up of tele dermatology in one of the nine Provinces of South Africa and may serve as a model for other developing world implementations. The paper proposes a TDSF and TDSF-IR and describes their final development, validation, and refinement.

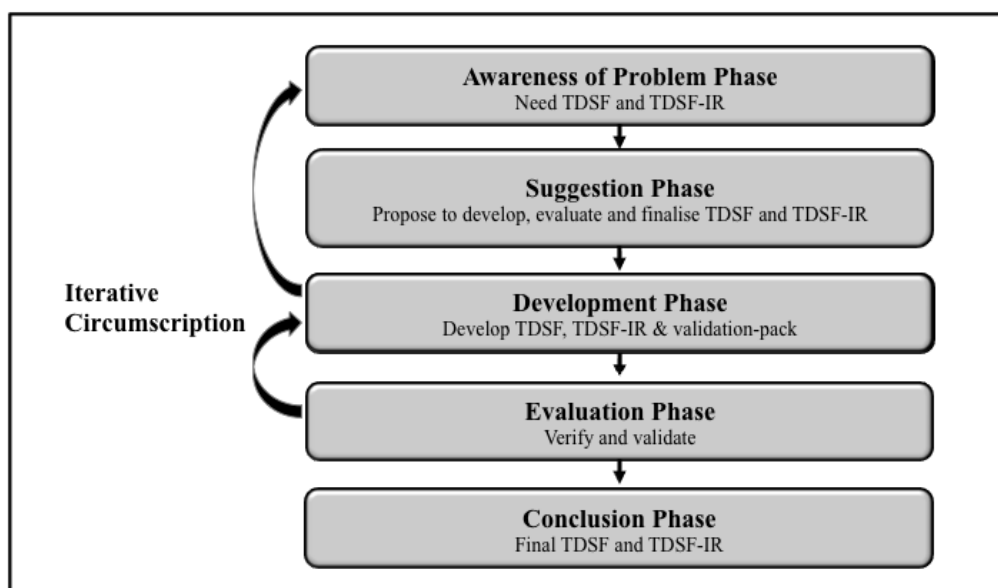
Methods

The design science research process model was adopted [18]. This model consists of 5 phases to solve the research problems and develop knowledge, namely Awareness, Suggestion, Development, Evaluation, and Conclusion (Figure 1). The process is both iterative and compounding, with output from prior phases providing input to succeeding phases, and each phase progressively informing the next, with the conclusion phase completing the process.

The “Awareness” and “Suggestion” phases were complimentary and identified and proposed approaches to resolve the research needs [18]. Evidence for these two phases was gathered through literature review and critique of existing scale-up frameworks. Tele dermatology stakeholder interviews, tele dermatology program observation, and expert opinion.

During the “Development Phase,” the proposed approaches were used to create specific solutions to meet the needs. This phase consisted of 3 major steps: design and development of a TDSF, design and development of a TDSF-IR, and development of a validation pack.

Figure 1. Overview of the design science research process model phases [18] used to meet the main need for an evidence-based teledermatology scale-up framework (TDSF) and implementation roadmap (TDSF-IR).



TDSF Design and Development was based on the previously established design requirements, which were interpreted and organized to define components and subcomponents for the TDSF. Interpretation involved unpacking the themes, categories, and requirement descriptions (including their reasoning and sources) to establish meaning. Thereafter they were organized by determining relationships and dependencies between the components to inform sequencing. In turn, TDSF-IR Design and Development involved creation of a tool to enable implementation of the TDSF. The TDSF-IR was developed by logically describing the organization and processes for implementing components and subcomponents of the TDSF, with estimated durations for completion. Circumscription [19] was used to channel the awareness of new constraints back into the “Awareness of Problem” phase.

The “Evaluation Phase” gave results from piloting the validation-pack with the international expert review panel, before applying the revised pack to key senior teledermatology stakeholders.

The validation-pack comprised a questionnaire and demonstration material to allow the utility of the TDSF and TDSF-IR to be assessed and presented. The questionnaire collected data about respondents’ experience in eHealth or telemedicine or teledermatology (Expert>5 years, Intermediate=2-5 years, Beginner≤2 years), their roles (Advisory, Implementer, User, Manager, Research), 9 validation statements using a 5-point Likert scale (5=Agree strongly, 4=Agree, 3=Unable to assess, 2=Disagree, 1=Disagree strongly), a comments field for every validation statement, and a general comments section at the end (see the Validation questionnaire in [Multimedia Appendix 1](#)). The demonstration material consisted of a slide show depicting the TDSF components and their organization during application of the TDSF.

Piloting of the validation-pack used 2 international eHealth experts. The experts were from Canada and Australia and are active in international eHealth research, implementation, and

academic fields. The pilot entailed using the validation pack (demonstration material and questionnaire) in one-to-one sessions (1 hour), followed by questions and answers. The experts completed and returned the questionnaires electronically. Based on the experts’ feedback, the validation-pack was revised.

The revised validation-pack was then presented to all 4 key senior teledermatology stakeholders who represented clinical and academic dermatology management and practice, as well as eHealth research. They were selected based on their current teledermatology management roles and their participation in the data collection phase. Except for the eHealth researcher, the key stakeholders represented the current teledermatology management team from the KwaZulu-Natal Department of Health and the Department of Dermatology at the University of KwaZulu-Natal. The management team had also participated in semistructured interviews that identified the TDSF design requirements from an earlier study [11].

Each step of the development phase was also evaluated to ensure the outputs met the associated need. Evaluation entailed verification and validation of the developed artifacts. To verify the design of the TDSF each design requirement was checked to confirm it was mapped to a TDSF component and subcomponent. To verify the TDSF-IR, the step-by-step instructions were reviewed to confirm correct mapping to TDSF action steps and sequencing of deliverables for implementation. Utility of the TDSF and TDSF-IR to assist the KwaZulu-Natal public health management with teledermatology scale-up was determined by using the validation-pack described above.

The final “Conclusion Phase” ensured the artifacts were consolidated into their final form (in accordance with feedback obtained from the evaluation phase), that all contributions were identified, and that the results were clearly communicated.

Results

Overview

The entire process from conception to completion is reflected in Figure 2. Steps numbered 1, 2, and 3 have been previously published [8,11,12] but are included in Figure 2 for completeness. Only results for Steps 4 onwards are described below, that is, those gained through the iterative and reflective process used within the context of the design science research process model and giving rise to specific content.

The results for the developed artifacts are presented in 3 subsections: TDSF, TDSF-IR, and validation of TDSF and TDSF-IR.

Tele dermatology Scale-Up Framework

The design requirements were interpreted to define 5 components (eHealth scale-up building blocks, eHealth strategic objectives and budget, scale-up continuum, scale-up drivers, and scale-up phases). In addition, 36 subcomponents were identified: 10 within “eHealth scale-up building blocks,” 4 within “Scale-up continuum,” 6 within “Scale-up drivers,” and 13 within “Scale-up activities,” embedded in the 3 “Scale-up phases” (Figure 3).

eHealth Scale-Up Building Blocks

Results of performing Step 4 of the Development Phase also included recognition of the need for 10 context specific eHealth scale-up building blocks to form a solid foundation for sustainable scaling. These building blocks are the presence of the following factors listed below:

1. National and provincial government operational objectives and budget
2. Political mandate and leadership
3. Legal and regulatory settings
4. Standards (eg, South Africa’s National Health Normative Standards for Interoperability)
5. Stakeholder management [20]
6. Public private partnerships (PPP)
7. eHealth Performance Indicators
8. eHealth Governance (ie, ICT and Health)
9. Architecture (eg, Health Patient Record System [1])
10. Project and program management capabilities

eHealth Strategic Objectives and Budget

The need for official politically— and financially—approved mandates is crucial for scale-up of proven pilot programs.

Figure 2. Overview of the complete design science research process model showing phases followed. The shaded areas (steps 1-3) form part of earlier published studies [8,11,12]. TD: Tele dermatology; TDSF: Tele dermatology Scale-Up Framework; TDSF-IR: Tele dermatology Scale-up Framework Implementation Roadmap.

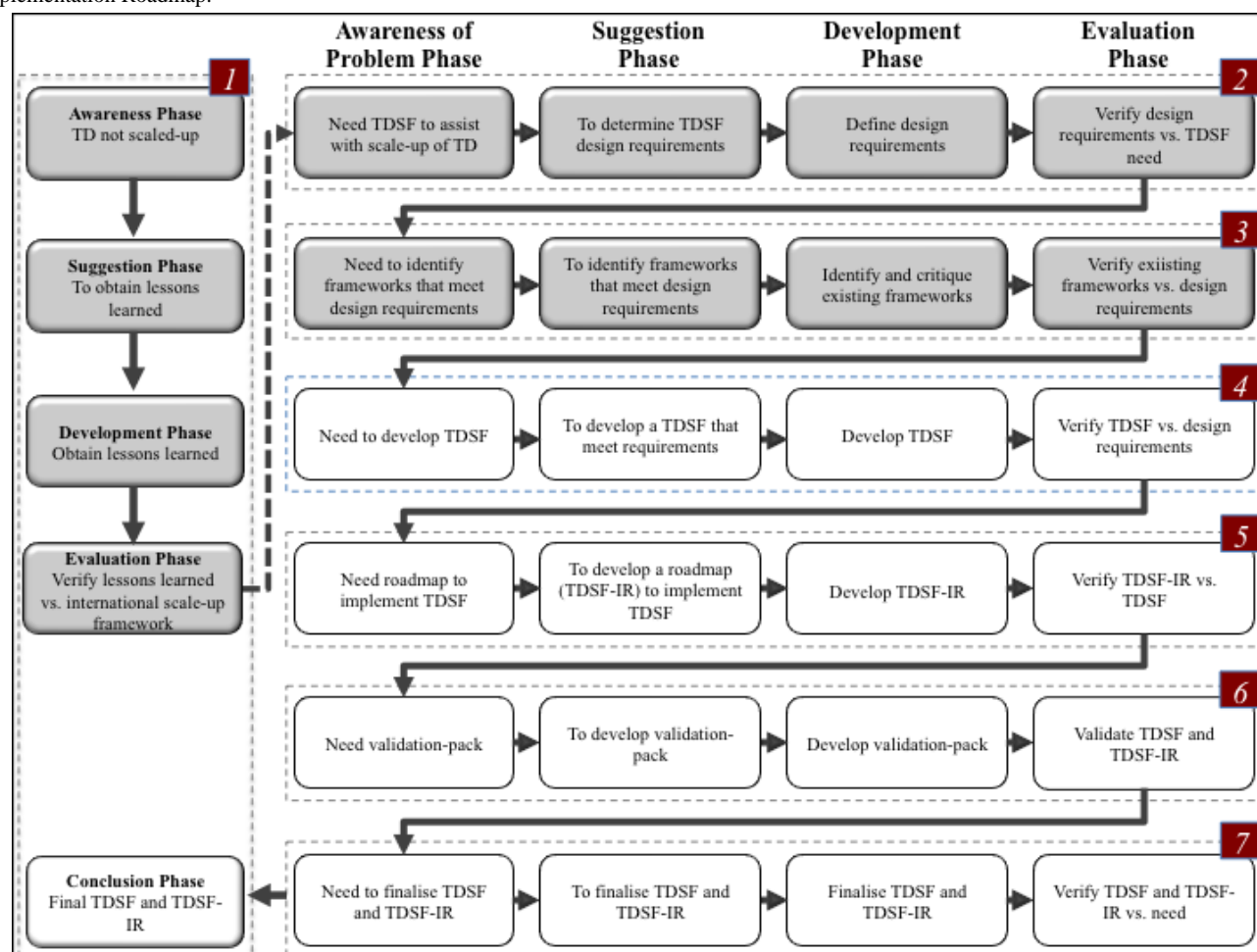
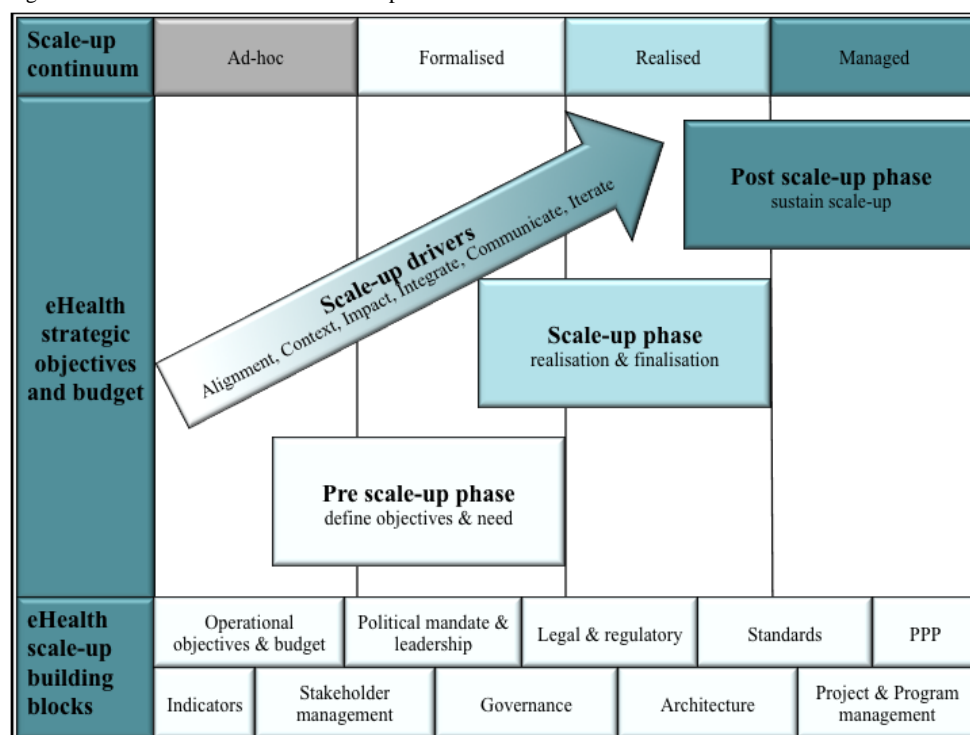


Figure 3. Illustration of how the design requirements fit within the conceptualized Teledermatology Scale-Up Framework (TDSF) and how the scale-up drivers assist with ensuring sustainable scale-up by realizing the teledermatology benefits at each scale-up phase along the continuum, all grounded by critical eHealth building blocks. PPP: Public-Private Partnerships.



Scale-Up Continuum

The continuum is comprised of 4 periods of evolution extending from “Ad-hoc” actions to “Formalized” actions where governance is influential, to a “Realized” period where specified objectives and benefits are seen, to ultimately the “Managed” period where the entire scale-up process and associated risks are managed (Table 1).

Scale-Up Drivers

Ongoing iterative and reflective analysis during Step 4 resulted in identification of 6 scale-up drivers to assist with ensuring sustainable scale-up. These were: intentional alignment, context sensitive, targeted integration, indicator impact, influential communication, and iterative process. These drivers are described in more detail in Table 2. While scale-up activities were related to each specific scale-up phase, the drivers also

coincidentally impacted all phases and assisted with moving the process along the scale-up continuum (Figure 3). For example, the pre-scale-up phase activity “Stakeholder management” also had activity associated with each driver (Table 2).

Scale-Up Phases

The interpretation process performed during TDSF design and development resulted in the inclusion of 3 scale-up phases (pre-scale-up, scale-up, and post-scale-up). Pre-scale-up activities focused on the definition of scale-up objectives and need activities, scale-up activities focused on scale-up realization and finalization, and post-scale-up activities focused on sustaining scale-up activities. This resulted in organization of the components and subcomponents according to prerequisites to ensure sequential phasing of scale-up activities.

Table 1. Description of the scale-up continuum.

Scale-up continuum periods	Description
Ad-hoc	Scale-up planning, implementation, management, and communication is happening on an as-needed basis and is not formally planned, approved and implemented as per TDSF ^a scale-up phases, activities and steps
Formalized	Scale-up governance processes and structures are planned, approved, communicated, monitored, and controlled as per TDSF pre-scale-up phase, activities and steps
Realized	Scale-up is formalized and functional; agreed health objectives and indicator benefits are realized; and risks are known and actively managed through implementation of plans, communication, monitoring and are controlled as per TDSF scale-up phase, activities, and steps
Managed	Scale-up is formalized, realized, communicated, monitored and controlled as per TDSF post scale-up phase, activities, and steps

^aTDSF: teledermatology scale-up framework.

Table 2. Scale-up drivers and their purpose, areas of application, and (using “Stakeholder management” as an example) relevant activity.

Drivers	Purpose	Areas	“Stakeholder Management” Activity
Intentional alignment	To intentionally support the overall goals and objectives of the health sector	Strategic and Tactical (Operational)	Identify and document the tele dermatology stakeholders for public and private sector
Context sensitive	To ensure that the proposed action is appropriate for the health care system and ICT ^a capabilities	Context	Determine the tele dermatology stakeholder requirements for key stakeholders (patients, health care system, ICT governance, and architecture)
Targeted integration	To ensure that proposed actions can leverage on existing eHealth ^b interventions	Alignment with existing initiatives	Identify and document existing eHealth stakeholders and assess tele dermatology value contribution opportunities to existing relationships
Indicator impact	To ensure that outcomes are measurable, recognised and aligned with health indicators	Sustainability, contribute to bottom line, economic, social and environment	Determine tele dermatology’s contribution to the need to increase access to equitable, effective, and efficient health care
Influential communication	To ensure that intent, progress, and contributions are communicated to the right people at the right time	Communicate to all levels of stakeholders at regular intervals	Regularly communicate with key stakeholders such as Department of Health and Health Professions Council the impact and status of tele dermatology and request feedback on enabling environment
Iterative process	To ensure that feedback is used to refine and enhance scale-up process along the continuum.	Continuous measure; refine feedback loops to encourage maturity	Assess scale-up status and take recommended action to progress in scale-up continuum

^aICT: information and communications technology.

^beHealth: electronic health.

Tele dermatology Scale-Up Framework Implementation Roadmap

The Step 5 Development Phase also resulted in formulation of the content for the TDSF-IR that supports TDSF implementation. The content of the TDSF-IR provides a step-by-step guide to ensure that implementation of the TDSF is executed in a logical sequence (Figure 4).

The sequence was the result of analyzing the dependencies of deliverables (outputs) of one TDSF scale-up activity to provide inputs to the subsequent activity. The 13 TDSF scale-up activities are:

1. Define scale-up need
2. Define scale-up stakeholders
3. Confirm scale-up compliance
4. Develop scale-up strategy
5. Develop detailed scale-up plans
6. Mobilize scale-up resources
7. Implement scale-up plans
8. Manage scale-up benefits
9. Manage scale-up risks
10. Confirm scale-up readiness
11. Finalize scale-up
12. Manage scale-up sustainability
13. Monitor and control scale-up

Within the TDSF-IR activities were mapped to an action plan with 8 fields: activity, inputs (what is needed to start the action), ownership (roles and responsibilities), action (what needs to be done), steps (how to get things done), rules and regulations (policies, standards, procedures and structures), timing

(estimated duration in years), and deliverables. The TDSF-IR ensures that the relationship and sequencing of TDSF activities support the implementation process and that deliverables of one activity feed into the next, for example TDSF activities 1, 2, and 3 in Figure 4 provide deliverables (tele dermatology scale-up business case, eHealth stakeholder map, and eHealth compliance register) that feed into activity 4.

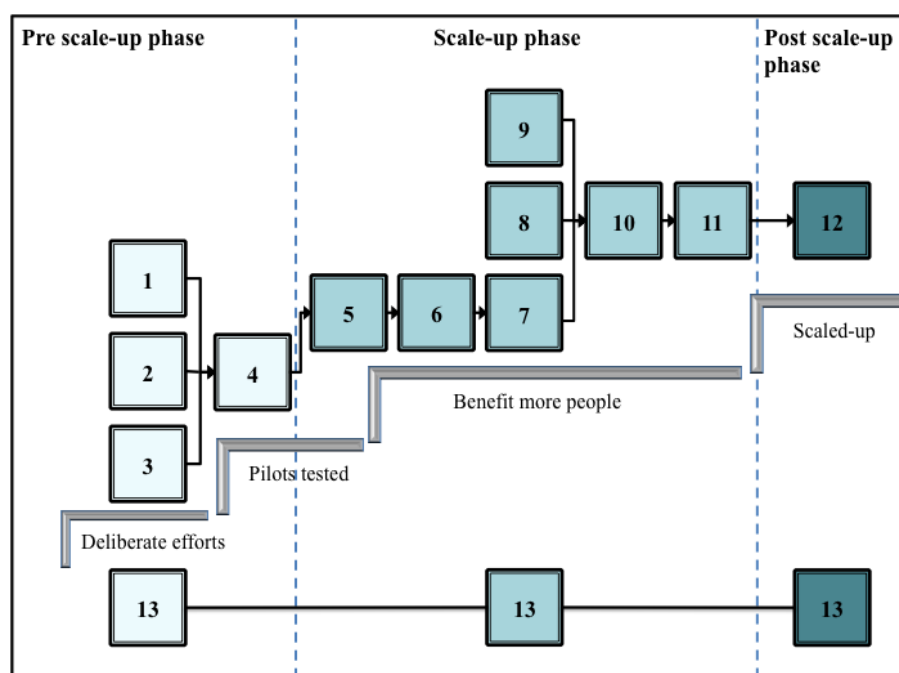
Validation of Tele dermatology Scale-Up Framework and Tele dermatology Scale-Up Framework Implementation Roadmap

The experts recommended the TDSF be more context sensitive to ensure that the needs of South Africa’s public health and key tele dermatology stakeholders are clearly addressed. This feedback from the experts was incorporated into the TDSF, and the demonstration process was revised to demonstrate the application of the TDSF to KwaZulu-Natal by using district specific health objectives and indicators. No other changes to the TDSF or TDSF-IR design were required.

The mean (SD) score for the 9 questions in the validation study was 4.0 (0.53) and 33 of 36 responses were either agree or strongly agree. One respondent did not feel that the TDSF could meet the requirements of the KwaZulu-Natal Department of Health, but stated that “I believe that the TDSF can meet the scale-up requirements of KwaZulu-Natal Department of Health” and that:

If properly implemented, TDSF would be able to guide such activities, but I am not sure that the framework on its own can do so. There is a whole eco-system that may impact on various aspect(s) of the framework’s realization

Figure 4. Outline of Teledermatology Scale-Up Framework Implementation Roadmap (TDSF-IR) depicting the relationships and sequencing of TDSF activities through the 3 phases from pre-scale-up to scale-up and post-scale-up. Numbers refer to the 13 TDSF scale-up activities: (1) Define scale-up need, (2) Define scale-up stakeholders, (3) Confirm scale-up compliance, (4) Develop scale-up strategy, (5) Develop detailed scale-up plans, (6) Mobilize scale-up resources, (7) Implement scale-up plans, (8) Manage scale-up benefits, (9) Manage scale-up risks, (10) Confirm scale-up readiness, (11) Finalize scale-up, (12) Manage scale-up sustainability, and (13) Monitor and control scale-up.



The eHealth researcher was unable to assess two questions and was uncertain whether the KwaZulu-Natal Department of Health management would understand the TDSF and whether the Department of Health would be able to leverage on existing eHealth equipment.

Some concerns were raised on the “need to provide more staff dedicated to run the telehealth aspect and training for the Medical Officers and Interns in the respective hospitals” (key stakeholder). It was considered that aspects of staffing are addressed under several components: the pre-scale-up phase (scale-up strategy [resources], eHealth governance [resources]), the scale-up phase (detail planning [change management], mobilization of resources), and the post-scale-up phase (operational plan; building blocks; operational objectives and budget), and no adjustment was made.

Discussion

Principal Findings

Using a design science research process model [18] this study addressed the need to use locally identified design requirements to develop an evidenced-based TDSF and supporting TDSF-IR. Each was centered on an objective realization management approach from needs definition through planning, implementation, and finalization, to sustaining scale-up.

The eHealth related scale-up frameworks identified earlier did not meet all the design requirements of the KwaZulu-Natal public health sector [12]. The American Academy of Dermatology [13] launched a toolkit that provides guidance and recommendations for implementers, and Ernst & Young developed a telemedicine adoption model [14,21]. However,

neither provided a conceptual framework nor supporting implementation roadmap to assist public health management to ensure sustainable scale-up of successful pilots into routine health care.

While limitations exist (eg, with so few people involved in teledermatology the sample size was restricted and the KwaZulu-Natal Department of Health lacks a telemedicine strategy limiting implementation possibilities), several aspects of the overall design are considered key. The eHealth scale-up building blocks form the foundation of the framework. The building blocks are a prerequisite for sustainable scaling [17] although a risk management approach [22] could be adopted for the absence of any one block. Also, the scale-up phases allow for a phased scaling compared to a big-bang approach where all the activities need to be completed in one implementation cycle. A phased approach to scaling is supported [23] as previously recommended for the TDSF [11]. The scale-up drivers provide the momentum and energy required to push the scaling process. The drivers are designed to guide the implementer through systematic scaling. Finally, the scale-up continuum periods are different from the typical maturity model approach with the intention that scale-up is a process with the goal of integrating teledermatology into routine practice in a sustainable manner. The Health Information System continuum [24] of the WHO compares well with the scale-up continuum although being more applicable for a wider eHealth strategic planning level.

The benefits that a scaled-up teledermatology service holds can now be realized through use of the TDSF and TDSF-IR. Furthermore, the uncommon approach to assist with ensuring sustainable teledermatology scale-up with objectives realization

management could potentially assist public health to realize the country's National Health Insurance goals towards achieving universal health coverage [25].

The proposed evidenced-based conceptual TDSF and supporting TDSF-IR could be considered for future eHealth scale-up framework development and scale-up implementation. The TDSF is based on realizing the objectives defined in strategic plans of South Africa's National and Provincial eHealth strategic plans (personal communication by Walters LEM, Mars M, Scott RE. An exploration of the use of benefits realisation management in tele dermatology related scale-up framework development; 2018).

Conclusions

A TDSF and TDSF-IR were developed, based on evidence obtained from key stakeholders, program observations, the literature, and the author's experience, and validated with eHealth management (clinical, academic, research, and general) that rated the TDSF and TDSF-IR as useful to assist the KwaZulu-Natal sector with sustainable tele dermatology scaling. These artifacts address a gap in published literature for an evidenced-based tele dermatology scale-up framework and supporting implementation roadmap.

The proposed approach and resultant TDSF and TDSF-IR could potentially be adapted to assist with ensuring sustainable scale-up for other eHealth practices in other locations. Future studies could implement the TDSF and TDSF-IR in the real world and use feedback to refine the artifacts.

Acknowledgments

Research reported in this publication acknowledges the contributions of KwaZulu-Natal Department of Health, University of KwaZulu-Natal Department of Dermatology, and the Council for Scientific and Industrial Research, especially, Professor N Dlova, Professor A Mosam, Professor P Kotze, and Dr R Eashwari for their and their team's contributions. The research was supported by the Fogarty International Center of the National Institutes of Health under Award Number D43TW007004 and the Council for Scientific and Industrial Research, Meraka Institute, Pretoria, South Africa. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or Fogarty International or Council for Scientific and Industrial Research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Validation questionnaire

[PDF File (Adobe PDF File), 56KB - [jmir_v20i6e224_app1.pdf](#)]

References

1. South Africa National Department of Health. National Department of Health. Pretoria, South Africa: National Department of Health; 2012. National eHealth strategy South Africa 2012/13-2016/17 URL: <http://www.health.gov.za/index.php/2014-08-15-12-54-26/category/95-2012s?download=171:ehealth-strategy-south-africa-2012-2016> [accessed 2018-04-02] [WebCite Cache ID 6yNgQOVEY]
2. American Telemedicine Association. 2018. Tele dermatology SIG URL: <http://www.americantelemed.org/main/membership/ata-members/ata-sigs/tele dermatology-sig> [accessed 2018-02-26] [WebCite Cache ID 6xWKyp3Zu]
3. Bashshur R, Shannon G, Tejasvi T, Kvedar JC, Gates M. The Empirical Foundations of Tele dermatology: A Review of the Research Evidence. *Telemed J E Health* 2015 Dec;21(12):953-979 [FREE Full text] [doi: [10.1089/tmj.2015.0146](https://doi.org/10.1089/tmj.2015.0146)] [Medline: [26394022](#)]
4. South Africa National Department of Health. Pretoria, South Africa: National Department of Health; 2016. Annual report 2015/16 URL: https://www.gov.za/sites/default/files/dohannual%20report%20%202016%20reduced_a.pdf [accessed 2018-04-02] [WebCite Cache ID 6yNgiPi3]
5. Hu J, McKoy K, Papier A, Klaus S, Ryan T, Grossman H, et al. Dermatology and HIV/AIDS in Africa. *J Glob Infect Dis* 2011 Jul;3(3):275-280 [FREE Full text] [doi: [10.4103/0974-777X.83535](https://doi.org/10.4103/0974-777X.83535)] [Medline: [21887061](#)]
6. Carruthers J. Dermatology World. 2013. Dermatology in South Africa URL: [{?page?:?16?.?issue_id?:152144}\(Archived \[WebCite Cache ID 6wIyrXNE9\]\)](http://digital.ipcprints services.com/publication/index.php?i=152144&m=&l=&p=39&pre=1)
7. Christiansen EK, Henriksen E, Lange M, Lapão L, Kaye R, Knarvik U, et al. Telemedicine Momentum. 2015. Moving telemedicine from pilot to scale 18 Critical Success Factors URL: <http://www.telemedicine-momentum.eu/wp-content/uploads/2015/07/MomentumLeaflet2015-BlueprintInANutshell.pdf> [accessed 2018-04-02] [WebCite Cache ID 6yNhGxjX]
8. Walters L, Mars M, Scott R. A Review and Critique of Tele dermatology in the South African Public Health Sector. *Stud Health Technol Inform* 2016;231:143-151 [FREE Full text] [Medline: [27782026](#)]

9. Naidoo S, Mars M. International Society for Telemedicine and eHealth. Luxembourg: International Society for Telemedicine and eHealth; 2015. Telemedicine in the Public Sector in South Africa: An Overview. Vol 8 URL: https://www.isfteh.org/files/media/Global_Telemedicine_and_eHealth_Updates_2015.pdf [accessed 2018-03-26] [WebCite Cache ID 6yCe1kpre]
10. Oninla O. Mucocutaneous Manifestations of HIV and the Correlation with WHO Clinical Staging in a Tertiary Hospital in Nigeria. *AIDS Res Treat* 2014;2014:360970 [FREE Full text] [doi: [10.1155/2014/360970](https://doi.org/10.1155/2014/360970)] [Medline: [25587439](https://pubmed.ncbi.nlm.nih.gov/25587439/)]
11. Walters L, Scott R, Mars M. Design requirements for a teledermatology scale-up framework. *SACJ* In-press 2018:39 (forthcoming).
12. Walters L, Scott R, Mars M. Teledermatology scale-up frameworks: A review and critique. *BMC Health Serv Res* In-press 2018:A (forthcoming).
13. American AOD. American Academy of Dermatology. 2017. Teledermatology Toolkit URL: <https://www.aad.org/practicecenter/managing-a-practice/teledermatology> [accessed 2017-02-20] [WebCite Cache ID 6yNhenIi6]
14. Ernst & Young LLP. HIMSS conference.: Ernst & Young LLP EY telemedicine adoption model unveiled at HIMSS conference URL: <http://www.ey.com/us/en/newsroom/news-releases/news-ey-telemedicine-adoption-model-unveiled-at-himss-conference> [accessed 2018-01-08] [WebCite Cache ID 6wJLVoEN3]
15. Jabareen Y. Building a Conceptual Framework: Philosophy, Definitions, and Procedure. *International Journal of Qualitative Methods* 2009 Dec;8(4):49-62 [FREE Full text] [doi: [10.1177/160940690900800406](https://doi.org/10.1177/160940690900800406)]
16. Miles M, Huberman A. Focusing and Bounding the Collection of Data: The Substantive Start. In: Holland R, editor. *Qualitative data analysis: An expanded source book*. Newbury Park, CA: Sage Publications; 1994:18.
17. Department of Reproductive Health and Research, World Health Organization. Geneva, Switzerland: World Health Organization and ExpandNET; 2010. Nine steps for developing a scaling-up strategy URL: http://www.who.int/immunization/hpv/deliver/nine_steps_for_developing_a_scalingup_strategy_who_2010.pdf [accessed 2018-04-02] [WebCite Cache ID 6yNiHFFHY]
18. Vaishnavi V, Kuechler W. Design Science Research in Information Systems. URL: <http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf> [accessed 2018-03-30] [WebCite Cache ID 6yJ7heHdU]
19. McCarthy J. Circumscription—A form of non-monotonic reasoning. *Artificial Intelligence* 1980 Apr;13(1-2):27-39 [FREE Full text] [doi: [10.1016/0004-3702\(80\)90011-9](https://doi.org/10.1016/0004-3702(80)90011-9)]
20. Project MI. A Guide to the Project Management Body of Knowledge (PMBOK Guide) Sixth Edition Vol Volume.: Project Management Institute, Inc; 2017. URL: <https://www.pmi.org/pmbok-guide-standards/foundational/pmbok/sixth-edition> [accessed 2018-01-08] [WebCite Cache ID 6wJJ6CDtt]
21. Sokolovich N, Fera B. HIMSS15 Annual Conference and Exhibition. Advancing telemedicine through an adoption model URL: <http://assets.fiercemarkets.net/public/healthit/himss15eytelemed.pdf> [accessed 2018-01-08] [WebCite Cache ID 6wJ6JuoK3]
22. Jones T, Broomhead. Tin Tree. 2011. eHealth and managing risk URL: <http://tintree.org/whitepapers/Managing%20Risk.pdf> [accessed 2018-01-08] [WebCite Cache ID 6wJ4wuMQX]
23. Simmons R, Fajans P, Ghiron L. World Health Organization. Geneva, Switzerland: World Health Organization; 2007. Scaling up health service delivery: From pilot innovations to policies programmes URL: http://www.expandnet.net/PDFs/Scaling-Up_Health_Service_Delivery-WHO-ExpandNet.pdf [accessed 2018-04-02] [WebCite Cache ID 6yNiiqqSp]
24. World Health Organization. 2011. HIS country ownership and leadership continuum URL: http://www.who.int/ehealth/resources/his_continuum.pdf [accessed 2018-01-08] [WebCite Cache ID 6wJ3egogL]
25. South Africa National Department of Health. South Africa National Department of Health. Pretoria, South Africa: Government Gazette; 2017. National health insurance for South Africa towards universal health coverage URL: https://www.gov.za/sites/www.gov.za/files/40955_gon627.pdf [accessed 2018-01-08] [WebCite Cache ID 6wJ2uRFUb]

Abbreviations

ICT: information and communication technology

TDSF: Teledermatology Scale-Up Framework

TDSF-IR: Teledermatology Scale-Up Framework Implementation Roadmap

Edited by G Eysenbach; submitted 25.01.18; peer-reviewed by F Kaliyadan, K Ndlovu, C Kovarik; comments to author 15.02.18; revised version received 12.04.18; accepted 17.04.18; published 20.06.18.

Please cite as:

Walters LEM, Scott RE, Mars M

A Teledermatology Scale-Up Framework and Roadmap for Sustainable Scaling: Evidence-Based Development

J Med Internet Res 2018;20(6):e224

URL: <http://www.jmir.org/2018/6/e224/>

doi: [10.2196/jmir.9940](https://doi.org/10.2196/jmir.9940)

PMID: [29925492](https://pubmed.ncbi.nlm.nih.gov/29925492/)

©Laticha Elizabeth Marolana Walters, Richard Ernest Scott, Maurice Mars. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.06.2018. 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.

Original Paper

Home-Based Rehabilitation With Telemonitoring Guidance for Patients With Coronary Artery Disease (Short-Term Results of the TRiCH Study): Randomized Controlled Trial

Andrea Avila¹, MD, MSc; Jomme Claes², BSc, MSc; Kaatje Goetschalckx³, MD; Roselien Buys², BSc, MSc, PhD; May Azzawi⁴, PhD; Luc Vanhees¹, PhD; Véronique Cornelissen¹, PhD

¹Department of Rehabilitation Science, KU Leuven, Leuven, Belgium

²Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium

³Department of Cardiology, University Hospital Leuven, Leuven, Belgium

⁴Cardiovascular Research Group, School of Healthcare Science, Manchester Metropolitan University, Manchester, United Kingdom

Corresponding Author:

Véronique Cornelissen, PhD

Department of Rehabilitation Science

KU Leuven

O&N IV Herestraat 49 - Bus 1510

Leuven, 3000

Belgium

Phone: 32 16 32 91 52

Email: veronique.cornelissen@kuleuven.be

Abstract

Background: Cardiac rehabilitation (CR) is an essential part of contemporary coronary heart disease management. However, patients exiting a center-based CR program have difficulty retaining its benefits.

Objective: We aimed to evaluate the added benefit of a home-based CR program with telemonitoring guidance on physical fitness in patients with coronary artery disease (CAD) completing a phase II ambulatory CR program and to compare the effectiveness of this program in a prolonged center-based CR intervention by means of a randomized controlled trial.

Methods: Between February 2014 and August 2016, 90 CAD patients (unblinded, mean age 61.2 years, SD 7.6; 80/90, 89.0% males; mean height 1.73 m, SD 0.7; mean weight 82.9 kg, SD 13; mean body mass index 27.5 kg/m², SD 3.4) who successfully completed a 3-month ambulatory CR program were randomly allocated to one of three groups: home-based (30), center-based (30), or control group (30) on a 1:1:1 basis. Home-based patients received a home-based exercise intervention with telemonitoring guidance consisting of weekly emails or phone calls; center-based patients continued the standard in-hospital CR, and control group patients received the usual care including the advice to remain physically active. All the patients underwent cardiopulmonary exercise testing for assessment of their peak oxygen uptake (VO₂ P) at baseline and after a 12-week intervention period. Secondary outcomes included physical activity behavior, anthropometric characteristics, traditional cardiovascular risk factors, and quality of life.

Results: Following 12 weeks of intervention, the increase in VO₂ P was larger in the center-based ($P=.03$) and home-based ($P=.04$) groups than in the control group. In addition, oxygen uptake at the first (P -interaction=.03) and second (P -interaction=.03) ventilatory thresholds increased significantly more in the home-based group than in the center-based group. No significant changes were observed in the secondary outcomes.

Conclusions: Adding a home-based exercise program with telemonitoring guidance following completion of a phase II ambulatory CR program results in further improvement of physical fitness and is equally as effective as prolonging a center-based CR in patients with CAD.

Trial Registration: ClinicalTrials.gov NCT02047942; <https://clinicaltrials.gov/ct2/show/NCT02047942> (Archived by WebCite at <http://www.webcitation.org/70CBkSURj>)

(*J Med Internet Res* 2018;20(6):e225) doi:[10.2196/jmir.9943](https://doi.org/10.2196/jmir.9943)

KEYWORDS

cardiac rehabilitation; telemonitoring; exercise; coronary artery disease

Introduction

Cardiovascular diseases (CVD) remain the leading contributor to global premature mortality and morbidity. In Europe, more than 4 million people die from CVD every year, with more than 1.4 million dying before the age of 75 years [1]. Today, secondary prevention of CVD, including coronary artery disease (CAD), by means of cardiac rehabilitation (CR) is considered a class IA recommendation by the European Society of Cardiology, American Heart Association, and American College of Cardiology [2]. CR is now recognized as an essential part of contemporary CAD management that has significantly contributed to the observed reduction in cardiovascular mortality and disability by facilitating the adoption of and adherence to healthy behaviors and promoting an active lifestyle [3]. However, the majority of patients fail to achieve secondary prevention targets in the long term [4]. Many patients receiving center-based CR adopt healthier lifestyles but relapse into old habits when returning to everyday life. After completion of a structured, supervised, exercise-based CR program without any extended support or follow-up, the assumption, of both the participant and CR staff, is that the patient will be able to self-maintain these appropriate health behaviors and optimal CVD risk profile. Unfortunately, studies have shown that patients exiting center-based CR have difficulty retaining the positive benefits derived from their participation [5]. Moreover, previous reports indicate decreased exercise adherence and increased body weight and serum lipid levels as early as 6 months after CR [4,6].

Consequently, there is a need for innovative CR methods to increase long-term adherence to a physically active lifestyle that will result in more sustained effects on health-related physical fitness and cardiovascular health, thus, reducing morbidity and mortality [7]. One attractive strategy is the use of home-based exercise training in combination with telemonitoring guidance. Home-based programs may overcome barriers associated with participation in a center-based exercise program, and they have been shown to provide comparable long-term effects on mortality, recurrent coronary event risk, and cardiovascular risk factors in patients with CVD [8]. This has been attributed partly to the fact that home-based interventions focus more on the development of self-regulatory techniques that create empowerment and perceived control, resulting in longer lasting effects on physical activity improvements [9]. That is, individuals who develop their own physical activity plans are more likely to adhere to these plans than those who have a structured exercise plan imposed on them [9]. The use of information and communication technology to augment home-based programs also enables the provision of additional feedback, education, and counseling [8].

A recent meta-analysis by Buckingham et al [10] found no significant differences in the short-term (<12 months) or long-term (>12 months) patient outcomes including exercise capacity, modifiable risk factors (blood pressure, blood lipid concentrations, and smoking), health-related quality of life

(HRQoL), and cardiac events (mortality, coronary revascularization, and hospital readmissions) among patients participating in home-based or center-based phase II CR. However, there is little evidence about the added benefits of a home-based exercise program for patients being discharged from center-based CR compared with advice only.

In this paper, we report on the secondary objective of the TeleRehabilitation in Coronary Heart disease (TRiCH) study. We aimed to investigate the short-term effect of an HR CR program with telemonitoring guidance on physical fitness and other secondary outcomes in CHD patients following completion of a center-based CR program. We also aimed to compare the effectiveness of this program with that of a prolonged center-based CR program by means of a randomized controlled trial (NCT02047942). The longer-term results of the TRiCH study will be published in a second report.

Methods**Study Design**

We conducted a randomized controlled trial using a three-arm, parallel group design among 90 low-to-moderate risk CAD patients completing a phase II CR program at the University Hospital Leuven (Belgium). The study protocol was approved by the medical ethical committee of the UZ Leuven/KU Leuven. The protocol has been described in detail elsewhere [7].

Patient Population and Randomization

Patients were recruited between February 2014 and August 2016 at the University Hospital Leuven (Belgium). The eligible patients included men and women (aged between 40 and 75 years) with angiographically-documented CAD or previous myocardial infarction, on optimal medical treatment for the last 6 weeks, who successfully completed a supervised ambulatory CR program and who had access to a computer with Internet connection. The exclusion criteria included known clinically significant ventricular arrhythmia or exercise-induced arrhythmia at screening, myocardial ischemia, other cardiac diseases (valve disease with significant hemodynamic consequences, hypertrophic cardiomyopathy, etc), significant illness for the last 6 weeks, co-morbidity that might represent a significant influence on 1-year prognosis (eg, cancer), and co-morbidity that limits exercise testing and/or training. The criteria for ischemia on the electrocardiogram during exercise included horizontal or downsloping ST depression ≥ 1 mm at 80 ms after the J-point or any ST depression >1 mm at 80 ms after the J-point [11]. The eligible patients were contacted in the last weeks of their in-hospital ambulatory CR program (phase II) and were provided verbal information about the TRiCH study. Agreeing patients subsequently received written information and were asked to provide written informed consent according to the principles of Good Clinical Practice and the Declaration of Helsinki.

Procedures

All the agreeing patients who had completed 40 sessions of their ambulatory CR program (phase II) were included and were subsequently randomized in a 1:1:1 ratio to one of three groups: home-based group, center-based group, or a usual care control group by means of a web-based random number generator.

The home-based CR group received training for the first three sessions under the supervision of the investigator. During this period, the patients received an individualized aerobic exercise prescription recommending at least 150 min of exercise per week (preferably 6-7 days/week) at an individually determined target heart rate corresponding to a moderate intensity (ie, 70%-80% of heart rate reserve [HRR]) in their home environment during the 12-week intervention. Furthermore, this group received instructions on how to use the heart rate monitor (Garmin Forerunner 210, Wichita USA) and how to upload their exercise data to the Garmin platform [12]. This application was used to review the training data by both the patient and the investigator [13]. Patients received feedback via phone or email once a week according to their preferences. These contact moments were used for the following purposes: 1) to check for adverse events and injuries, 2) to provide feedback on performed exercise during the preceding week, 3) to discuss the exercise program regarding duration and intensity, and 4) to discuss adherence and barriers to adherence if necessary.

Patients randomized to center-based CR continued their exercise program at the outpatient clinic of UZ Leuven under the direct supervision of physiotherapists. The patients were asked to perform three exercise sessions per week totaling approximately 150 min of endurance exercise. Each training session consisted of predominantly endurance training (2×7 min of cycling, 2×7 min of treadmill walking/running, 7 min of arm ergometry or rowing, and 2×7 min of dynamic calisthenics) and was followed by relaxation. The endurance exercise workload was individually controlled by heart rate monitoring, which was performed by palpation by the physiotherapist during the last minute of each round of exercise. Exercise load was adjusted to maintain target heart rate (70%-80% of the HRR). Patients randomized to the control group received usual care including the standard advice to remain physically active.

Primary Outcome Measure

Primary outcome was change in the exercise capacity following the intervention. Exercise capacity (defined as the maximum amount of physical exertion that a patient could sustain) [14] was determined at baseline and at the end of the intervention using a maximally graded test on a bicycle with breath-by-breath respiratory gas analysis (Ergometrics 800S, Ergometrics, Bitz, Baden-Württemberg, Germany). Peak exercise capacity was defined as the 30-s average oxygen uptake (VO_2) at the highest workload [7]. Ventilatory thresholds (VTs), peak respiratory exchange ratio, and peak heart rate were also established [7].

Secondary Outcome Measures

Secondary outcomes included daily physical activity, measured using a Sensewear Mini Armband (BodyMedia, Inc., Pittsburgh, PA, USA). Steps, sedentary time (duration of sedentary activity

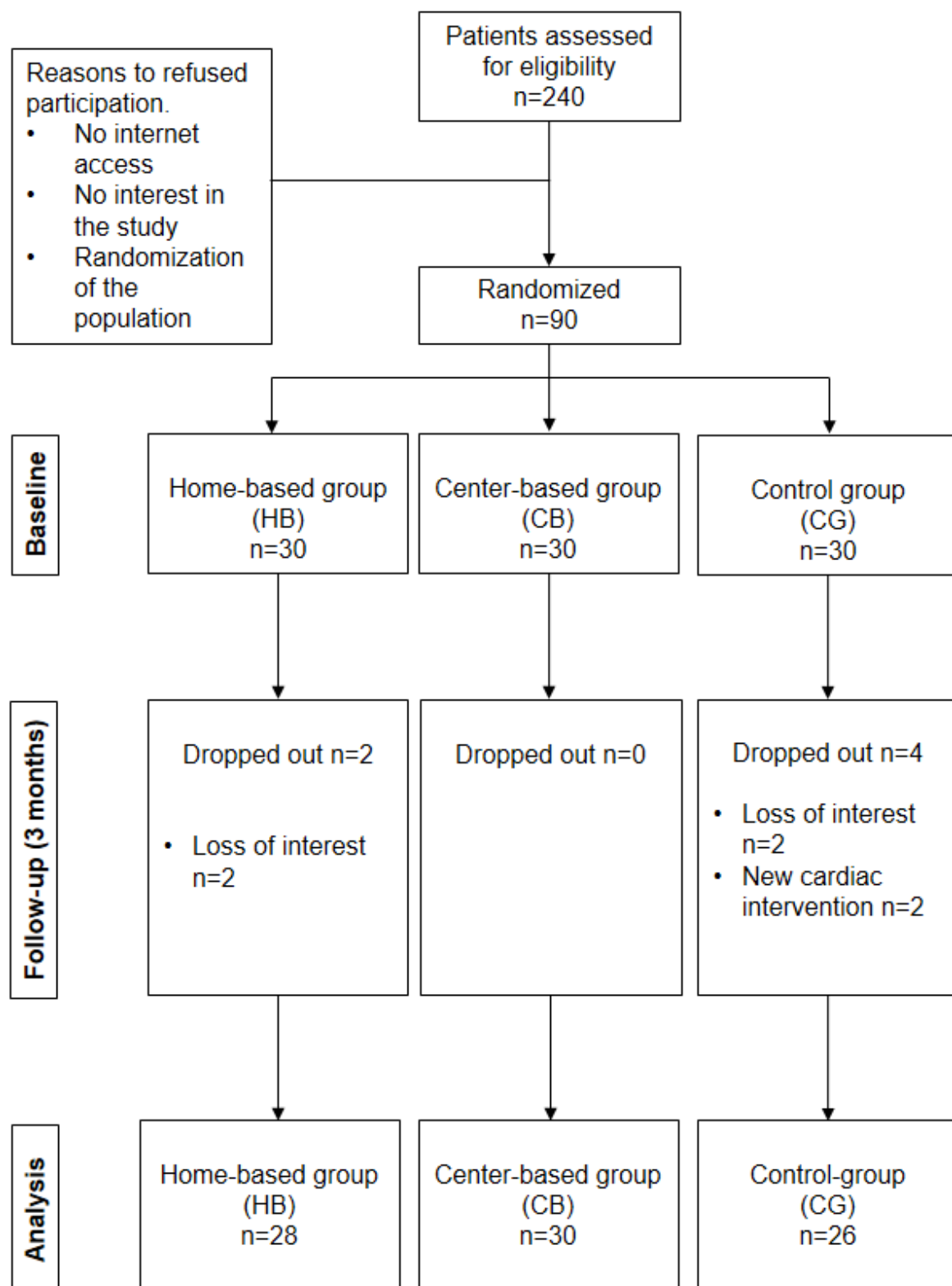
at an intensity of ≤ 1.5 metabolic equivalents of task [METs], min), active energy expenditure (physical activity at an intensity of ≥ 3 METs, kcal), and duration of moderate and vigorous physical activity (≥ 3 METs, min) were used in the analyses [15]. Oxygen uptake on-kinetics were established at least 48 h after the maximal exercise test and was calculated algebraically and expressed as mean response time [7]. Sitting-rising test (SRT), handgrip strength (JAMAR grip strength dynamometer), and quadriceps maximal isometric knee extension strength and endurance (Biodex Medical Systems Inc., 840-000 System 4, New York, USA) were obtained along with traditional cardiovascular risk factors such as anthropometric measures (body mass index, waist and hip circumference) and biochemical parameters of a fasting blood sample (glucose, total cholesterol, low-density lipoprotein [LDL] cholesterol, high-density lipoprotein [HDL] cholesterol, and triglycerides). Additionally, homeostasis assessment model (HOMA) [16] index was calculated using the following formula: fasting plasma glucose (mmol/L) times fasting serum insulin (mU/L) divided by 22.5. Low HOMA-IR values indicated high insulin sensitivity, whereas high HOMA-IR values indicated low insulin sensitivity (insulin resistance). For this study, patients with HOMA-IR ≥ 3.8 were considered to be insulin resistant [17]. Finally, HRQoL was obtained by means of the standard version of the Short Form 36 [7].

Analysis

All data were expressed as mean (SD) or median, range, or percentages (for categorical variables). Statistical analyses were performed using SPSS (version 20; SPSS for windows; SPSS Inc., Chicago, IL). Shapiro-Wilk test was used to assess normality. At baseline, the groups were compared using one-way analysis of variance or chi-square tests. For follow-up data, a linear mixed modeling method was used to evaluate time, group, and time \times group interaction effects. The analysis was complemented with a matrix syntax code including a least significant difference post-hoc test when a significant time \times group interaction identified a group that significantly differed over time. An intention-to-treat analysis was performed on the primary outcome (peak oxygen uptake, VO_2 P), and on-treatment analysis was used for secondary outcomes. Spearman correlation coefficients (p) were calculated between VO_2 P and active energy expenditure and physical activity duration at 12 weeks. A probability level of $P \leq 0.05$ was considered significant.

Results

A total of 90 CAD patients agreed to participate and were randomized to home-based group ($n=30$), center-based group ($n=30$), and control group ($n=30$). Figure 1 shows the flow of patients throughout the study. Six patients, (4 men: control group, $n=4$; home-based, $n=2$) dropped out during the 3-month intervention period. Reasons for dropout included loss of interest (control group, $n=2$; home-based, $n=2$) and a new cardiac intervention (ie, percutaneous coronary intervention) (control group, $n=2$). No serious adverse events related to exercise occurred in any of the groups.

Figure 1. Flow of patients through the study.

The basic characteristics of the study population are described in Table 1. The mean age of the participants was 61.4 (SD 7.3) years (range: 42-73 years). A total of 10 (11.1%) women participated in the study, and patients were on average slightly overweight, 27.5 (SD 3.4 kg/m²). Overall, exercise capacity

was normal, 101.1% (SD 21.1) compared with reference values [18]. Baseline characteristics were comparable between the groups regarding physical characteristics, reason for referral, and pharmacological therapy.

Primary Outcome Measure

Changes in cardiorespiratory parameters are described in Table 2. As can be appreciated from peak respiratory exchange ratios (RER), participants in all three groups exerted a similar maximal effort at baseline and follow-up. The pattern of change in VO_2 P differed significantly over time among the three groups (group \times time interaction, $P=.04$), with a larger improvement following home-based ($P=.03$) and center-based ($P=.04$) interventions than control group interventions. Group \times time interactions were also established for O_2 uptake at the first ventilatory threshold (VT_1 ; P -interaction=.03) and the second ventilatory threshold (VT_2 ; P -interaction=.03), with larger improvements in the home-based group than in the control group.

Secondary Outcome Measures

Changes in daily physical activity are shown in Table 3. Physical activity behavior remained constant after the intervention (P -time=.73). Of all the patients, 97.0% (84/90) met the international guidelines of 150 min or more of moderate physical activity per week [19]. In addition, a significant increase in sedentary time in the center-based group (P -interaction=.02) was found. No significant correlation of change in VO_2 P with change in active energy expenditure (Spearman $\rho=-.14$; $P=.40$) or change in physical activity duration ($\rho=.09$; $P=.44$) was found. However, a significant correlation of VO_2 P with physical activity duration ($\rho=.53$; $P<.001$) at 12 weeks as well as with active energy expenditure ($\rho=.37$; $P<.001$) was found.

Table 1. Baseline characteristics of patients.

Characteristics	Home-based (n=30)	Center-based (n=30)	Control (n=30)
General characteristics			
Age (years), mean (SD)	58.6 (13)	61.9 (7.3)	61.7 (7.7)
Female, n (%)	4 (13)	3 (10)	3 (10)
% of Predicted peak VO_2^a , mean (SD)	99.9 (23.1)	99.3 (20.1)	105.2 (20.2)
Reason for referral, n (%)			
CABG ^b	18 (60)	18 (60)	20 (67)
PCI ^c	12 (40)	12 (40)	10 (33)
Cardiovascular risk factors, n (%)			
Familial predisposition	12 (40)	8 (27)	13 (43)
Hypertension	14 (47)	11 (37)	17 (57)
Diabetes	12 (40)	8 (27)	4 (13)
Dyslipidemia	15 (50)	17 (57)	19 (63)
Smoking			
Never-smoker	11 (37)	14 (47)	15 (50)
Ex-smoker	16 (53)	15 (50)	15 (50)
Current-smoker	3 (10)	1 (3)	0 (0)
Medication, n (%)			
Anti-hypertensive ^d	23 (77)	27 (90)	24 (80)
Beta Blockers	21 (70)	23 (77)	25 (83)
Statins	28 (93)	29 (97)	28 (93)
Aspirin	29 (97)	27 (90)	29 (97)
Anti-thrombotic	19 (63)	18 (60)	23 (77)
Anti-arrhythmic	1 (3)	1 (3)	0 (0)
Hypoglycemic	4 (13)	8 (27)	4 (13)
Vasodilators	0 (0)	1 (3)	2 (7)

^a VO_2 : oxygen uptake.

^bCABG: coronary artery bypass graft.

^cPCI: percutaneous coronary intervention.

^dAnti-hypertensive medication: warfarine and clopidogrel.

Table 2. Changes in cardiorespiratory parameters at baseline and 3-month follow-up.

Parameter	Home-based (n=28), mean (SD)		Center-based (n=30), mean (SD)		Control (n=26), mean (SD)		P value		
	Baseline	3-Month	Baseline	3-Month ^a	Baseline	3-Month	Time	Group	Interaction
VO ₂ peak (mL•kg ⁻¹ •min ⁻¹)	26.7 (6.55)	27.8 (6.83)	25.4 (7.32)	26.7 (7.90)	26.6 (4.97)	26.4 (5.42)	.08	.69	.04 ^b
VT1 ^a (mL•kg ⁻¹ •min ⁻¹)	19.5 (1.07)	21.5 (1.07)	19.5 (1.04)	20.4 (1.04)	19.9 (1.08)	19.3 (1.11)	.81	.06	.03 ^b
VT2 ^c (mL•kg ⁻¹ •min ⁻¹)	24.9 (5.25)	26.3 (6.98)	22.7 (6.95)	24.2 (7.13)	24.7 (5.08)	22.9 (4.19)	.41	.49	.03 ^b
Duration (s)	570 (136)	587 (157)	549 (133)	552 (157)	600 (126)	579 (116)	.89	.52	.23
Peak heart rate (bpm)	140 (18.8)	139 (17.8)	141 (21.5)	140 (21.1)	140 (18.9)	140 (16.6)	.75	.97	.95
Peak load (watts)	198 (49)	200 (54)	191 (50)	191 (54)	206 (41)	197 (38)	.39	.67	.15
Peak RER ^d	1.24 (0.89)	1.21 (0.10)	1.23 (0.80)	1.24 (0.10)	1.20 (0.8)	1.20 (0.13)	.47	.32	.36
Borg scale	15.8 (1.16)	15.8 (1.33)	16.2 (1.04)	16 (1.17)	15.9 (1.05)	16.2 (1.02)	.87	.45	.38

^aVT1: first ventilatory threshold.^bP- interaction<.05.^cVT2: second ventilatory threshold.^dRER: respiratory exchange ratios.**Table 3.** Changes in daily physical activity at baseline and 3-month follow-up.

Physical activity	Home-based (n=24), mean (range)		Center-based (n=28), mean (range)		Control (n=26), mean (range)		P value		
	Baseline	3-Month	Baseline	3-Month	Baseline	3-Month	Time	Group	Interaction
Steps per day	7896 (2018-12554)	6469 (473-12828)	7608 (2474-13281)	7065 (489-14785)	6419 (2227-13181)	6408 (296-12041)	.18	.56	.18
Sedentary time (≤1.5 METs; min/day)	1039 (688-1260)	1032 (790-1455)	1005 (122-1290)	1094 (857-1254) ^a	1100 (825-1355)	1062 (484-1402)	.56	.43	.02 ^b
Active energy expenditure (>3METs; kcal)	1336 (351-3217)	1307 (661-2246)	1137 (484-2539)	1244 (549-2745)	1223 (401-2253)	1196 (181-2510)	.56	.40	.45
Physical activity duration (>3METs; min/day)	145 (34-299)	141(51-259)	146(28-417)	134(29-366)	114(30-311)	114(6-382)	.73	.27	.47
Moderate physical activity duration (3-6 METs; min/day)	136(34-238)	134(49-241)	140(28-391)	128(27-348)	109(29-303)	115(25-368)	.57	.24	.62
Vigorous physical activity duration (>6 METs; min/day)	8(0-33)	7(0-24)	6(0-26)	6(0-24)	5(0-20)	2(0-27)	.50	.21	.59

^aMETs: metabolic equivalents of task.^bP- interaction <.05.

As shown in table [Table 4](#), isometric handgrip strength (HG), isometric quadriceps strength, and endurance, as well as exercise-onset oxygen uptake on-kinetics remained stable during the follow-up period. Additionally, cardiovascular risk factors ([Figure 2](#)) and anthropometrics ([Table 5](#)) were similar between the groups at baseline and remained stable during the follow-up period, except for an increase in HOMA index (P -time=.05), which was not different between the groups. Finally, there were no significant changes in the overall score for HRQoL (P -interaction=.57) as well as the physical (P -interaction=.50) and mental (P -interaction=.85) composite scores. [Table 6](#) shows HRQoL from baseline and follow-up evaluations.

Training Data

Patients in the home-based group completed an average of 2.5 sessions per week (range: 12-60 sessions for 12 weeks), whereas those in the center-based group completed an average of 2.0 sessions per week (range: 4-36 sessions for 12 weeks). Patients in the home-based group exercised for an average 164 min per week at an average intensity of 46.8% of HRR (76.7 min within the prescribed zone). Patients in the center-based group exercised for an average 90 min per week at an average intensity of 61.2% of HRR.

Table 4. Changes in muscle strength and exercise-onset oxygen uptake (VO₂) kinetics.

Parameter	Home-based (n=23), mean (SD)		Center-based (n=29), mean (SD)		Control (n=19), mean (SD)		P value		
	Baseline	3-Month	Baseline	3-Month	Baseline	3-Month	Time	Group	Interaction
Muscle strength									
Handgrip strength (kg)	43.1 (10.5)	44.7 (12.3)	40.2 (8.6)	41.2 (8.3)	41.6 (8.3)	43.8 (9.3)	.99	.23	.23
Isometric quadriceps extension (60° Nm)	151.8 (28)	164.1 (37)	150.5 (44.9)	155 (43.4)	148.7 (30)	148.8 (28.3)	.23	.47	.86
Extension total work (J)	1614 (680)	1976 (718)	1758 (756)	1893 (717)	1694 (796)	1906 (689)	.09	.94	.52
Exercise-Onset VO₂ kinetics									
Average MRT ^a (s)	45.5 (16.2)	39.8 (9.3)	38.7 (8.1)	40.8 (9.1)	39.8 (16.9)	43.6 (22)	.98	.64	.19

^aMRT: mean response time.

Figure 2. Cardiovascular risk factors. No significant changes were found in cardiovascular risk factors for total cholesterol (P -interaction=.82), HDL-cholesterol (P -interaction=.69), LDL-cholesterol (P -interaction=.79), triglycerides (P -interaction=.27), fasting glucose (P -interaction=.71), HOMA index (P -interaction=.93). Dark gray column: home-based group, White column: center-based group, Light grey column: control group. HDL: high-density lipoprotein; HOMA: homeostasis assessment; LDL: low-density lipoprotein.

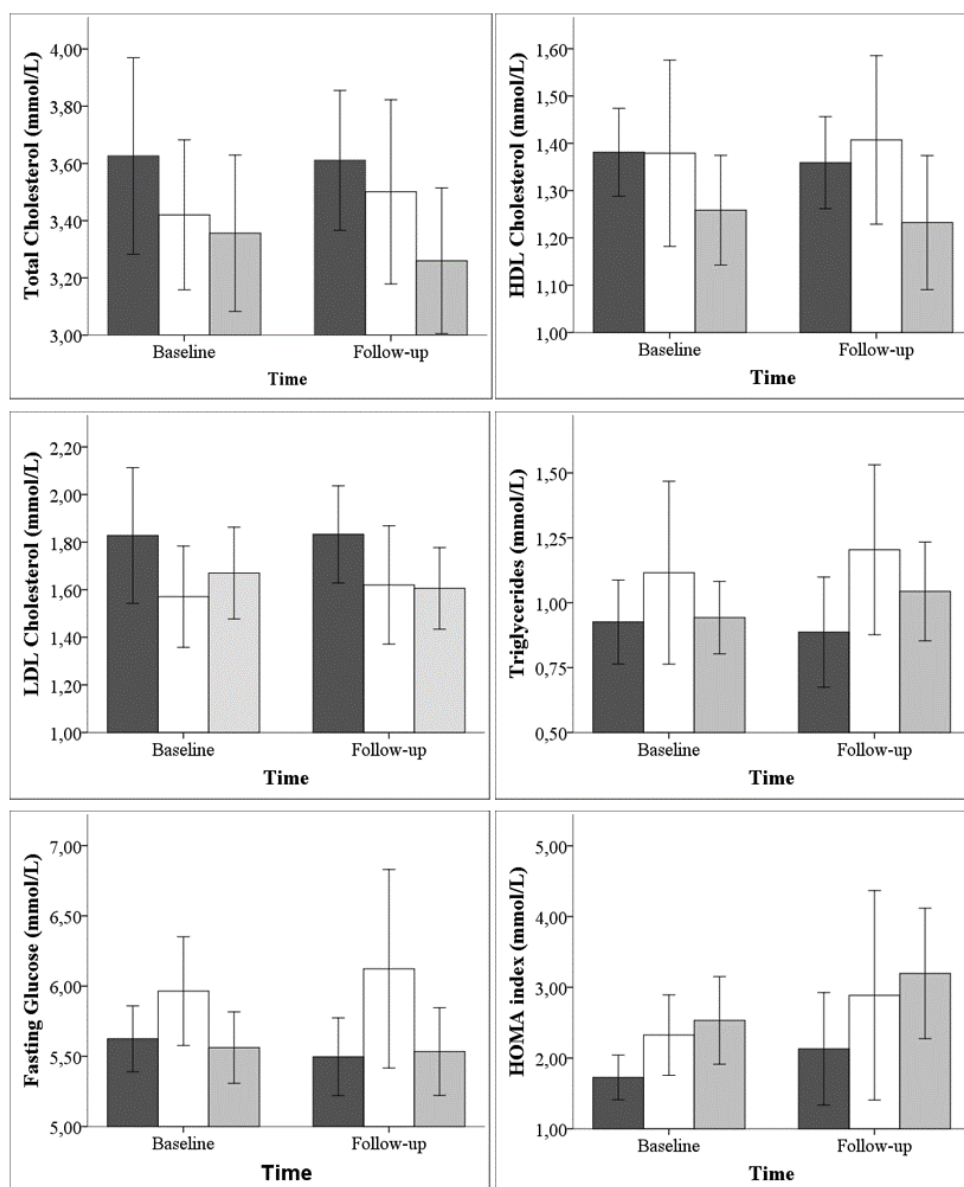


Table 5. Outcomes at baseline and 3-month follow-up regarding anthropometric parameters.

Anthropometrics	Home-based (n=28), mean (SD)		Center-based (n=30), mean (SD)		Control (n=26), mean (SD)		P value		
	Baseline	3-Month	Baseline	3-Month	Baseline	3-Month	Time	Group	Interaction
Weight (kg)	80.4 (10.3)	79.8 (10.3)	82.9 (15.3)	82.4 (15)	85 (12.7)	84.1 (11.9)	.36	.22	.55
Body mass index (kg/m ²)	26.6 (2.5)	26.4 (2.5)	27.8 (4)	27.6 (4.1)	28 (3.3)	27.6 (2.6)	.28	.21	.53
Body fat (%)	26.8 (5.7)	26.5 (6.1)	29.5 (5.5)	28.4 (6.4)	29.5 (5.1)	28.8 (4.2)	.16	.12	.60
Waist Circumference (cm)	96.8 (8.8)	95.2 (8.2)	98.7 (11.2)	99 (11.4)	99.5 (9.8)	99.6 (9.6)	.81	.29	.11
Hip circumference (cm)	101.1 (5.9)	101 (5.4)	102.6 (7.4)	102.7 (7.6)	102.5 (4.1)	102.1 (5.1)	.31	.42	.82
Systolic blood pressure (mmHg)	125.3 (15.6)	124.1 (13.9)	127.4 (15)	124.1 (13.8)	123.6 (13)	122.9 (14)	.29	.75	.73
Diastolic blood pressure (mmHg)	75.4 (9.5)	75.8 (9)	76 (8.3)	74.7 (8.2)	75.5 (8.9)	76.9 (7.9)	.81	.91	.37
Heart rate at rest (bpm)	56.8 (9.1)	56.4 (7.4)	57.1 (8.2)	57.7 (10.1)	56 (8.1)	56.7 (14)	.90	.68	.84

Table 6. Changes in health-related quality of life (HRQoL).

HRQoL measures	Home-based (n=28), mean (SD)		Center-based (n=30), mean (SD)		Control (n=26), mean (SD)		P value		
	Baseline	3-Month	Baseline	3-Month	Baseline	3-Month	Time	Group	Interaction
Physical Function	90 (10.7)	91.9 (7.6)	85.3 (15.6)	87.4 (13.6)	81.4 (15.7)	84.3 (16.1)	.19	.04	.49
Role-physical function	83.6 (27.3)	87.5 (28.4)	74.2 (36.6)	83.9 (31.3)	58 (41.4)	61.4 (38.9)	.15	.00	.43
Bodily pain	84.2 (13.7)	80.9 (17.7)	82.5 (19.2)	83.8 (23.3)	72.7 (23.3)	77.6 (23.2)	.83	.11	.70
General health	74.2 (13.1)	75.2 (18.7)	75.7 (13.6)	72.4 (22.9)	64.2 (16.6)	71.5 (19.1)	.41	.17	.17
Vitality	72.8 (13.9)	71 (13.4)	69.3 (15.6)	74.6 (15.4)	65.3 (14.6)	63.9 (20)	.70	.06	.10
Social function	87.6 (16.7)	90.6 (16.5)	86.6 (17.3)	89.1 (15.7)	83.7 (18.2)	89.5 (17.5)	.20	.81	.90
Role-emotional function	84.6 (27)	86.8 (22.8)	84.5 (33.3)	86.8 (29.1)	86.2 (30.2)	84.7 (31)	.73	.99	.79
Mental health	78.6 (15.6)	77.5 (17.6)	79.3 (13.3)	81.4 (15)	78.2 (15.6)	76.6 (21.1)	.91	.72	.59
Physical Composite Score	80.9 (12.3)	81.3 (13.1)	77.2 (15.2)	80.8 (16.3)	67.9 (17.1)	71.8 (18.3)	.008	.07	.50
Mental Composite Score	79.8 (13.7)	80.2 (13.5)	77.5 (18)	81.3 (15.5)	75.1 (13.8)	77.2 (16.2)	.06	.38	.85
SF-36 ^a	82.2 (13.3)	82.6 (13)	79.8 (16.1)	82.6 (15.8)	73.3 (15.1)	76.4 (16.4)	.07	.06	.57

^aSF-36: Short Form 36. Scores of the domains of the SF-36: 0=worst and 100=best score.

Discussion

The significant finding of our randomized controlled study is that a 3-month home-based rehabilitative intervention with telemonitoring guidance results in further improvement of exercise capacity (VO₂ P) in CAD patients who had recently completed a phase II ambulatory program, and a home-based program is as effective as a prolonged center-based CR. The observed improvements of 1.30 mL·kg⁻¹·min⁻¹ and 1.10 mL·kg⁻¹·min⁻¹ in VO₂ P of center-based and home-based groups, respectively, are likely to be clinically relevant. It has been shown in earlier studies that a 1 mL·kg⁻¹·min⁻¹ increase in exercise capacity is associated with a 10% reduction in cardiovascular mortality [20]. Thus, our results support the added value of a structured continued rehabilitation program.

There are only few studies in the literature that have investigated the effectiveness of a home-based, telemonitored, phase III CR

program starting immediately after completion of a phase II center-based CR program. A small proof-of-concept study by Brubaker et al [5] randomly assigned 31 patients to home-based, center-based, or standard care. In line with our results, they found that the home-based program was as effective as the center-based program at improving and maintaining oxygen consumption among patients 9 months after exiting a CR program. In the Telerehab III trial [21], 140 patients were randomized to a telerehabilitation program in addition to conventional CR or conventional CR alone. This study also reported that a 6-month patient-specific comprehensive telerehabilitation program initiated 6 weeks after the start of ambulatory rehabilitation leads to a bigger improvement in VO₂ P and confirmed our results of lack of an additional weight loss, blood pressure reduction, lipid profile improvement, and glycemic control.

In the last decade, several meta-analyses have been published demonstrating the effectiveness of home-based programs for

CAD patients implemented as a phase II CR program [22-24]. Although the increase in VO_2P of 4%-5% in our phase III center-based and home-based groups is less than what has been seen in previous phase II programs, this is still of clinical relevance, as the purpose of phase III (maintenance phase) CR is to preserve, or if possible, enhance the health benefits gained in phase II. Our results further show that although patients in the home-based group exercised only 75 min per week at the prescribed intensity and the average intensity was below the recommended thresholds, patients were still able to further increase their exercise capacity compared with those in the control group who receive only advice on how to maintain a physically active lifestyle. This is in line with meta-analytic results of Swain and Franklin [25], demonstrating that in healthy individuals with a mean baseline $\text{VO}_{2\text{max}} < 40 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, there is no clear minimal intensity threshold to increase their aerobic capacity and that patients already show improvements when exercising at an intensity of 40% of HRR. Yet, there is abundant evidence that larger effects on health and fitness are established when individuals exercise at higher intensities and larger duration [26].

Regarding physical activity, Ayabe et al [27], reported that 6500-8500 steps/day should be considered as the minimal and optimal goal of physical activity for secondary prevention of CVD. With the number of steps ranging between 7612 (home-based) and 7700 (center-based), the patients in the home-based and center-based groups were within this target zone, whereas those in the control group (5566) did not seem to reach this goal. However, we were not able to promote an additional increase in the number of daily steps. These results are in line with previous studies that have demonstrated how exercise interventions focused on physical fitness improvement in cardiac patients do not influence the improvement of physical activity [21,28,29].

We observed, however, a small but significant increase in sedentary time in the center-based group. Evidence suggests that those participating in exercise-focused interventions are not likely to reduce their sedentary time by a meaningful amount [30]. King et al [31] explained this behavior as a compensatory effect for exercise. That is, the simple fact of enrolling in a supervised exercise program might reduce physical activity levels throughout the rest of the day. This compensatory effect acts in different ways by promoting adjustments in energy expenditure in order to save energy or recover from the exercise training, consequently, increasing the sedentary time [32]. Growing evidence suggest that prolonged sedentary behavior can affect cardiovascular and all-cause mortality risk independent of physical activity [33]. Probably, recommendations on sedentary behavior can be included in CR programs, given the current state of the evidence [30].

Small improvements in weight reduction and body composition as well as in blood pressure were observed post-intervention, although this did not reach significance. This is consistent with the findings in the Telerehab III trial [21]. According to Frederix et al, digital health interventions seem to be able to improve cardiovascular risk factors in primary prevention, but not in secondary prevention, programs [21]. Furthermore, we

hypothesize that when the pharmacological management of the groups is close to optimal, like in our patients, the incremental benefit of secondary prevention programs over usual care is very small [34].

One of the main objectives of CR is to optimize patients' physical functionality as a means to improve quality of life. In our study, no effects of our interventions on different HRQoL domains could be found. Contrary to our results, Smith et al [35] found clinically significant improvements in HRQoL in their home-based group compared with their hospital-based group after a 6-month intervention. The authors considered that 6 months of CR, regardless of location, was associated with improvements in physical HRQoL. Thus, it is possible that with a longer intervention for both the groups, some differences could have been obtained in our study.

One area that is not commonly considered with CR is the functional status or abilities of the patient. The typical patient in CR is over the age of 60 and presents with multiple cardiovascular risk factors, such as inactivity and obesity, and is recovering from a recent cardiac event. All these factors can lead to deficits in balance, mobility, and function. Sumide et al, found that musculoskeletal fitness and flexibility, measured by SRT, was a significant predictor of mortality in 51- to 80-year-old participants [36]. In this study, no significant difference was seen in SRT between the groups ($P=.36$). Our results are in line with those of Oerkild et al [37]. Regarding components of muscular fitness, Mroszczyk et al [38] and Thomaes et al [39] found increased HG strength in patients following 3 months of a (predominantly aerobic exercise training) CR program. In our study, however, no changes were observed after 12 weeks of intervention. It is possible that the largest gains in HG strength appear during phase II of CR, while our intervention targeted phase III patients. In addition, we found no significant difference in knee extension strength or resistance after 3 months of home-based or center-based training, which may be explained by the focus on aerobic training during the intervention. These results are also consistent with those of previous studies [40,41]. Conraads et al collected muscle strength data in 75 CAD patients before and after 12 weeks of an aerobic interval training or continuous training CR program, finding that muscle strength did not improve. The authors considered as a possible cause the use of statins that has been associated with negative side effects on the muscles. In our study, 94% (79/84) of the patients are treated with statins, although insufficient evidence exists to prove that statins really affect muscle strength [42]. Currently, there is very limited evidence regarding the effects of telerehabilitation on muscle strength, and more research is needed [43].

Finally, while short-term changes are of interest, it is important to establish whether the benefits are maintained over time; thus, further research should focus on the long-term effects of home-based CR with telemonitoring guidance.

Limitations

Our study should be interpreted within the context of its limitations. First, next to physical activity training, CR includes other important core components such as nutritional counseling, risk factor management, and psychosocial management.

Although physical activity training comprises 30%-50% (up to >70%) of all CR activities, it should be acknowledged that this study evaluates the effect of physical activity telemonitoring rather than telerehabilitation [3]. Second, heart rate monitors were used only in the home-based group as we opted to not change the traditional center-based program where heart rate is measured by palpation by physiotherapists. We were not able to precisely define the exact number of minutes patients spend within the prescribed training zone.

Another limitation of this study is the lack of blinding of test personnel. However, as the main outcome measure was VO_2P , the effort of the participants can be objectively quantified by

means of RER and subjectively by means of the Borg scale [44]. The study, as in most randomized controlled trials, has missing outcome data. Regarding muscle strength, 19 values were completely missing at random due to technical problems, and regarding physical activity, the data was incomplete and, thus, excluded for 12 patients.

Conclusion

The results of our study show that home-based CR with telemonitoring guidance can be an effective alternative to center-based CR for further improving exercise capacity following phase II CR in CHD patients.

Acknowledgments

AA is supported by a doctoral research grant funded by the European Commission through MOVE-AGE, an Erasmus Mundus Joint Doctorate programme (2011-2015). RB received a research grant from the Belgian Fund for Cardiac Surgery.

The authors thank the patients who participated in the TRiCH study as well as the staff of the cardiac rehabilitation center in UZ Leuven for their enthusiastic support of this project.

Authors' Contributions

VC contributed to the conception and design of the work. AA and JC contributed to the acquisition, analysis, or interpretation of data for the work. VC and AA drafted the manuscript. JC, RB, KG, MA, and LV critically revised the manuscript. All the authors gave final approval and agree to be accountable for all aspects of work ensuring integrity and accuracy.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 804KB](#) - [jmir_v20i6e225_app1.pdf](#)]

References

1. Perk J, De BG, Gohlke H, Graham I, Reiner Z, Verschuren M, European Association for Cardiovascular Prevention & Rehabilitation (EACPR), ESC Committee for Practice Guidelines (CPG). European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J 2012 Jul;33(13):1635-1701 [[FREE Full text](#)] [doi: [10.1093/eurheartj/ehs092](#)] [Medline: [22555213](#)]
2. Piepoli MF, Corrà U, Adamopoulos S, Benzer W, Bjarnason-Wehrens B, Cupples M, et al. Secondary prevention in the clinical management of patients with cardiovascular diseases. Core components, standards and outcome measures for referral and delivery: a policy statement from the cardiac rehabilitation section of the European Association for Cardiovascular Prevention & Rehabilitation. Endorsed by the Committee for Practice Guidelines of the European Society of Cardiology. Eur J Prev Cardiol 2014 Jun;21(6):664-681. [doi: [10.1177/2047487312449597](#)] [Medline: [22718797](#)]
3. Frederix I, Van DN, Hansen D, Berger J, Bonne K, Alders T, et al. Increasing the medium-term clinical benefits of hospital-based cardiac rehabilitation by physical activity telemonitoring in coronary artery disease patients. Eur J Prev Cardiol 2015 Feb;22(2):150-158. [doi: [10.1177/2047487313514018](#)] [Medline: [24249840](#)]
4. Janssen V, De GV, van EH, Maes S. Beyond resolutions? A randomized controlled trial of a self-regulation lifestyle programme for post-cardiac rehabilitation patients. Eur J Prev Cardiol 2013 Jun;20(3):431-441. [doi: [10.1177/2047487312441728](#)] [Medline: [22396248](#)]
5. Brubaker PH, Rejeski WJ, Smith MJ, Sevensky KH, Lamb KA, Sotile WM, et al. A home-based maintenance exercise program after center-based cardiac rehabilitation: effects on blood lipids, body composition, and functional capacity. J Cardiopulm Rehabil 2000;20(1):50-56. [Medline: [10680098](#)]

6. Lear SA, Spinelli JJ, Linden W, Brozic A, Kiess M, Frohlich JJ, et al. The Extensive Lifestyle Management Intervention (ELMI) after cardiac rehabilitation: a 4-year randomized controlled trial. *Am Heart J* 2006 Aug;152(2):333-339. [doi: [10.1016/j.ahj.2005.12.023](https://doi.org/10.1016/j.ahj.2005.12.023)] [Medline: [16875919](#)]
7. Avila A, Goetschalckx K, Vanhees L, Cornelissen V. A Randomized Controlled Study Comparing Home-Based Training with Telemonitoring Guidance Versus Center-Based Training in Patients with Coronary Heart Disease: Rationale and Design of the Tele-Rehabilitation in Coronary Heart Disease (TRiCH) Study. *J Clin Trials* 2014;04(04):1-5. [doi: [10.4172/2167-0870.1000175](https://doi.org/10.4172/2167-0870.1000175)]
8. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. *Heart* 2016 Dec 01;102(15):1183-1192. [doi: [10.1136/heartjnl-2015-308966](https://doi.org/10.1136/heartjnl-2015-308966)] [Medline: [26936337](#)]
9. Ferrier S, Blanchard CM, Vallis M, Giacomantonio N. Behavioural interventions to increase the physical activity of cardiac patients: a review. *Eur J Cardiovasc Prev Rehabil* 2011 Feb;18(1):15-32. [doi: [10.1097/HJR.0b013e32833ace0e](https://doi.org/10.1097/HJR.0b013e32833ace0e)] [Medline: [20502341](#)]
10. Buckingham SA, Taylor RS, Jolly K, Zawada A, Dean SG, Cowie A, et al. Home-based versus centre-based cardiac rehabilitation: abridged Cochrane systematic review and meta-analysis. *Open Heart* 2016;3(2):e000463 [FREE Full text] [doi: [10.1136/openhrt-2016-000463](https://doi.org/10.1136/openhrt-2016-000463)] [Medline: [27738516](#)]
11. Hendel R, Carey K. In: Hendel R, Carey K, editors. *Cardiology Procedures: A Clinical Primer*. London: Imprint Springer; 2017:A.
12. Garmin International Inc. Garmin Connect. URL: <https://connect.garmin.com/nl-NL/> [WebCite Cache ID 6xRyfpipN]
13. Kraal JJ, Peek N, Van DAMME, Kemps HM. Effects of home-based training with telemonitoring guidance in low to moderate risk patients entering cardiac rehabilitation: short-term results of the FIT@Home study. *Eur J Prev Cardiol* 2014 Nov;21(2 Suppl):26-31. [doi: [10.1177/2047487314552606](https://doi.org/10.1177/2047487314552606)] [Medline: [25354951](#)]
14. Walker H, Hall W, Hurst J. *Clinical methods: the history, physical, and laboratory examinations*. 3rd ed. Boston: Butterworths; 1990.
15. Pattyn N, Vanhees L, Cornelissen VA, Coeckelberghs E, De MC, Goetschalckx K, et al. The long-term effects of a randomized trial comparing aerobic interval versus continuous training in coronary artery disease patients: 1-year data from the SAINTEX-CAD study. *Eur J Prev Cardiol* 2016 Dec;23(11):1154-1164. [doi: [10.1177/2047487316631200](https://doi.org/10.1177/2047487316631200)] [Medline: [26858279](#)]
16. Wallace TM, Matthews DR. The assessment of insulin resistance in man. *Diabet Med* 2002 Jul;19(7):527-534. [Medline: [12099954](#)]
17. Marques-Vidal P, Mazoyer E, Bongard V, Gourdy P, Ruidavets J, Drouet L, et al. Prevalence of insulin resistance syndrome in southwestern France and its relationship with inflammatory and hemostatic markers. *Diabetes Care* 2002 Aug;25(8):1371-1377. [Medline: [12145237](#)]
18. Wasserman KH, Sietsema K, Sue D, Stringer W, MD, Whipp B, MD. In: Wasserman KH, editor. *Principles of exercise testing and interpretation; including pathophysiology and clinical applications*, 5th ed. Portland: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2012; 2015.
19. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc* 2011 Jul;43(7):1334-1359. [doi: [10.1249/MSS.0b013e318213fefb](https://doi.org/10.1249/MSS.0b013e318213fefb)] [Medline: [21694556](#)]
20. Kavanagh T, Mertens DJ, Hamm LF, Beyene J, Kennedy J, Corey P, et al. Peak oxygen intake and cardiac mortality in women referred for cardiac rehabilitation. *Journal of the American College of Cardiology* 2003;17;42(12):2139-2143 <https://www.sciencedirect.com/science/article/pii/S0735109703013056?via%3Dihub> [FREE Full text] [doi: [10.1016/j.jacc.2003.07.028](https://doi.org/10.1016/j.jacc.2003.07.028)] [Medline: [14680741](#)]
21. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-Term Effectiveness of a Comprehensive Internet-Based and Patient-Specific Telerehabilitation Program With Text Messaging Support for Cardiac Patients: Randomized Controlled Trial. *J Med Internet Res* 2015;17(7):e185 [FREE Full text] [doi: [10.2196/jmir.4799](https://doi.org/10.2196/jmir.4799)] [Medline: [26206311](#)]
22. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. *Mayo Clin Proc* 2015 Apr;90(4):469-480. [doi: [10.1016/j.mayocp.2014.12.026](https://doi.org/10.1016/j.mayocp.2014.12.026)] [Medline: [25841251](#)]
23. Huang K, Liu W, He D, Huang B, Xiao D, Peng Y, et al. Telehealth interventions versus center-based cardiac rehabilitation of coronary artery disease: A systematic review and meta-analysis. *Eur J Prev Cardiol* 2015 Aug;22(8):959-971. [doi: [10.1177/2047487314561168](https://doi.org/10.1177/2047487314561168)] [Medline: [25488550](#)]
24. Frederix I, Vanhees L, Dendale P, Goetschalckx K. A review of telerehabilitation for cardiac patients. *J Telemed Telecare* 2015 Jan;21(1):45-53. [doi: [10.1177/1357633X14562732](https://doi.org/10.1177/1357633X14562732)] [Medline: [25475219](#)]
25. Swain DP, Franklin BA. VO(2) reserve and the minimal intensity for improving cardiorespiratory fitness. *Med Sci Sports Exerc* 2002 Jan;34(1):152-157. [Medline: [11782661](#)]
26. Mezzani A, Hamm LF, Jones AM, McBride PE, Moholdt T, Stone JA, European Association for Cardiovascular Prevention/Rehabilitation, American Association of Cardiovascular/Pulmonary Rehabilitation, Canadian Association of

- Cardiac Rehabilitation. Aerobic exercise intensity assessment and prescription in cardiac rehabilitation: a joint position statement of the European Association for Cardiovascular Prevention and Rehabilitation, the American Association of Cardiovascular and Pulmonary Rehabilitation and the Canadian Association of Cardiac Rehabilitation. *Eur J Prev Cardiol* 2013 Jun;20(3):442-467. [doi: [10.1177/2047487312460484](https://doi.org/10.1177/2047487312460484)] [Medline: [23104970](https://pubmed.ncbi.nlm.nih.gov/23104970/)]
27. Ayabe M, Brubaker PH, Dobrosielski D, Miller HS, Kiyonaga A, Shindo M, et al. Target step count for the secondary prevention of cardiovascular disease. *Circ J* 2008 Feb;72(2):299-303 [FREE Full text] [Medline: [18219170](https://pubmed.ncbi.nlm.nih.gov/18219170/)]
 28. Kraal JJ, Van DAMME, Abu-Hanna A, Stut W, Peek N, Kemps HM. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: Results of the FIT@Home study. *Eur J Prev Cardiol* 2017 Aug;24(12):1260-1273 [FREE Full text] [doi: [10.1177/2047487317710803](https://doi.org/10.1177/2047487317710803)] [Medline: [28534417](https://pubmed.ncbi.nlm.nih.gov/28534417/)]
 29. ter HN, Huisstede BMA, Stam HJ, van DRT, Sunamura M, van DBRJG. Does cardiac rehabilitation after an acute cardiac syndrome lead to changes in physical activity habits? Systematic review. *Phys Ther* 2015 Feb;95(2):167-179. [doi: [10.2522/ptj.20130509](https://doi.org/10.2522/ptj.20130509)] [Medline: [25278337](https://pubmed.ncbi.nlm.nih.gov/25278337/)]
 30. Prince SA, Saunders TJ, Gresty K, Reid RD. A comparison of the effectiveness of physical activity and sedentary behaviour interventions in reducing sedentary time in adults: a systematic review and meta-analysis of controlled trials. *Obes Rev* 2014 Nov;15(11):905-919 [FREE Full text] [doi: [10.1111/obr.12215](https://doi.org/10.1111/obr.12215)] [Medline: [25112481](https://pubmed.ncbi.nlm.nih.gov/25112481/)]
 31. King NA, Caudwell P, Hopkins M, Byrne NM, Colley R, Hills AP, et al. Metabolic and behavioral compensatory responses to exercise interventions: barriers to weight loss. *Obesity (Silver Spring)* 2007 Jun;15(6):1373-1383 [FREE Full text] [doi: [10.1038/oby.2007.164](https://doi.org/10.1038/oby.2007.164)] [Medline: [17557973](https://pubmed.ncbi.nlm.nih.gov/17557973/)]
 32. de MBP, Marins JCB, Franceschini SDCC, Reis JS, Amorim PRDS. Aerobic exercise did not have compensatory effects on physical activity levels in type 2 diabetes patients. *J Sports Sci* 2015 Sep;33(6):545-551. [doi: [10.1080/02640414.2014.951875](https://doi.org/10.1080/02640414.2014.951875)] [Medline: [25259751](https://pubmed.ncbi.nlm.nih.gov/25259751/)]
 33. Biswas A, Faulkner GE, Oh PI, Alter DA. Patient and practitioner perspectives on reducing sedentary behavior at an exercise-based cardiac rehabilitation program. *Disabil Rehabil* 2017 Jun 06;1-8. [doi: [10.1080/09638288.2017.1334232](https://doi.org/10.1080/09638288.2017.1334232)] [Medline: [28583032](https://pubmed.ncbi.nlm.nih.gov/28583032/)]
 34. Clark AM, Hartling L, Vandermeer B, McAlister FA. Meta-analysis: secondary prevention programs for patients with coronary artery disease. *Ann Intern Med* 2005 Nov 1;143(9):659-672. [Medline: [16263889](https://pubmed.ncbi.nlm.nih.gov/16263889/)]
 35. Smith KM, Arthur HM, McKelvie RS, Kodis J. Differences in sustainability of exercise and health-related quality of life outcomes following home or hospital-based cardiac rehabilitation. *Eur J Cardiovasc Prev Rehabil* 2004 Aug;11(4):313-319. [Medline: [15292765](https://pubmed.ncbi.nlm.nih.gov/15292765/)]
 36. Brito LBB, Ricardo DR, Araújo DSMSD, Ramos PS, Myers J, Araújo CGSD. Ability to sit and rise from the floor as a predictor of all-cause mortality. *Eur J Prev Cardiol* 2014 Jul;21(7):892-898. [doi: [10.1177/2047487312471759](https://doi.org/10.1177/2047487312471759)] [Medline: [23242910](https://pubmed.ncbi.nlm.nih.gov/23242910/)]
 37. Oerikild B, Frederiksen M, Hansen JF, Simonsen L, Skovgaard LT, Prescott E. Home-based cardiac rehabilitation is as effective as centre-based cardiac rehabilitation among elderly with coronary heart disease: results from a randomised clinical trial. *Age Ageing* 2011 Jan;40(1):78-85. [doi: [10.1093/ageing/afq122](https://doi.org/10.1093/ageing/afq122)] [Medline: [20846961](https://pubmed.ncbi.nlm.nih.gov/20846961/)]
 38. Mrosczyk-McDonald A, Savage PD, Ades PA. Handgrip strength in cardiac rehabilitation: normative values, interaction with physical function, and response to training. *J Cardiopulm Rehabil Prev* 2007;27(5):298-302. [doi: [10.1097/01.HCR.0000291297.70517.9a](https://doi.org/10.1097/01.HCR.0000291297.70517.9a)] [Medline: [17885508](https://pubmed.ncbi.nlm.nih.gov/17885508/)]
 39. Thomaes T, Thomis M, Onkelinx S, Goetschalckx K, Fagard R, Cornelissen V, et al. Muscular strength and diameter as determinants of aerobic power and aerobic power response to exercise training in CAD patients. *Acta cardiologica* 2012 Aug;1;67(4):399-406. [doi: [10.2143/AC.67.4.2170680](https://doi.org/10.2143/AC.67.4.2170680)]
 40. Sumide T, Shimada K, Ohmura H, Onishi T, Kawakami K, Masaki Y, et al. Relationship between exercise tolerance and muscle strength following cardiac rehabilitation: comparison of patients after cardiac surgery and patients with myocardial infarction. *J Cardiol* 2009 Oct;54(2):273-281 [FREE Full text] [doi: [10.1016/j.jjcc.2009.05.016](https://doi.org/10.1016/j.jjcc.2009.05.016)] [Medline: [19782265](https://pubmed.ncbi.nlm.nih.gov/19782265/)]
 41. Servantes DM, Pelcerman A, Salvetti XM, Salles AF, de APF, de SFCA, et al. Effects of home-based exercise training for patients with chronic heart failure and sleep apnoea: a randomized comparison of two different programmes. *Clin Rehabil* 2012 Jan;26(1):45-57. [doi: [10.1177/0269215511403941](https://doi.org/10.1177/0269215511403941)] [Medline: [21937519](https://pubmed.ncbi.nlm.nih.gov/21937519/)]
 42. Conraads VM, Pattyn N, De MC, Beckers PJ, Coeckelberghs E, Cornelissen VA, et al. Aerobic interval training and continuous training equally improve aerobic exercise capacity in patients with coronary artery disease: the SAINTEX-CAD study. *Int J Cardiol* 2015 Jan 20;179:203-210 [FREE Full text] [doi: [10.1016/j.ijcard.2014.10.155](https://doi.org/10.1016/j.ijcard.2014.10.155)] [Medline: [25464446](https://pubmed.ncbi.nlm.nih.gov/25464446/)]
 43. Hwang R, Bruning J, Morris N, Mandrusiak A, Russell T. A Systematic Review of the Effects of Telerehabilitation in Patients With Cardiopulmonary Diseases. *J Cardiopulm Rehabil Prev* 2015;35(6):380-389. [doi: [10.1097/HCR.0000000000000121](https://doi.org/10.1097/HCR.0000000000000121)] [Medline: [26034937](https://pubmed.ncbi.nlm.nih.gov/26034937/)]
 44. Aamot I, Forbord SH, Gustad K, Løckra V, Stensen A, Berg AT, et al. Home-based versus hospital-based high-intensity interval training in cardiac rehabilitation: a randomized study. *Eur J Prev Cardiol* 2014 Sep;21(9):1070-1078. [doi: [10.1177/2047487313488299](https://doi.org/10.1177/2047487313488299)] [Medline: [23613224](https://pubmed.ncbi.nlm.nih.gov/23613224/)]

Abbreviations

CAD: coronary artery disease
CHD: coronary heart disease
CR: cardiac rehabilitation
CVD: cardiovascular diseases
HDL: high-density lipoprotein
HG: hand grip
HOMA: homeostasis assessment
HRQoL: health-related quality of life
HRR: heart rate reserve
LDL: low-density lipoprotein
RER: respiratory exchange ratios
SRT: sitting-rising test
TRiCH: TeleRehabilitation in Coronary Heart
VO₂P: peak oxygen uptake
VT: ventilatory threshold

Edited by G Eysenbach; submitted 25.01.18; peer-reviewed by JJ Kraal, R Widmer; comments to author 15.02.18; revised version received 12.04.18; accepted 24.04.18; published 22.06.18.

Please cite as:

Avila A, Claes J, Goetschalckx K, Buys R, Azzawi M, Vanhees L, Cornelissen V

Home-Based Rehabilitation With Telemonitoring Guidance for Patients With Coronary Artery Disease (Short-Term Results of the TRiCH Study): Randomized Controlled Trial

J Med Internet Res 2018;20(6):e225

URL: <http://www.jmir.org/2018/6/e225/>

doi: [10.2196/jmir.9943](https://doi.org/10.2196/jmir.9943)

PMID: [29934286](https://pubmed.ncbi.nlm.nih.gov/29934286/)

©Andrea Avila, Jomme Claes, Kaatje Goetschalckx, Roselien Buys, May Azzawi, Luc Vanhees, Véronique Cornelissen. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.06.2018. 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.

Original Paper

Effects of the Digital Transformation: Qualitative Study on the Disturbances and Limitations of Using Video Visits in Outpatient Care

Linda Sturesson^{1*}, MA; Kristina Groth^{2,3*}, MSc, PhD

¹Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

²Centre for Innovation, Karolinska University Hospital, Stockholm, Sweden

³Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

* all authors contributed equally

Corresponding Author:

Kristina Groth, MSc, PhD

Centre for Innovation

Karolinska University Hospital

C1:77 Huddinge

Stockholm,

Sweden

Phone: 46 700856664

Email: kristina.groth@sll.se

Abstract

Background: Video mediated meetings with patients were introduced in outpatient care at a hospital in Sweden. New behaviours and tasks emerged due to changes of roles, work processes and responsibilities. The study investigates effects of digital transformation, in this case how video visits in outpatient care change work processes and introduces new tasks, in order to further improve the concept of video visits.

Objective: Through real-time, social interactional features of preparing for and conducting video visits, the study examines clinicians' perceived limitations and disturbances, and how the conditions between patients and clinicians may change when using video visits instead of face-to-face meetings in outpatient care.

Methods: Qualitative methods have been used including 14 observations of video visits at two different clinics and 14 followup interviews with clinicians. Transcriptions of interviews and field notes were thematically analysed, discussed and synthesised into themes.

Results: Disturbances and limitations related to the technology were related to time; a flexibility to schedule the meeting unbound of place, frustrations when the other part was late for the scheduled meeting, and that more experienced users of video visits usually waited longer before logging in. They were also related to sound; problems getting the sound to work satisfactory during the video visits, and problems with the image. Disturbances and limitations related to the surroundings were related to both the patient's and the clinician's environment; the principle of video technology in itself may affect the experience and the content of the consultation, and the surrounding chosen changes the conditions for and reduces the participants' field of view.

Conclusions: We could see 1) a transformation of roles and responsibilities when turning from face-to-face meetings to video visits, 2) that video visits add new circumstances, with a risk of introducing disturbances and limitations, that in turn affects the content of the meeting, 3) that avoiding negative disturbances during a video visit, requires a sensibility from the clinician's side as well as a trust in the patient's judgement, 4) that both expected and unexpected disturbances and limitations during a video visit affect the clinician's behaviour, feelings, the content of the meeting and how the clinician's relate to the different components of the concept, and 5) that there is a change of roles introduced when conducting video visits, eg, the clinician taking the first line support if both (s)he and the patient encounter problems with the technology.

(*J Med Internet Res* 2018;20(6):e221) doi:[10.2196/jmir.9866](https://doi.org/10.2196/jmir.9866)

KEYWORDS

Video visit, outpatient care, disturbance, perceived limitation, telemedicine, telehealth, ethnography

Introduction

Over the last few decades, telemedicine has evolved as a solution for the healthcare sector to meet the challenges of an increasing patient population [1,2]. One telemedicine solution is video-mediated meetings or consultations (hereafter called video visits) between patients (including relatives) and clinicians. Video visits provide a new way of conducting clinical meetings, potentially reducing costs and providing services for patients living in remote or rural areas [1,3]. There is no doubt that video visits can be beneficial for different stakeholders from different perspectives [1,3], but the implementation of telemedicine also causes changes in the organization of work; tasks and processes; and identities, roles, and authority, potentially affecting the power relationships among the participants; and participants' expectations of the meeting (cf [4-7]). Recent research has shown that video visits in outpatient care settings appear to be safe, effective, and convenient; however, there are complex challenges related to their adoption by clinicians [8]. Implementation of new technology in consultations can be met with skepticism by clinicians. Some view technology as something that interferes with good clinical practice and the exercise of professional judgment [1]. In comparison to home visits, video visits carry advantages for clinicians, such as being cost- and time-efficient [9]. Satisfaction with video technology has been explored in surveys [9], but problems with technology are seldom explored in-depth [1]. Experiences regarding the implementation of video visits have not been examined very well [3], but recent research addresses the question of implementation and has generated five key recommendations concerning how technology should be introduced: iterative introduction with the involvement of staff, time for reflection with staff and patients, relations with the information and communication technology (ICT) department to establish roles and processes, understanding patients' conditions, and flexibility of use to fit patients' needs [8]. If video visits are going to be beneficial at the macro level, they need to attract users (ie, clinicians and patients) at the micro level by considering professional dimensions and meeting the users' needs. Hence, the users' interplay, expectations, experiences, and perceived benefits should be accounted for during concept development.

The technology used, together with its principles and procedures employed when conducting video visits, can produce limitations and add disturbances that may, positively or negatively, affect how the meeting is conducted and its outcome. A video visit implies a geographical separation between clinician and patient (cf [10]) with a new location added to the healthcare consultation. Introducing non-clinical settings may affect those involved, the consultation, and the outcome of the meeting because a healthcare environment manifests social orders [5]. Also, medical spaces and physical settings facilitate the maintenance of professional and patient roles. Hence, video visits introduce new spaces that may challenge or reinforce the established performances, relations, and hierarchies [5]. Petersson [11] highlighted how telehealth "allows health professionals to unfold new spaces of visibility," since it provides a link between the patient's setting and healthcare institution. However, research on telemedicine seldom includes

the place, as it appears to be considered irrelevant in the discourse of telemedicine [10]. The impact of the patient's environment and surroundings, when conducting video visits, appears to have been less thoroughly explored, even though there is evidence that the place matters [10-13]. For example, even though the complex communication of a video visit can be affected by the environment [13], the video visit concept seldom includes recommendations for the physical environment. Other aspects however are considered, such as recommendations that everyone present during a video consultation should be identified. This is because otherwise sensitive information could be disclosed to individuals not in the field of view of the webcam [12]. Privacy considerations are generally handled by the clinician during a physical meeting, but it might be more difficult when the patient is located in another place due to the limited view of the camera. What happens in the patient's environment and surroundings may cause disturbances and create limitations, which can sometimes be easily resolved [8], but may also affect the outcome of the consultation. It is well known that the physical environment affects patients' satisfaction levels, attitudes, and work performances during the meeting [14,15]. When the space of care is no longer shared physically, and the connection between the clinician and patient is mediated by technology, new questions arise: What might take place that cannot be seen? What disturbances and limitations may occur? What effects can such disturbances and limitations cause?

The aim of this study was to explore the disturbances and limitations experienced by clinicians when conducting video visits and how these disturbances and limitations affected roles, content, and perceptions. We explored what happened when the space of care was spatially shared between two environments and mediated by video, with the overarching goal of improving the concept of video visits, thereby meeting the need for a more efficient care. The study was conducted in an outpatient care center at a university hospital in Sweden.

Methods

The study was qualitative and explorative in its approach. Interviews with clinicians and observations of video visits were conducted to generate data. The focus was on the situatedness in the use of video visits and situated actions when clinicians conducted such visits. Additionally, informal and formal settings in everyday work and ad hoc individual conversations related to video visits were observed and used to understand the phenomenon of video visits and their role in a wider context.

Approach to the Research Area

Theoretical perspectives in symbolic interactionism provided a source of inspiration and a starting point, with frameworks suitable for analyzing the social reality and understanding human behavior and human feelings. The social interaction can be influenced by moods, weather, locations, and environments. The individual defines the situation both consciously and unconsciously and human behavior is seen in relation to the whole context [16]. Diversity, as well as commonalities, are sought with an open mind, with attention given "to what falls out of view or falls between the cracks" [17]. In our study, the

video visits were part of a treatment program that included several consecutive meetings. The consultation is a social interaction between a clinician and a patient and/or relative, where at least one of them has a predetermined goal for the meeting. What happens among those involved can be understood as social acting and, more specifically, as an instrumental or planned action. For example, a clinician may have the goal of discovering the patient's behavior since the last meeting, progress, or side effects, etc. To achieve this goal, the clinician will prepare by reading the patient's medical record and making notes on what to address during the consultation. However, each consultation session is a link in a longer treatment chain—a path where each situation affects the outcome of each session (cf [18]).

Clinicians develop skills based on physical consultations, and their face-to-face visits become the norm for clinical meetings [7]. The clinicians' frame of reference is, thus, the traditional physical meeting or a follow-up by phone. When introducing video visits, clinicians are therefore likely to compare video visits to traditional clinical meetings. Anything that is perceived as a deviation from the norm can be understood as a disturbance. A disturbance is defined as something that differs from the norm, from what clinicians are used to, and can be perceived as negative or positive. Disturbances can cause perceived limitations, which can be seen as disadvantages or advantages. In our study, we defined a limitation as an abstract feeling or experience of something not being enough or being a restriction. For example, surroundings or situations can serve as limitations. Limitations can lead to disturbances and disturbances can lead to limitations; there is an interplay between the two concepts. Disturbances and limitations can be unexpected in some cases and planned for in others. The technology, procedure, and principles of video visits can cause both disturbances and limitations.

In our study, we explored video visits by gathering examples of disturbances, both negative and positive, and limitations (as perceived by clinicians, not by patients/relatives). We examined

what gave rise to the disturbance, how the disturbance was interpreted, and its consequences. This includes understanding the video visits as present situations happening in a context including space and place, even though the participants are located at a distance from each other.

Ethical Approval and Consent

Ethical approval for the study was given by the Regional Ethical Review Board in Stockholm before data gathering (reference number: 2016/1027-31). The clinicians obtained written informed consent for participation and for publication (including information about participation, anonymity, purpose and objectives of the study, and responsible researcher) from patients, relatives, and guardians. Participants were offered video visits instead of physical meetings. The consent form was either sent by email or given by hand to the patient/relatives. The clinicians signed a written consent for participation following review by the researcher.

Context

Two patient flows were involved in this study, named Clinic A and Clinic B. Both clinics treated patients with obesity. The clinics had congruent goals, agendas, and philosophies for their treatment. The content of care was mainly based on a humanistic perspective of health and disease, with lesser focus on biomedical data such as weight and body composition. However, these variables were still used as treatment outcome assessments. Video visits at the clinics were part of a treatment program that included several consecutive meetings aimed at helping patients to successfully implement lifestyle changes. Clinicians supported patients in their efforts to achieve behavioral and lifestyle changes. Between visits, patients were asked to work actively on lifestyle changes by themselves. Both clinics shared the same view about using video visits as complements to face-to-face visits and for follow-ups. The staff consisted of doctors, nurses, psychologists, nutritionists, occupational therapists, and physiotherapists. Clinicians at Clinic B also had competence in cognitive behavior therapy. Differences between the clinics are described in Table 1.

Table 1. Comparison of patient population, implementation stages, and settings of Clinic A and Clinic B.

Aspect of the setting	Clinic A	Clinic B
Patient population	Children and adolescents with obesity (2–18 years old)	Adults with obesity (>18 years)
Responsible for and involved in the treatment	Relatives were responsible for and provided an important role in the treatment. Relatives of young children visited the clinic together with the child. Follow-ups and reconciliations were made over phone with relatives of young children and not with the child. Teenage patients were assessed by the clinicians to decide if they were mature enough to take responsibility for their own treatment. If so, the relative usually did not participate in follow-ups.	Patients were responsible for their own treatment. Relatives were not present during meetings.
Stages of implementation of video visits	Video visits began when the research study started.	6-month history of carrying out video visits
Setting for video visits	One room was used for video visits. The room was equipped with a computer, camera, and headset. Clinicians booked the room before the video visits.	The clinicians used their own room, with their computer equipped with camera and a headset.

Table 2. Number of observations and interviews conducted at Clinic A and Clinic B, with clinicians, patients, relatives, or both patients and relatives.

Method	Total	Clinic A					Clinic B		
		Total	Clinician	Patient	Relative	Both	Total	Clinician	Patient
Observation	13	9	6	5	3	1	4	2	4
Interview	14	10	6	— ^a	—	—	4	2	—

^aPatients or relatives were not interviewed.

Technology and Devices

The concept used for the video visits was developed for less complex meetings in outpatient care. The technology included an ordinary video conferencing tool with encrypted communication, capable of producing adequate quality for seeing and hearing each other and for sharing documents. The technology could not be used to connect sensors used for monitoring parameters, and the quality of the video was not high enough to provide details of, for example, skin issues. A complex video visit, such as when a neurologist needs to see small detailed movements during care for patients with Parkinson's disease or demonstrate exercises to a patient [19], may require equipment of higher quality, as well as additional space in front of the video camera for specific exercises. The patient/relative typically used his or her own device, such as a computer, mobile phone, or tablet, with a webcam, speaker, internet connection and web browser, or the video conferencing app.

Respondents and Recruitment

In preparation for the study, two clinics were selected to participate. They were identified from the second author's work of introducing video visits in outpatient care settings at hospitals. One clinic was selected because it had successfully adopted the concept of video visits earlier in the year. The other clinic was selected because it had shown interest and carried out test video visits, but had not yet started. Two clinicians from the first clinic, already conducting video visits, and six clinicians from the second clinic, who wanted to start video visits, agreed to participate in the study.

At both clinics, the staff selected patients or relatives for video visits. Video visits were only offered to patients who had physically presented to the clinic at the beginning of their treatment. The clinicians offered video visits to the selected patients either during a physical meeting or through a telephone contact. The patients had the opportunity to accept or decline video visits. During the study period, there were patients who declined. The clinicians who conducted the video visits had previously met face-to-face with the patients. Selecting patients for video visits and the criteria used in the process have been described in a paper sent for publication (Sturesson and Groth, in preparation).

If patients accepted a video visit, the clinicians asked them if they wanted to participate in the research study. The question was asked of the patient/relative during a face-to-face meeting, a phone call, or a previous video visit (at Clinic B, where video visits were used before the research study started). The staff, patients, and any guardian provided written informed consent to participate in the research.

Data Collection

The data collection consisted of a total of 13 observations and 14 interviews; see Table 2 for more details. Six of the clinicians conducted two video visits each and were, therefore, observed and interviewed twice. However, one of the interviews was conducted without an observation (see below), resulting in a total of 13 observations and 14 interviews.

Each observation started before the actual video visit and included the time for the clinician's immediate preparation. The researcher was located in the same room as the clinician and was visually and verbally presented to the patient/relative at the beginning of the video visit, giving each patient a chance to withdraw his or her consent. During the video visit, the researcher observed the meeting from a position out of sight of the webcam, that is, the patient/relative could not see the researcher. The observations were partly exploratory and partly structured. Some aspects, such as start and end time, patient's location, and number of participants, were predetermined and noted in the observation protocol. These were combined with field diaries that contained the exploratory observation notes. The observations were not recorded, photographed, or filmed.

The interviews were in-depth, contextual, and semi-structured and were conducted with the clinicians after, and in addition to, each video visit. Of the 14 interviews, 13 were conducted face-to-face and one by phone. One of the interviews occurred without an observation, since the patient withdrew consent to participate in the study as the observation was about to start. The interview was still conducted after the video meeting. The interviews were recorded and transcribed verbatim.

In addition, the researcher attended formal encounters (eg, treatment conferences with clinicians) as a passive observer and participated in informal gatherings (eg, lunches and other breaks), taking field notes to capture the clinical discourse and clinicians' perceptions and thoughts about video visits and without interfering in the discussions taking place. All data were gathered during a contiguous period of 3 months during 2016.

Analysis

The analysis process followed a qualitative approach [20], in which the transcripts of interviews and field notes were read through several times to familiarize with the content. During the reading, themes were identified and noted on a blank sheet. Corresponding transcripts and field notes were read iteratively to gain a full picture of the collected data. A conceptual framework was created. After this initial process, the transcripts of interviews and field notes were thematically analyzed [21]. Data were then read through again and coded to match the themes in the developed conceptual framework.

Table 3. Being late or being on time, based on observational data on each video visit.

Observation data	Total	Clinic A										Clinic B			
Video visit number	N/A ^a	1	2	3	4	5	6	7	8	9	10	11	12	13	14
The clinician was logged in on time	9	Y ^b	Y	Y	N ^c	Y	N	Y	N	N	Y	Y	N	Y	Y
The patient/relative was logged in when the clinician logged into the virtual meeting room.	3	N	N	N	N	N	N	N	N	Y	N	N	Y	N	Y
The clinician calls patient/relative	6	Y	N	Y	Y	N	Y	Y	N	N	Y	N	N	N	N
The video visit started at the scheduled time	4	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	Y
The video visit started later than the scheduled time	10	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N
Number of minutes after scheduled time that the video visit started	— ^d	—	—	12	26	—	23	19	9	3	28	3	3	6	—

^aN/A: not applicable.^bY=Yes.^cN=No.^d—" indicates that the video visit started on time.

Spreadsheets were used to organize and sort the data. To find and keep track of patterns in the material, themes were separated into different rows in the spreadsheet, and each interview and the corresponding observation was sorted into different columns. Pieces of the text were sorted to the appropriate cells. The principle of spreadsheets was also used to analyze and find patterns in the quantitative data (such as the data described in Table 3). The data, themes, and sorting were continuously discussed throughout the analysis.

Video visit number 6 only consisted of an interview since the patient declined to participate in the study at the last minute. The interview was conducted immediately after the video visit.

The themes were synthesized into two overarching categories: "Selecting patients for video visits" and "Disturbances and limitations." From the analysis, it became clear that selecting patients has added a new task for clinicians and that video meetings have introduced disturbances and limitations related to both the technology and surroundings. This paper focuses on the second category: Disturbances and limitations. The themes sorted under this category were issues of time and senses related to the technology used and issues of space and place related to the surroundings used by the clinicians and patients/relatives.

In the Results section, quotes are used to illustrate situations causing a disturbance or where a limitation was identified. The quotes chosen represent situations that occurred once or several times, illustrating an effect of something that may occur in other situations. When illustrating a situation related to the theme with excerpts from the data, we use the notation Clinic X, Int_Y, or Obs_Y, where Int stands for interview and Obs for observation. The interview that followed an observation of a video visit was given the same number as the observation, that is, the number of the video visit.

Results

Overview

We have identified several situations wherein clinicians experienced disturbances and limitations when conducting video

visits. Disturbances and limitations were related either to the technology or surroundings.

As described above, we defined a disturbance as something that differed from the norm and from what the clinicians were used to. Disturbances can cause perceived limitations and can be perceived as negative but also, on occasion, as positive by participants. In addition, the technology and principles of video visits occasionally caused limitations. As pointed out in the Methods, we defined limitations as an abstract feeling or experience of something not being enough or as something being restricted due to, for example, the surroundings or situation.

Disturbances and Limitations Due to the Technology

We identified disturbances and limitations related directly or indirectly to the technology used during video visits, which are presented based on two themes: *Time* and *Senses*. The time-related aspects presented below are *Flexibility in Time and Space*, *Being Late and Being on Time*, and *Waiting and Logging in*. Aspects of senses presented below are *Hearing and Being Heard* and *Seeing and Being Seen*.

Time: Flexibility in Time and Space

Clinicians at both clinics saw advantages related to flexibility in time when using video visits. Video visits made it possible for those involved to schedule the meeting unbound from place, eg, if the patient lived or worked far from the clinic. Clinicians who were more experienced in using video visits also saw benefits for themselves. For example, some were able to redistribute their working hours by working longer one day, as they no longer depended on the clinic's regular business hours or the number of staff present at the clinic. According to safety regulations, at least three members of the staff needed to be present in the office when a clinician physically met with a patient. In turn, working longer one day made it possible for some clinicians to go home earlier another day. This flexibility in time could be seen as a positive disturbance.

Time: Being Late and Being on Time

During video visits, we observed several examples of worries and frustrations when the clinician or patient was late for the scheduled meeting. Table 3 specifies how many video visits started at the scheduled time and, if not, how many minutes late the visit started, who logged in on time, and whether the clinician called the patient.

Clinicians at Clinic B did not call patients who were late, whereas clinicians at Clinic A called the patients during six of the observations. The clinicians waited up to 7 minutes before calling if the patient was late. The delays observed were caused by (time) planning aspects, external circumstances, and technical issues. The clinicians always sent the number and code for the web-based video room to the patient before the meeting took place, but the routine differed among the clinicians with regard to how far in advance they sent the code in relation to the scheduled meeting. Some sent the information when the meeting was booked, and some sent it the same day as the meeting. The patients learned their clinicians' routines.

In one of our observations, the clinician sent the information to the patient several hours later than usual. The information was actually sent after the meeting was scheduled to start. This caused the patient to become worried and confused about the time of the video visit (Clinic B, Obs_12). When a patient physically arrives at a clinic, he or she reports to the reception. This may give the patient a sense of security about not being forgotten. In another of our observations, the video visit started 20 minutes late because a previous face-to-face meeting ended later than planned. This was distressing to the clinician, who did not see any opportunity to notify the subsequent patient of the unexpected delay. When the meeting started, the clinician perceived frustration from the patient and relative. Further, the patient had plans after the meeting, leading to a shorter meeting than that for which the clinician had planned (Clinic A, Int_6 and Obs_6).

Being late due to external circumstances could also be related to the weather or to technical problems:

One of the clinicians at Clinic A arrived at the clinic 5 minutes before the video visit was scheduled to start because of heavy snowfall causing problems in the morning traffic. The clinician started to read the patient journal. Nine minutes after the scheduled time, the clinician entered the room dedicated for video visits, logged into the web-based video room, and noticed that the patient has not yet logged in. The clinician wanted to call the patient but there was no contact number with the hospital's central server on the computer, and the patient's phone number could not, therefore, be reached. Instead, the clinician had to take out the SITHS card from the computer, walk back to her office to find the patient's contact details on her own computer, and then go back to the video visit room and log in to the computer again. The clinician called the patient, who had still not logged in. It turns out that they were logged into different web-based video rooms, something that, in principle, should not be possible. After 23 minutes, they managed to connect. [Clinic A, Obs_4]

In this case, the meeting started late because the clinician was late and also because the patient was logged into the wrong meeting room. These delays caused a ripple effect of delays, and thus, frustration and stress were experienced by all participants. Valuable time was lost and the meeting was shorter than planned.

Time: Waiting and Logging In

If they were not delayed, the clinicians at Clinic A usually logged into the web-based video room approximately 4–10 minutes before the meeting was scheduled to start. The clinicians who were more experienced at using video visits (at Clinic B) usually waited longer before logging into the web-based video room but also gave the following, somewhat contradictory, explanation for logging in 2–5 minutes in advance:

If I log in 5 minutes in advance [...] then I may have replied to an email or so, but also 2 minutes, that is too short a time to start with something, and then you just sit there and look at the screen, which feels a bit meaningless. [Clinic B, Int_11]

The same clinician noted in the interview that she became restless while sitting and waiting for the video visit to start. She likened it to waiting for a bus and stated that she starts to think about other things and then suddenly realizes that the patient has logged in. Another clinician, from the same clinic, reported similar reflections. She reportedly felt stressed if she logged in too early because she then had difficulties doing other tasks, since she needed to keep an eye on the screen to be able to see when the patient logged in (Clinic B, Int_13).

Another clinician reflected that, if a patient was delayed for a face-to-face meeting, then they could do other things because they were in their own office and could be notified by reception when the patient arrived. In a video visit, the clinician had to actively check the screen now and then while waiting (Clinic A, Int_3).

Hence, video visits introduced limitations related to time management while waiting for the patient to log in. One strategy clinicians at Clinic B developed to overcome this was to log in as late as possible.

Senses: Hearing and Being Heard

On several occasions, there were problems getting the sound to work satisfactorily during video visits, affecting both the clinicians and the video visit itself. Problems with sound affected communication in that clinicians had to repeat themselves and ask “what?” more often. In some cases, system usability caused problems, in others, it was connectivity. In both cases, clinicians had to take on the role of first-line support, trying to figure out the cause of the problem: “The clinician asked if the patient has an icon for sound on the phone. The patient nodded and the clinician said: ‘try to push it [...] I can see that the sound is not on, from your side’” (Clinic A, Int_3).

Problems with the sound caused irritation: “it’s happened so many times that the sound hasn’t been working and it is a frustration” (Clinic A, Int_1). In one case, the clinician made a joke of it and blamed the bad weather and they then talked about a technical solution, discussing volume settings and differences

between conducting video visits using a mobile phone and a laptop (Clinic B, Obs_13).

Problems with the technology not working as expected created barriers to the adoption of video visits. At the same time, when the sound did not work as expected, clinicians created situation-based solutions to these in situ problems. At Clinic A, clinicians would bring a phone to the video visit room: “The sound didn’t work and it’s happened once before [...] but now I’ve solved this by having a phone with a speaker during the video visit” (Clinic A, Int_3). The in situ problem thus led to a pre-conceived notion of sound problems, for which the clinicians planned in advance. We observed that sound-related problems were shared with colleagues during breaks, resulting in clinicians who had never performed a video visit taking a phone when they began to carry out video visits.

Senses: Seeing and Being Seen

Video visits differed from phone consultations in that the clinician and patient could see each other. During video visits, the image was sometimes missing at the beginning of the visit or it sometimes disappeared during the visit for various reasons. This affected the clinician and the video visit in different ways.

One example illustrates how this could affect the clinician: “If the screen closes down all the time [...] if you have poor connectivity, it might shut down like that and it can be a bit annoying because then maybe you’re talking about that instead ‘oh, now you’ve disappeared, but now you’re back’” (Clinic B, Int_13). In one case, the same clinician conducted a video visit and anticipated bad connectivity. She adapted the video visit to technical circumstances by planning the session’s content differently, knowing that visits like these tended to be “more choppy,” and she also lowered her own expectations; thus, the video visit became more “undemanding.”

Video problems could occur if the user was inactive on the computer, resulting in a locked screen mode, but with the web-based video meeting room still running. The clinician and the patient/relative could still hear each other but not see each other. For a clinician new to video visits, this could be disturbing, not knowing what happened or how to get back into the meeting. To avoid this, the experienced clinician moved the mouse quickly and entered the log in code, without affecting the meeting.

In one of our observations, a patient/relative suddenly disappeared from the screen but it was not related to being inactive. In this case, it may have been intentional (see example from Clinic A, Obs_1 under *the patient’s surrounding*, in the next section), but such a disturbance could also be caused by technology; for example, if the bandwidth goes down. Another reason for the patient/relative disappearing from sight involved the clinician sharing a presentation, eg, of a growth curve, or when a clinician opened a medical journal or calendar on the screen, thereby hiding the video conferencing window. This kind of planned disturbance could still cause inconveniences or confusion when the shared presentation was ended or the other windows were closed. The patient/relative could, if using a mobile phone or tablet, be in a different surrounding with other disturbances and limitations than before (Clinic A, Obs_9).

Technology caused limitations and disturbances that were perceived as negative from the clinician’s point of view. On the other hand, video visits also served as an extended eye for the clinician and were perceived as something positive since they provided the opportunity to see the patient’s context: their home, workplace, or school. This was expressed as something that provided more insight into the patient’s personality and the feeling of achieving a closer contact: “now I notice that I get a more personal picture of the person when I come into their home, because then suddenly it becomes real [...] it becomes a closer contact somehow” (Clinic B, Int_13). The same clinician also said,

You can kind of take part in their private life. Then, whether it’s important, I can’t say. I think it can have a kind of effect for the compliance ... maybe becomes positive when I say... what a nice color, and you see that everybody becomes happy. It’s really something personal, and it’s just small comments... I think it can be important for compliance, but I don’t know if it has any effect on the treatment.” [Clinic B, Int_13]

The clinician experienced this differentially in comparison to face-to-face visits and perceived it as a positive aspect of video visits. The extent to which the patient saw this as a positive or negative disturbance or limitation was not stated.

Disturbances and Limitations Due to the Surroundings

We identified disturbances and limitations related directly or indirectly to the surroundings used during video visits, which are presented based on the themes *the patient’s surroundings* and *the clinician’s surroundings*. The surroundings are related to both place and space.

The Patient’s Surroundings

Video visits imply that the patient and clinician are in geographically different locations. Using videos as a tool for communication brings a new field of view to the healthcare meeting: the patient’s/relative’s physical environment. When the clinician sees the patient in his or her environment and, at the same time, uses the sense of hearing, the patient’s/relative’s surroundings are brought to the meeting. What is happening in a specific situation, that can be both seen and not seen, is interpreted and affects both the clinician and patient, their actions, the content of the meeting, and its outcome.

Hence, the principle of video technology, which makes it possible to see and hear, may affect the experience and content of the consultation, as can be seen in the following example:

The clinician is conducting a video visit with the relative of a younger patient. After 10 minutes’ discussion, it becomes clear that the patient has gained a relatively large amount of weight since the last meeting. When the clinician asks “why?,” the relative said, “I’ve no idea, don’t ask me.” They continued to talk about the weight gain, about what may have happened, and so on. Suddenly the screen turned black for a while and when it came back the audio on the phone was off (the phone was being used because of audio problems with the video connection). The clinician called the relative and they came back into the meeting, but soon, the screen again turned black, this time for a couple of minutes. When the clinician called

back, the relative said they needed to end the meeting because the patient became sick and required care (Clinic A, Obs_1).

In the follow-up interview, the clinician said that she found it difficult to be sure whether it was a technical problem that caused the interruption or it was caused intentionally by the relative or the patient because of an unwillingness to discuss his or her recent weight gain. The clinician said that she found it peculiar that the patient suddenly became sick, but since she could neither see nor hear the patient during the video visit, she did not know what was going on in the patient's surroundings (Clinic A, Int_1).

In another example, the patient's surroundings provided a space that was unknown to the clinician other than that the patient would be in school: The video visit began and the patient, a teenager, was outside. From knowing that the patient would be in school, the clinician assumed the patient was in the schoolyard. The clinician felt that the patient had a roving eye, and thus assumed the patient "felt uncomfortable in the situation and uncertain; either [the patient] was scanning the environment [...] or was just uncomfortable with the situation of sitting and looking into the camera" (Clinic A, Obs_8 and Int_8).

The clinician was not aware of whether there were other people in the surrounding area. However, the clinician's assumption of a schoolyard may have affected the interpretation of the patient's behavior, creating a sense of disturbance in the consultation: "I believe that [the patient] was perhaps a little bit stressed" (Clinic A, Int_8). As a result, the clinician did not ask further questions about sensitive subjects and "did not stay with every question as long as usual" (Clinic A, Int_8). Also, the clinician knew the patient was going back to class "so there was a time aspect as well, somehow the teacher was expecting him back."

The environment may be a disturbing component of the meeting, in an interplay with what the clinicians can see, what they interpret about the unseen environment from the patient's behavior, and time, all of which are bound to the patient's environment. These three issues affected the actions of the clinician and limited the content of the consultation. In some cases, the clinician asked beforehand where the patient would be during the video visit. In one case, the patient answered the video call when in a locker room. The clinician asked if they might be disturbed, but the patient, a teenager, said the risk of disturbance was low. However, the patient and clinician agreed to use a code word if someone entered the locker room during the video visit, so they could close down the meeting without the person entering understanding the meeting's content.

During one observation, the patient was in a parked car. Something happened during the video visit, something the clinician in the follow-up interview described as a "glitch," which she interpreted as someone entering the car. Since the meeting time was almost over, the clinician ended it a little more quickly than she would have done if it had been a face-to-face meeting. The clinician noticed that she did not receive the same focus after the "glitch," and she felt that the patient was indicating that the meeting time was almost over (Clinic A, Obs_2 and Int_2).

In another example, the patient, a child, was participating in a video visit at home when all of a sudden the child looked to the side several times and began to smile and laugh. The clinician asked the patient if there was another person who wanted to participate in the meeting. The patient said the younger sibling's name (Clinic A, Obs_7). The clinician perceived this as a disturbance and said that the patient made signs to the sibling to send them out of the room:

I saw [...] what happened there, talking about what you see and can't see [...] but then I wonder, [The relative] was clear with, ok now I'll leave you [...] you can sit in peace and quiet [...], but now the sibling [...] I could say that, the next time, maybe ask [the relative] [...] to talk to [the sibling] and say that this is [the patients'] time so he doesn't feel that [the sibling] is standing there and listening. [Clinic A, Int_7]

These examples imply that the place itself may not only affect the meeting; rather, the meeting can be affected by events that occur in that place. Two of the interviews exemplified how the clinicians planned for this beforehand:

You can prepare when you book the video visit, [and ask the patient] which environment do you feel safe in [...] where do you feel you can talk freely? Is it at home or [...] a private room at the library or whatever? [Clinic A, Int_8]

To minimize moments of distraction in the same way as if you were meeting here, you close the space around you so it's only you and the patient. If he was sitting in a kitchen with younger siblings around, then I would recommend that he goes to a calm place where it's easier for him to focus. Whether it's his room or anywhere else, that doesn't matter. [Clinic A, Int_7]

The Clinician's Surroundings

The clinicians at Clinic A all conducted video visits in one room dedicated to such visits, also used as a patient kitchen during face-to-face meetings. At Clinic B, the clinicians conducted video visits in their own rooms. This resulted in clinicians at Clinic A having to perform the extra task of double-checking the availability of the room before scheduling a video visit, something that was perceived as a disturbance or a barrier for video visits.

A video visit adds in the patient's and/or relative's environment and surroundings, but through the video technology used, it also changes the conditions and reduces the patients' field of view. For clinicians at Clinic B, who were using their own rooms and were more experienced, insights about the field of view also affected their behavior. For example, we observed two cases in which clinicians brought coffee into the video visit. When asked about this, the clinician stated that this would never happen if the patient had visited face-to-face as that would not be professional. However, since the clinician knew that the coffee was not visible to the patient, it was perceived as acceptable to have it in the room during the video visit.

Table 4. Summary of findings.

Theme	Disturbances	Limitations
Time		
Flexibility in time and space	<ul style="list-style-type: none"> Unbound from place 	<ul style="list-style-type: none"> Possibility to schedule meetings outside the opening hours of reception (Positive) Easier to schedule meetings with patients living far from the clinic (Positive)
Being late and being on time	<ul style="list-style-type: none"> Frustration 	<ul style="list-style-type: none"> Lack of functionality to communicate delays (Negative)
Waiting and logging in	<ul style="list-style-type: none"> “Dead-time” when waiting for the meeting to start Need to actively check if the patient has logged in 	<ul style="list-style-type: none"> Lack of functionality to be alerted when other parties are logged in (Negative)
Senses		
Hearing and being heard	<ul style="list-style-type: none"> Technology problems Technology problems shared among colleagues 	<ul style="list-style-type: none"> Barriers to adopting video visits (Negative) Adjust the meeting to known problems (Negative)
Seeing and being seen	<ul style="list-style-type: none"> Being inactive on the computer Technology issues known beforehand Patient disappearing out of sight Extending the eye of the clinician into the patient’s context 	<ul style="list-style-type: none"> Locked screen mode (Negative) Adjust the plans of the meeting (Negative) Difficult for clinician to understand the cause of this action (Negative) Gaining insight into the patient’s personality, adding a feeling of getting closer to the patient (Positive)
Patient’s surroundings	<ul style="list-style-type: none"> Patient disappearing out of sight Field of view changes Unknown spaces where disturbances are difficult to understand The environment itself and what can and cannot be seen 	<ul style="list-style-type: none"> Lack of understanding about the cause of action (Negative) Adding the context of the patient (Positive) The content of the meeting (Negative)
Clinician’s surroundings	<ul style="list-style-type: none"> The need to check the availability of room Decrease in the patient’s field of view 	<ul style="list-style-type: none"> Barriers to adopting video visits (Negative) Only parts of the room needed to appear professional (Positive) Only visible clothing needed to be professional (Positive)

Another example was when a clinician took off her shoes during a meeting (Clinic B, Obs_13). In the follow-up interview, the clinician stressed that this was not something that would happen in a face-to-face visit; then, she would always keep her shoes on. Also, one clinician said that she did not need to bother about a clean desk, as long as it was out of sight of the patient: “they don’t see my desk and that’s pretty nice, because then I can have it a little messier” (Clinic B, Int_11). She also noted not having to put private belongings somewhere else; a suitcase was standing in the middle of the room during the video visit and her coat was on a hanger. If a patient was visiting face-to-face, these objects would be hidden. This was seen as a part of being a professional, in the sense that patients should not be exposed to items from the clinician’s personal life including pictures, coats, and bags. Similarly, one of the clinicians at Clinic B said that when she used video visits, she used the opportunity to stand up at her (height adjustable) desk while using the computer. The awareness of the patient’s field of view and how the surroundings and behavior could change compared to a face-to-face meeting reportedly came with

experience, when the focus was not on becoming familiar with the technology.

Summary of Findings

Table 4 summarizes the findings based on the theme for each category and what disturbance caused what limitation. The text in parentheses denotes whether the disturbances and limitations were interpreted as positive or negative.

Discussion

Principal Findings

Our results show that a number of disturbances and perceived limitations related to the technology and surroundings occurred during video visits. These, in turn, affected how both the clinicians acted before and during video visits and the content of the visits themselves. Disturbances and perceived limitations were added to the healthcare consultation when using video communication. These disturbances and perceived limitations were caused by the technology and the patients’ and clinicians’

locations during the consultation (cf [8]). Our findings about disturbances and limitations that were related to technology addressed time (aspects of flexibility, being late or on time, waiting and logging in) and senses (hearing and seeing). The disturbances they experienced were mostly perceived as negative by the clinicians, and the limitations were mostly seen as disadvantages. However, there were some exceptions. For example, clinicians perceived video visits as easier to schedule, and the limited field of view affected the behavior of the clinician (eg, being relaxed enough not to wear shoes or to have a cup of coffee on the table). The situations we observed may not occur frequently, but they all illustrate situations that should be considered, since they may happen. In the following sections, we expand upon some features of our study to connect with topics that have been less thoroughly addressed in the literature: responsibilities that change due to a new place and space, location-related consequences, and how video visits can be improved. Our findings are in line with the “key recommendations” given by Greenhalgh et al. [8], for example, clinicians need time to develop new tasks and changes in work processes that come with a new concept. Further, new support processes and roles need to be developed and established, and there is an ongoing need to understand patients’ conditions and need for flexibility.

From One Place to Another—Transformation of Responsibility

From our results, we can see that place still matters when using telemedicine solutions (cf [10]). The choice of place and environment for conducting video visits is important, especially as non-medical settings are added to the clinical environment (cf [15]). A video visit extends the clinical space and brings new context to a consultation, a context that leads to assumptions when a situation is interpreted. The disturbances and limitations experienced generate knowledge about what is needed to avoid them, thereby creating new requirements. The process of selecting patients for video visits and the criteria used in that process (Sturesson and Groth, in preparation) can be understood as actions to minimize the risk of disturbances and limitations in advance. Thus, to avoid disturbances and limitations, the clinicians develop requirements for the patient’s environment, or rather for the patient’s space, wanting this to be secure and comfortable for the patient (cf [8,12]). Since a video visit involves social interaction between people in different contexts, mediated by video technology, disturbances and limitations cannot always be avoided.

Both the place and the patient’s feelings in that place are important. As long as the patient feels secure and comfortable with the situation, and as long as the clinician interprets the situation as such, the patient can be at home, at work, in school, outside, at the library, in a car, or at a bus stop. Therefore, the place itself may not be a problem, as long as the risk of interruption can be minimized. However, choosing a place outside may introduce disturbances and limitations to the meeting (eg, due to background noise that might affect the ability to hear or cause other unpredictable interruptions). Patients who are new to video visits may not be aware of how their choice of space may influence the meeting. Clinicians can guide their patients in choosing a place, but the final

responsibility lies in the hands of the patient. We have observed patients who were in the schoolyard, in a car, and at home in the kitchen or the bedroom. Hence, the environment and specific place where the video visit is conducted is complex and is affected by many factors that need to be accounted for (cf [13]). We have observed and described a number of situations in which the place selected by the patient caused disturbances and limitations. It is difficult for the clinician to control the patient’s space, but they can inform the patient of the requirements that need to be fulfilled to ensure the best outcome for the consultation.

There is a responsibility to select and provide a secure and comfortable environment in which the video visit can be conducted. In traditional face-to-face meetings, this responsibility is entirely the clinician’s. However, introducing video visits transfers part of this responsibility to the patient, that is, to choose a space in an environment where the patient can feel safe and comfortable for the purpose of meeting. Video visits also introduce new environmental consideration, not only in terms of being able to see and hear each other, but also what may happen in those places.

The clinicians video visit experience are aware of what can happen if the patient has not selected an optimal place for the video visit, and they can offer advice before a patient uses video visits for the first time. Clinicians who are only starting to use video visits as an alternative to physical meetings may be occupied getting everything to work, focusing on the technology, and may need support on this matter (eg, through check lists; cf [8]).

Consequences of the Choice of Place

The video visits take place in a specific situation during a treatment program, and every visit can be seen as a social interaction between the clinician and the patient. The care is patient-centered and the meetings are based on topics the patient selects. During a meeting, the clinician thus not only adapts the content of the video visit to the patient but also to other external circumstances, that is, parts of what is happening that are not planned in advance (cf [18]). Therefore, the treatment sessions cannot be planned in detail. Using video communication adds new circumstances that can affect the content of the meeting (cf [5]). This complexity makes it important to minimize the risk of disturbances and interruptions, which may affect the meeting in a negative way.

The place, or rather what is happening in the place, that the patient has chosen can have a significant effect on the meeting and the treatment program (cf [11,13,15]). The clinician can only see parts of the place from which the patient participates. That is, for the clinician, the place includes one visible and one invisible space. Even though the patient has a responsibility for the chosen place, the clinician is still responsible for adjusting the meeting content to the selected space to meet the requirements of the environment. This requires sensitivity from the clinician’s side as well as trust in the patient’s judgment to avoid negative disturbances in terms of unaccounted limitations.

The limitations of the invisible space at the clinician’s location may be an advantage for the clinician. The physical place or

the room used for video visits does not have to be designed for patient visits. Only the space visible to the video camera needs to provide a professional clinical environment. Hence, there are fewer requirements for the environment that is not part of the meeting space. In addition, the opportunity to see the environment and context of the patient's life was considered a positive consequence of video visits because it can provide insights into the patient's personality and the feeling of gaining a closer contact with the patient. However, this may also affect the patient in making his or her home less of a private sphere [10].

Improvements in the Concept of Video Visits

Implementing video visits in outpatient care requires new ways of working (eg, developing criteria for patient selection). Also, new tasks are introduced when using new technologies (eg, for video communication), and a new division of labor is introduced (eg, when responsibility for one of the locations used is transferred to the patient). These changes threaten and disrupt spatial, professional, and organizational orders in the work and organization (cf [5]). These changes also introduce disturbances and limitations through both the technology (cf [5]) and surroundings (cf [8]), which need to be addressed and managed through an interplay between the technology and locations used during the video visits and the participants in the meeting.

Implications of Flexibility in Time and Space

In general, video visits, in comparison to face-to-face meetings, provide flexibility in time and space. This flexibility means that video visits may increase clinicians' power over their own working time, giving them more control over their working situation. In the long run, this could reduce stress. However, video visits may also lead to stress if the visits are not easy to carry out and can thus lead to disturbances and limitations. Our analysis showed that being late caused an in situ disturbance and a perceived limitation for the clinician of not being able to communicate the delay to the patient, since there was no process or functionality for this. If the patient had been physically at the clinic, a receptionist or secretary would probably have been able to inform the patient about the delay. The feeling of not being able to notify the patient was stressful and frustrating for the clinician. The clinicians also sensed frustration from their patients in that specific situation. Hence, the virtual meeting place lacks functionality that is usually available in a physical meeting place, such as a meeting room with all its services. The place itself is important and affects the satisfaction, attitudes, and work performance of all the participants, not only the patient (cf [15]).

The awareness of being late or being on time is reciprocal. When the patient was not logged in at the scheduled time, this created uncertainty regarding if the patient had forgotten the video visit, the patient had not received the credentials to access the virtual room, or the patient was experiencing problems with technology. This uncertainty may have increased the need for the clinician to be on time and thus reduced temporal flexibility. Logging in earlier imposed a limitation on performing other duties, which in some cases led to feelings of stress and restlessness. The clinicians who were comfortable with the technology and with video visits, and who knew that their patients could also operate

the technology, preferred to log in only a couple of minutes before the scheduled time of the video visit. Thus, temporal flexibility increased as participants became more comfortable with the concept.

Flexibility of place is more obvious, but has different dimensions. While logged in and waiting for the patient to log in, the clinician's spatial flexibility and opportunities to engage in other tasks are reduced to the space captured by the video conferencing system. To notice when the patient logs in, the clinician needs to pay attention to the video conferencing window on the screen, which limits the clinician's ability to engage in other tasks. Such limitations can be reduced if the clinician is experienced and is using his or her own office, making it possible to complete other tasks while waiting. Technical functionality could further reduce these limitations, for example, by sending a notification when the patient has logged in.

Flexibility of time and place is hence relative and exhibits different dimensions that relate to the clinician's experiences while using video visits. This therefore affected the selection criteria used when choosing patients who are considered suitable for video visits (Sturesson and Groth, in preparation). This also needs to be considered when developing tools for, or the process of, performing video visits.

Implications of Issues Related to Usability of Audio and Video

Even when the clinician and patient were both logged in on time, difficulties with managing the technology, which were not unusual, could delay or otherwise affect the start of the consultation session (cf [1]). When these disturbances occurred, the clinician had to guide the patient to turn on the video or audio. Sometimes, the clinician had to guide the patient on how to log in as well. Hence, the clinician automatically took on the responsibility for first-line support, which also required that he or she had sufficient skills to guide the patient in overcoming obstacles, which were caused either by not reading the instructions properly or by the need for intricate user functionality, due to using an off-the-shelf product adapted to hospital security issues.

Disturbances and limitations, whether known or unknown beforehand, also affected the expectations and experiences of the video visit, for both the clinician and the patient. This in turn affected the content of the meeting. Anticipated disturbances and limitations were based on the clinician's own experiences or on their colleagues' thoughts and experience. Both anticipated and unanticipated disturbances and limitations affected the clinician's behavior and feelings, and the content of the meeting and how the clinicians related to the different components of the concept: the technology, his or her role, the patient and his or her role, the surroundings, and the content of the treatment. Through experience, clinicians developed know-how that they were able to take to the next video visit and share with their colleagues. This experience was also used when developing selection criteria to include or exclude patients for video visits. The same know-how should be used to develop advice for the patient to use when he or she selects the place for the meeting (see above about the transformation of

responsibility). Clinicians could provide this as a hand-out when offering video visits to patients.

Know-how about disturbances and limitations made it possible to adapt the video visit in advance to the technical and practical circumstances and changed personal expectations about the video visit and its outcome. Adapting the video visit to known circumstances is usually based on an individual's or colleague's experience of sudden disturbances; for example, knowing that audio will occasionally fail, clinicians always take their mobile phones to the meetings.

Implications for Further Research

Further research is required to understand the full effects of video visits. Our study provides one piece of the puzzle and can guide other researchers in studying the disturbances and limitations of the digital transformation in similar or other settings. The transformation of responsibility and a focus on patient empowerment are interesting topics in today's digital world that need further exploration. Another area that may be relevant to study, which is not addressed in this paper, is ergonomics, especially when the consultations are more complex, with a need for more detailed information.

Acknowledgments

The authors would like to thank the respondents at Clinic A and Clinic B for their time and for engaging in fruitful discussions with us. The project was co-funded by the county council of Stockholm and Vinnova.

Conflicts of Interest

None declared.

References

1. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. *BMJ Open* 2016 Jan 29;6(1):e009388 [FREE Full text] [doi: [10.1136/bmjopen-2015-009388](https://doi.org/10.1136/bmjopen-2015-009388)] [Medline: [26826147](https://pubmed.ncbi.nlm.nih.gov/26826147/)]
2. Wilson LS, Maeder AJ. Recent Directions in Telemedicine: Review of Trends in Research and Practice. *Healthc Inform Res* 2015 Oct;21(4):213-222 [FREE Full text] [doi: [10.4258/hir.2015.21.4.213](https://doi.org/10.4258/hir.2015.21.4.213)] [Medline: [26618026](https://pubmed.ncbi.nlm.nih.gov/26618026/)]
3. Powell RE, Stone D, Hollander JE. Patient and Health System Experience With Implementation of an Enterprise-Wide Telehealth Scheduled Video Visit Program: Mixed-Methods Study. *JMIR Med Inform* 2018 Feb 13;6(1):e10 [FREE Full text] [doi: [10.2196/medinform.8479](https://doi.org/10.2196/medinform.8479)] [Medline: [29439947](https://pubmed.ncbi.nlm.nih.gov/29439947/)]
4. Groth K, Olin K, Gran O, Permert J. The role of technology in video-mediated consensus meetings. *J Telemed Telecare* 2008;14(7):349-353. [doi: [10.1258/jtt.2008.007006](https://doi.org/10.1258/jtt.2008.007006)] [Medline: [18852315](https://pubmed.ncbi.nlm.nih.gov/18852315/)]
5. Halford S, Lotherington A, Obstfelder A, Dyb K. GETTING THE WHOLE PICTURE? Information, Communication & Society 2010 Apr;13(3):442-465. [doi: [10.1080/13691180903095856](https://doi.org/10.1080/13691180903095856)]
6. Nicolini D. Stretching out and expanding work practices in time and space: The case of telemedicine. *Human Relations* 2016 Apr 22;60(6):889-920. [doi: [10.1177/0018726707080080](https://doi.org/10.1177/0018726707080080)]
7. Pappas Y, Seale C. The opening phase of telemedicine consultations: an analysis of interaction. *Soc Sci Med* 2009 Apr;68(7):1229-1237. [doi: [10.1016/j.socscimed.2009.01.011](https://doi.org/10.1016/j.socscimed.2009.01.011)] [Medline: [19201514](https://pubmed.ncbi.nlm.nih.gov/19201514/)]
8. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-World Implementation of Video Outpatient Consultations at Macro, Meso, and Micro Levels: Mixed-Method Study. *J Med Internet Res* 2018 Dec 17;20(4):e150 [FREE Full text] [doi: [10.2196/jmir.9897](https://doi.org/10.2196/jmir.9897)] [Medline: [29625956](https://pubmed.ncbi.nlm.nih.gov/29625956/)]
9. Taylor A, Morris G, Pech J, Rechter S, Carati C, Kidd MR. Home Telehealth Video Conferencing: Perceptions and Performance. *JMIR Mhealth Uhealth* 2015 Sep 17;3(3):e90 [FREE Full text] [doi: [10.2196/mhealth.4666](https://doi.org/10.2196/mhealth.4666)] [Medline: [26381104](https://pubmed.ncbi.nlm.nih.gov/26381104/)]
10. Oudshoorn N. How places matter: telecare technologies and the changing spatial dimensions of healthcare. *Soc Stud Sci* 2012 Feb;42(1):121-142. [doi: [10.1177/0306312711431817](https://doi.org/10.1177/0306312711431817)] [Medline: [22530385](https://pubmed.ncbi.nlm.nih.gov/22530385/)]
11. Petersson J. Technospacialities and telehealthcare: unfolding new spaces of visibility. *Information, Communication & Society* 2015 Jul 17;19(6):824-842. [doi: [10.1080/1369118x.2015.1061575](https://doi.org/10.1080/1369118x.2015.1061575)]
12. Allen DT, Caldwell P, Komesaroff PA, Loh PK, Mozer R, Sabesan S, Royal Australasian College of Physicians Telehealth Working Group. Practical aspects of telehealth: set-up and preparation for video consultations. *Intern Med J* 2013 Oct;43(10):1133-1136. [doi: [10.1111/imj.12264](https://doi.org/10.1111/imj.12264)] [Medline: [24134169](https://pubmed.ncbi.nlm.nih.gov/24134169/)]
13. Sabesan S, Allen D, Caldwell P, Loh PK, Mozer R, Komesaroff PA, Royal Australasian College of Physicians Telehealth Working Group. Practical aspects of telehealth: doctor-patient relationship and communication. *Intern Med J* 2014 Jan;44(1):101-103. [doi: [10.1111/imj.12323](https://doi.org/10.1111/imj.12323)] [Medline: [24450527](https://pubmed.ncbi.nlm.nih.gov/24450527/)]
14. Laursen J, Danielsen A, Rosenberg J. Effects of environmental design on patient outcome: a systematic review. *HERD* 2014;7(4):108-119. [Medline: [25303431](https://pubmed.ncbi.nlm.nih.gov/25303431/)]
15. LeRouge C, Garfield MJ, Collins RW. Telemedicine: technology mediated service relationship, encounter, or something else? *Int J Med Inform* 2012 Sep;81(9):622-636. [doi: [10.1016/j.ijmedinf.2012.04.001](https://doi.org/10.1016/j.ijmedinf.2012.04.001)] [Medline: [22579395](https://pubmed.ncbi.nlm.nih.gov/22579395/)]

16. Trost J, Levin I. Att första vardagen: Med ett symboliskt interaktionistiskt perspektiv: Lund:Studentlitteratur; 2010.
17. Pigg SL. On sitting and doing: Ethnography as action in global health. *Social Science & Medicine* 2013;99:127-134. [doi: [10.1016/j.socscimed.2013.07.018](https://doi.org/10.1016/j.socscimed.2013.07.018)] [Medline: [23953135](#)]
18. Suchman L. Plans and Situated Actions: The Problem of Human-machine Communication. New York, NY: Cambridge University Press; 1987.
19. Willows T, Dizdar N, Nyholm D, Widner H, Grenholm P, Schmialek U, et al. Initiation of Levodopa-Carbidopa Intestinal Gel Infusion Using Telemedicine (Video Communication System) Facilitates Efficient and Well-Accepted Home Titration in Patients with Advanced Parkinson's Disease. *J Parkinsons Dis* 2017;7(4):719-728 [FREE Full text] [doi: [10.3233/JPD-161048](https://doi.org/10.3233/JPD-161048)] [Medline: [28984615](#)]
20. Ritchie J, Spencer L, O'Connor W. Carrying out Qualitative Analysis. In: Ritchie J, Lewis J, editors. *Qualitative Research Practice. A Guide for Social Science Students/Researchers*, ed. J.L. London • Thousand Oaks • New Delhi: SAGE Publications; 2007:219-262.
21. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? *Int J Qual Stud Health Well-being* 2014;9:26152 [FREE Full text] [Medline: [25326092](#)]

Abbreviations

ICT: information and communication technology

Edited by G Eysenbach; submitted 17.01.18; peer-reviewed by J Moll, B Kane; comments to author 22.02.18; revised version received 12.03.18; accepted 08.05.18; published 27.06.18.

Please cite as:

Sturesson L, Groth K

Effects of the Digital Transformation: Qualitative Study on the Disturbances and Limitations of Using Video Visits in Outpatient Care
J Med Internet Res 2018;20(6):e221

URL: <http://www.jmir.org/2018/6/e221/>

doi: [10.2196/jmir.9866](https://doi.org/10.2196/jmir.9866)

PMID: [29950290](https://pubmed.ncbi.nlm.nih.gov/29950290/)

©Linda Sturesson, Kristina Groth. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 27.06.2018. 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.

Review

Applications of Space Technologies to Global Health: Scoping Review

Damien Dietrich^{1*}, MD; Ralitza Dekova^{1*}, MSc; Stephan Davy¹, BS, Med; Guillaume Fahrni¹, MD; Antoine Geissbühler¹, MD, PhD

Hopitaux Universitaires de Genève, eHealth and Telemedicine Division, Geneva, Switzerland

* these authors contributed equally

Corresponding Author:

Damien Dietrich, MD

Hopitaux Universitaires de Genève

eHealth and Telemedicine Division

Rue Gabrielle-Perret-Gentil 4

Geneva, 1205

Switzerland

Phone: 41 22 372 62 01

Email: Damien.Dietrich@hcuge.ch

Abstract

Background: Space technology has an impact on many domains of activity on earth, including in the field of global health. With the recent adoption of the United Nations' Sustainable Development Goals that highlight the need for strengthening partnerships in different domains, it is useful to better characterize the relationship between space technology and global health.

Objective: The aim of this study was to identify the applications of space technologies to global health, the key stakeholders in the field, as well as gaps and challenges.

Methods: We used a scoping review methodology, including a literature review and the involvement of stakeholders, via a brief self-administered, open-response questionnaire. A distinct search on several search engines was conducted for each of the four key technological domains that were previously identified by the UN Office for Outer Space Affairs' Expert Group on Space and Global Health (Domain A: remote sensing; Domain B: global navigation satellite systems; Domain C: satellite communication; and Domain D: human space flight). Themes in which space technologies are of benefit to global health were extracted. Key stakeholders, as well as gaps, challenges, and perspectives were identified.

Results: A total of 222 sources were included for Domain A, 82 sources for Domain B, 144 sources for Domain C, and 31 sources for Domain D. A total of 3 questionnaires out of 16 sent were answered. Global navigation satellite systems and geographic information systems are used for the study and forecasting of communicable and noncommunicable diseases; satellite communication and global navigation satellite systems for disaster response; satellite communication for telemedicine and tele-education; and global navigation satellite systems for autonomy improvement, access to health care, as well as for safe and efficient transportation. Various health research and technologies developed for inhabited space flights have been adapted for terrestrial use.

Conclusions: Although numerous examples of space technology applications to global health exist, improved awareness, training, and collaboration of the research community is needed.

(*J Med Internet Res* 2018;20(6):e230) doi:[10.2196/jmir.9458](https://doi.org/10.2196/jmir.9458)

KEYWORDS

satellite imagery; satellite communications; public health; remote sensing technology; global positioning system; geographic information systems; telemedicine; spaceflight; space medicine; global health

Introduction

Background

The space-earth frontier is no longer afforded to a narrow niche of individuals. Compared with over 50 years ago when the first humans reached outer space, and satellite function only concerned a small number of scientists, today many programs and research projects in multiple fields exist that make use of outer space technologies. The field of global health too—interdisciplinary by definition—has innovated over the years and has made strides in the advancement of health aims using space technologies. Examples include using remote sensing technology to detect environmental changes that have a significant effect on local population health, satellite communication for medical endeavors and management of natural disasters, advancing medical knowledge through space medicine programs, and tapping into the benefits of localization through global navigation satellite systems (GNSSs). The UNISPACE+50 conference, taking place in 2018, marks the 50th anniversary of the start of the United Nations (UN) conferences that engaged states to cooperate in their outer space engagements. After half a century of cooperation and innovation, it is an appropriate time to take stock of where the global health field has ventured into its use of space technologies.

Objectives

The Expert Group on Space and Global Health of the UN Office for Outer Space Affairs (UNOOSA), in its 2016 work plan, mandated one of its members, Antoine Geissbühler, to produce a compilation of practices and initiatives [1] in the form of a scoping review, including both a literature review and stakeholders' involvement, via a self-administered questionnaire to identify (1) The main stakeholders in the field, (2) The key applications of space technologies to global health, and (3) The gaps, challenges, and perspectives.

This work uses a scoping review methodology, including both a literature review and stakeholders' involvement via a self-administered questionnaire. These are used to identify (1) The main stakeholders in the field, (2) The key applications of space technologies to global health, and (3) The gaps, challenges, and perspectives.

Key stakeholders of the fields are first briefly presented. Then, main themes in which space technologies are of benefit to global health are identified and illustrated in four technological domains. Finally, findings are summarized, and gaps, challenges, and perspectives are discussed.

Methods

Scoping Review

The general aim of a scoping review is to “map rapidly the key concepts underpinning a research area and the main sources and types of evidence available and can be undertaken as [a] stand-alone project in [its] own right, especially where an area is complex or has not been reviewed comprehensively before” [2]. As opposed to systematic reviews, scoping reviews can include a diversity of sources and, in particular, are not

necessarily limited to scientific articles. This allows researchers to gain a better overview on a broad subject but prevents precisely answering a well-defined question.

Accordingly, the scoping review methodology matches our objectives and was chosen for this work [3,4]. The Expert Group on Space and Global Health identified four key technological domains that are applied or could be applied to global health [5,1]: *domain A*: remote sensing, *domain B*: GNSS, *domain C*: satellite communications, and *domain D*: human space flight. Our scoping work was conducted using these domains as a framework. A distinct literature search was conducted for each of the four key technological domains on PubMed, with eventual further insights gathered from RERO, the Western Switzerland online network for libraries, and Google Scholar. Additionally, stakeholders' insights were collected through an emailed, self-administered questionnaire.

Literature Review

Search Strategy

Searches were conducted per technological domain. PubMed was the main search engine used. Complementary searches were performed on RERO and Google Scholar. Resources retrieved by these search engines were included only if they brought insights that were not identified in the original search. The keywords used for each domain are listed in Table 1. For each domain, the basic search structure was “domain-associated technology” AND “health.” Medical Subject Headings (MeSH) terms were not systematically used as some did not properly refer to the technology we were searching for. For Domain C, the search term “eHealth” was used in addition to “health” as it is a MeSH entry term for “telemedicine.” The year-parameter of the search was unbounded to access published material that could date back to the start of outer space technology and its application to global health activities. The “Similar Articles” feature of PubMed, as well as the list of references of included articles were used to identify additional resources. Finally, key stakeholders' websites were assessed for ongoing projects (listed in Multimedia Appendix 1).

Material Inclusion

Presentations, books, websites, and articles identified by the searches were included if they satisfied all of the following criteria:

1. Reporting research, or an applied program related to health
2. Use of space technology based on one of the four domains (remote sensing, GNSS, satellite communication, and inhabited space flight)
3. Only for RERO and Google Scholar: global health application not already described in a resource identified through the PubMed search

In each domain, duplicates found across the various search engines were excluded. Included resources were entered in Endnote (Clarivate Analytics) by domain and exported on spreadsheets (one for each domain). Importantly, resources written in languages other than English but whose abstracts were translated to English were included in the review. However, for these, full texts were not read.

Analysis and Reporting

On the spreadsheets described above, global health applications were identified for each resource. Then, main themes of global health applications per technological domain were identified. The numbers of resources per theme were counted in an attempt

to weight the different themes (Tables 2-5) for a particular technological domain. Articles dealing with more than one theme were allocated according to the dominant theme. If this was not possible, they were classified as “miscellaneous.” The different themes were then described by domain in the main text and illustrated by the citation of relevant articles.

Table 1. Search keywords used in this study.

Domain	Keywords used
Remote sensing	<ul style="list-style-type: none"> Remote sensing Health
Global navigation satellite systems	<ul style="list-style-type: none"> Satellite Global Positioning System (GPS) Global Navigation Satellite Systems (GNSS) Geographic Information Systems (GIS) Health
Satellite communication	<ul style="list-style-type: none"> Satellite communication Satellite Telemedicine Global Health Health eHealth
Inhabited space flight	<ul style="list-style-type: none"> Human spaceflight Manned spaceflight

Stakeholder Involvement

To gather additional insights, a brief semistructured, self-administered questionnaire (Multimedia Appendix 2) was created and sent by email to 16 stakeholders identified after the initial literature searches. Two reminders were eventually sent to nonresponders. The questionnaire was created following the same structure and logic as our overall work. Four open-ended questions were used, asking participants about:

1. Key applications of space technologies to global health for each domain and eventual other domains
2. Gaps, challenges, and opportunities
3. Key events related to the topic
4. Other important remarks they may have

Comments on the current state of the space-technology-global-health interface are included at the end of the Results section, whereas gaps and potential solutions are presented in the Discussion section.

Results

Data Collected

After the whole literature review process, 222 articles were included for domain A, 82 articles for domain B, 144 articles for domain C, and 31 articles for domain D. In total, 473 articles were included (6 of those were included in 2 domains). Most of the included resources were peer-reviewed scientific articles (96%, 213/222 for domain A; 99%, 81/82 for domain B; 84%, 121/144 for domain C; and 100%, 31/31 for domain D), and other types of sources included mainly book sections and Web pages. The mean publication year and the minimal and maximal

publication years were 2010 (1985; 2017) for domain A, 2010 (1996; 2016) for domain B, 2004 (1986; 2016) for domain C, and 1999 (1981; 2011) for domain D. Of note, in accordance with the scoping methodology used for this work and described in the Methods section, we used different combinations of keywords; included resources via the “Similar Articles” feature of PubMed and the list of references of included articles and navigation on stakeholders’ websites.

Regarding questionnaires, 3 out of 16 sent were answered and included for the analysis.

Presentation of the Stakeholders

Using insights from the literature review and the questionnaires, we performed a nonexhaustive listing of stakeholders implicated in the space and global health fields.

We categorized stakeholders per their nature: National Space Institutes; UN entities and specialized agencies; entities fostering data availability, usage, analysis and exchange; and journals, other consortia, and associations. These stakeholders are depicted in Figure 1.

National Space Institutes are usually public institutes that are responsible for applying their countries’ spatial programs. Their missions are space exploration, education, research, and development that can sometimes be translated into commercial applications, or eventually for terrestrial use. Nonexhaustively, we identified the US’s National Aeronautics and Space Administration (NASA), the Russian Federal Space Agency, the Japan Aerospace Exploration Agency, the French Centre National d’Etudes Spatiales, and the Canadian Aeronautics and Space Institute as being engaged at the space and global health interface.

The UN comprises several entities that deal with space and global health. The UN platform for Space-based Information for Disaster Management and Emergency Response (UN-SPIDER) and the UN Operational Satellite Applications Program (UNOSAT) aims at providing all countries and international organizations with space-based information useful for disaster risk management and emergency response. This is also one of the goals of the UN Economic and Social Commission for Asia and the Pacific. UN-Space is an interagency coordinating body aiming at fostering collaboration and synchronization between the various agencies implicated in space and global health. The Committee on the Peaceful Use of Outer Space (COPUOS) was set up by the general assembly in 1959 to govern the exploration and use of space for the benefit of all humanity: for peace, security, and development. The Expert Group on Space and Global Health that guided this review is part of COPUOS and has a focused scope on global health applications of space technologies. Of note, UNOOSA is a governing office that comprises UN-SPIDER, UN-Space, and COPUOS. It is also in charge of organizing the UNISPACE+50 conference that will mark the 50th anniversary of the first UN conference on the peaceful uses of outer space that engaged states to cooperate in their outer space uses. Applications of space technologies to global health is also an important interest of the World Health Organization (WHO), a specialized UN agency.

In addition, we identified entities aiming at fostering satellite data availability, analysis, visualization, interoperability, and exchange. As an example, the Group on Earth Observations (GEO) is a partnership of governments and organizations whose one activity among others is to build the Global Earth Information System of Systems. This platform offers access via a Web-based interface to earth-observation data coming from multiple sources, including satellites. It acts as a connector between different data sources and thus, increases data availability for researchers, public health professionals, and international organizations. The Global Disaster Alert and Coordinating System is a cooperative framework under the UN umbrella that connects to various services and platforms (the majority of which are listed in this section) to create a comprehensive solution that aims to create early alerts in the case of a disaster, to assess the impact of the disaster, to coordinate the response, and to provide disaster maps and satellite images. Black Sky is a service of Spaceflights Industries (a private company) that provides access to satellite imagery in addition to other sources of data (eg, radio communication and social media). It also offers spatial analysis based on those datasets and several algorithms. Humanitarian Data Exchange is an open platform for data sharing in the humanitarian context. The OSGeo foundation is a foundation that supports the creation and usage of an open source geospatial software. Finally, the National Oceanic and Atmospheric Administration provides environmental data, some of which are acquired via satellites. It is to be noted that most of the national space institutes listed previously are data providers too.

Some stakeholders are consortiums or associations active in the field of space and global health. We included the University Corporation for Atmospheric Research that regroups North

American colleges and universities focused on research and training in the atmospheric and related Earth system sciences. The Space Generation Advisory Council is a nongovernmental organization that promotes the access of students and young professionals to UN agencies and National Space Institutes.

Finally, we included as part of [Figure 1](#) a nonexhaustive list of journals that are implicated in the field of space and global health.

Domain A: Remote Sensing

Definition

Remote sensing refers to data collection at distance, usually from a satellite or an aircraft, as opposed to on-site sensing.

How It Works

A sensor, carried by a satellite or an aircraft, detects electromagnetic radiation coming from Earth and its characteristics. The electromagnetic radiation may be the reflection of an external source of energy (usually the sun) or of a source of energy carried by the satellite or aircraft itself. The terms passive or active remote sensing are used, respectively [6].

The detected signal is then processed through algorithms of various complexities to derive the parameters of interest. Example of parameters that can be derived via remote sensing include land temperature, altitude, humidity, rainfall, cloud coverage, air pollutants, livestock density, vegetation indices, sea temperature, sea salinity, sea nutrient concentration, sea algae concentration, sea bacteria concentration, urbanization, population density, and bare soil coverage. This list is nonexhaustive.

Insights From the Literature Review

Overall, remote sensing was useful for global health in three major ways:

- Identifying associations between diseases (or disease vectors) and remotely sensed parameters
- On the basis of these associations, model development and forecasting of the spatio-temporal evolution of diseases, thus allowing rational public health strategies
- Direct monitoring of certain microorganisms

Two major themes and two secondary themes were identified and are presented in [Table 2](#).

Main themes of global health applications in the remote sensing domain were identified, and the total number of resources per theme were counted as described in the Methods section.

Remote sensing was most used to identify determinants of infectious diseases and to develop models to predict their evolution (Theme A-1). For example, Midekisa et al [99] quantified the degree of association between malaria cases and remotely sensed environmental parameters such as rainfall, vegetation indices, and temperature. On this basis, they developed and tested a model able to predict malaria evolution and thus, guide public health decisions. Applications of spatial technologies for malaria transmission modeling and control were reviewed in 2015 by Gebreslasie [48]. In addition to

malaria [8-10,17,24,26,31,33-35,38-42,46,48,49,55,56,76,87, 90-93,99,100,102,103,105,111-115,118,119,122, 129-132,141] and schistosomiasis [15,37,53,54,63,95,124,126, 127,142-145,153,154,156,158], dengue fever [7,12,13,16,23, 44,89,94,98,101,117,140], cholera [43,71,72,74,75,80,88], and cyanobacterias [28,81,82,123,137,138,148,150,155] were the most studied.

Figure 1. Nonexhaustive collection of stakeholders and journals in the intersection of space technology and global health.

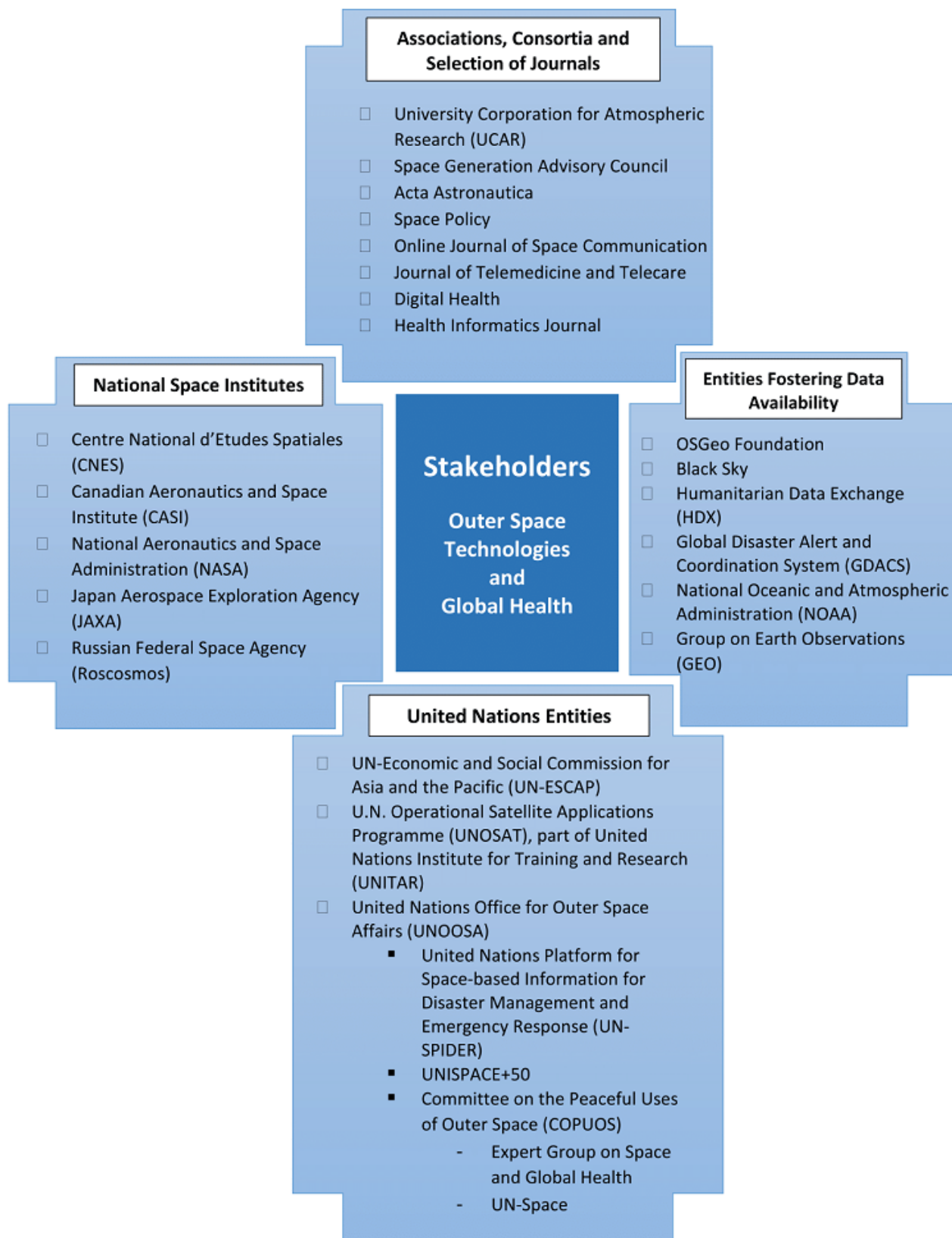


Table 2. Main themes of global health application for the remote sensing domain.

Theme category	Theme	Articles identified
Main themes-A		
A-1	Infectious diseases	153 [7-159]
A-2	Air pollutants and noncommunicable diseases (NCDs)	37 [160-196]
Secondary themes-B		
B-1	Other environmental pollutants and NCDs	8 [197-204]
B-2	Other environmental parameters and NCDs	12 [205-216]
B-3	Miscellaneous	12 [217-224,6,225-227]

Other studied diseases or pathogens included meningitis [225]; brucellosis [70]; *C. imicola* [67]; avian pathogens [25,134,136,50]; *V. vulnificus* [52]; *V. parahaemolyticus* [52]; *Fasciola hepatica* [36]; hand, foot, and mouth disease [20]; Helminth infections (not limited to schistosomiasis) [120,85,21,22]; Lyme disease [108,45,79,110]; Guinea worm [30]; Nipah virus [133]; onchocerciasis [68]; opisthorchiasis [146]; rotavirus [69]; typhoid fever [32]; Rift Valley fever [139,125,84]; Murray Valley encephalitis virus [121]; West Nile fever virus [96,159]; and hanta virus [149,152].

In an important number of studies, disease vectors (and not disease cases) were the outcomes predicted based on sensed environmental parameters. These vectors included *Anopheles* [10,55,103,141] (transmitting malaria) and *Aedes* [16,44,101,117] (transmitting dengue) mosquitos, as well as ticks [45,64,77,79,110,147] (transmitting Lyme's disease among other tick-borne diseases).

Of note, remote sensing techniques allow to directly derive the concentrations of some bacteria. Cyanobacteria produce various toxins that have been linked to the occurrence of amyotrophic lateral sclerosis and nonalcoholic liver disease [155,138]. They also have distinct fluorescent properties that can be exploited in active remote sensing to monitor their concentration [28,81,82,123,137,138,148,150,155].

The second main theme (A-2) was the use of remote sensing to monitor air pollutants and eventually link them to noncommunicable diseases (NCDs) such as respiratory diseases (asthma [161,188,191] and others [185,162,208]), coronary artery disease [165], premature birth [195], and low birth weight [181]. Particulate matter (PM_{2.5} and PM₁₀) [188,181,161,195,196,175,194,193,164,189,174,170,165,163,173,171,172,160,178,183], O₃ [218], NO₂ [180,225,218], pollens [209], asbestos [202], volcanic ash [184], and wildfire smoke [168,190,176] are among the air pollutants that can be effectively detected by remote sensing. Temperature and humidity are usually included as additional parameters when monitoring air pollutants as they may affect both respiratory diseases and air pollutants behavior. If many articles successfully describe the use of remote sensing for the monitoring of air quality, only a few establish a direct

link between air pollutants and health outcomes [188,162,181,191,161,195,165,185,208]. Moreover, results may be controversial, such as for asthma, where one study found a correlation between childhood asthma hospital admission that disappears after multivariate analysis [161], another one finds no correlation between air pollution and asthma prevalence [188], and a last one finds a correlation between PM_{2.5} concentration and salbutamol (treatment used in asthma and chronic obstructive pulmonary disease) use [191].

The remaining articles identified for remote sensing dealt with monitoring environmental pollutants (B-1) or parameters (B-2) and their links with NCDs. For example, studies investigated the link between urban greenness and birth outcomes [212] or cardiovascular diseases [206]. Others investigated the link between drought and respiratory illnesses [208] or between heat and elderly health [215,211] or childhood diarrhea [205]. Additional parameters or pollutants that can be sensed by remote sensing include artificial lights [207], soil contaminants (heavy metals [203], nitrates, nitrogens [197]), water quality [128,97,14], and arsenic [198].

Domain B: Global Navigation Satellite System

Definition

GNSS is the generic term for satellite navigation systems that provide autonomous geo-spatial positioning with global coverage [228]. GNSSs are satellite ensembles that allow any user on or near the Earth to determine their position with a precision from some meters to some centimeters.

The term global positioning system (GPS) is specific to the US' GNSS, the NAVSTAR GPS. The Global Orbiting Navigation Satellite System (GLONASS) is the Russian Federation's GNSS. As of 2013, these two are the only fully operational GNSSs.

Other GNSSs in various stages of development and deployment include:

- Galileo, the European Union's positioning system
- IRNSS, India's next generation regional system
- QZSS, the Japanese regional system
- China's BeiDou (COMPASS) GNSS

Table 3. Main themes in the global navigation satellite systems (GNSS) domain.

Theme category	Theme	Articles identified
Main themes-C		
C-1	Noncommunicable diseases	22 [230-251]
C-2	Communicable diseases	18 [252,253,25,254,255,38,48,256-258,95,259-265]
C-3	Innovative methods for research	26 [266-291]
Secondary themes-D		
D-1	Autonomy improvement	6 [292-297]
D-2	Transportation	5 [298-302]
D-3	Health care access	4 [303-306]
D-4	Accurate timekeeping	1 [307]

How It Works

Each system (GPS, GLONASS, Galileo, COMPASS, etc) consists of a constellation of satellites that send a continuous signal toward the Earth. Individuals wanting to use GNSS to determine their position must have an antenna that receives the signals coming from the satellites and a receiver that translates these signals. The antenna position will be deduced from the measurements of the time delay between the emission time (satellite) and the reception time (receiver) for at least four signals coming from different satellites [229]. Most importantly, the atomic clocks onboard the satellites are all synchronized so that the signals coming from the different satellites of the same constellation share the same reference time scale.

Although a GNSS is the space technology that is highlighted in this review, often mentioned in global health applications is the use of a geographic information system (GIS). A GNSS allows a user to determine the location of an object or individual, whereas a GIS is the system for storing, combining, and displaying data (partly coming from GNSS) on a map. It allows users to easily visualize spatial data, analyze them, and interpret trends or patterns.

Insights From the Literature Review

Seven themes were identified after the literature review and are shown in Table 3.

Main themes of global health applications in the GNSS domain were identified, and the total number of resources per theme were counted as described in the Methods section.

GNSS was used in epidemiological studies, often in combination with GIS and remote sensing. NCDs were the focus of many studies, whether directly as a measured outcome, or because of their risk factors being studied [239,236,232,245]. Physical activity (PA) was a very popular research area [243,237,247,233,250], most notably in children and adolescents [230,242,240,248,234,244,231]. Edwards and authors [234] assessed adolescents' use of public parks with regards to the features of the parks. The parks were characterized using GIS and a desktop auditing tool that uses remote sensing techniques, whereas the adolescents were surveyed to assess their activities. In two other studies in the United States and Switzerland [244,231], participants wore GPS receivers and accelerometers,

enabling researchers to assess and compare the intensity and location of the PAs. Links between different locations (home, playground, sidewalk, and more) and the intensity of PA were identified. In addition to PA, the built and natural environment were studied for their associations with NCDs. Researched environmental determinants of health ranged from air pollution [291,251] and water quality monitoring [246,241] to the complex ways in which climate change impacts global health [259]. For this purpose, researchers used GNSS and satellite imagery in a variety of ways. Interestingly, happiness was also studied as a health outcome. MacKerron and Mourato (2013) [238] used GPS to locate individuals at various, spontaneous moments while they answered questions about their subjective well-being. They found that participants were substantially happier in natural rather than urban environments. The variety of ways in which GIS can be used in environmental epidemiological studies was reviewed by Nuckols et al (2004) [282], who concluded that GIS and GPS are useful tools in providing precise locations of subjects and studying proximity and level of exposure to environmental contaminants.

GNSSs have been used often in the field of communicable diseases too, including person-to-person transmissible varieties [253,255], vector-borne diseases [21,48,38,252,258,260,257,264,256,95,263], and zoonoses [262,260,254,261,25]. In our search, the most studied communicable disease was malaria. Predicting vector breeding sites [21,38,257], malaria incidence, and adherence to medication [263] using GNSS, often in combination with GIS and remote sensing, were some practical applications. Additionally, distance to health facility was also used for malaria risk mapping [265]. Studies of zoonotic communicable diseases were limited to avian pathogens in this domain. Newman et al [260] marked two hosts of H5N1, a highly pathogenic avian influenza, with GPS transmitters and found links between flu outbreaks in humans and the hosts' travel patterns.

The use of GNSS as a new tool for epidemiological research was discussed in a variety of articles [282,275,270,274,280,279,281,287,288,271,269,284,268]. GNSS use was reported to construct random sampling frames for surveys, mapping households, or determining population estimates [267,278,272,277,273,266,285,289,276]. The potential future impacts of

GPS devices on medicine is discussed in Pager's article, *Impacts for medicine of global monitoring* [283].

Geolocation of individuals has been used in the assistance of mentally or physically impaired individuals to improve their autonomy [294,297]. Alisky [293] presents hypothetical scenarios whereby GPS devices can be of assistance. For instance, in the case of an individual with partial complex seizure disorder, the individual can wear a GPS-enabled watch that will notify a health management center in the case of a seizure. Gally et al [292] give a review of GPS technologies that have already been available to aid visually impaired individuals to navigate their surroundings. They discuss several limitations, eg, that the GPS receiver does not work well unless satellite coverage is satisfactory, and this is affected by climatic conditions as well as the user's location. GPS can also be of assistance for persons suffering from dementia. This could be achieved through orientation and safety cues, daily reminders of activities, protection against wandering, and direct links to medical assistance in case of incapacitation. Potential benefits are decreased stress and workload for formal and informal caregivers, decreased institutionalization, and thus, lower costs.

Geolocation is also helpful in promoting health care access in different settings [306,303]. In Bolivia, Perry et al [305] used GPS techniques and satellite imagery of the remote, impoverished, and mountainous region of Andean Bolivia to create a GIS that enabled them to assess the physical accessibility of several populations to health care services and auxiliary nurses [305]. Their findings demonstrate how medical geography can be used for better informed health care policy and planning decisions. Tassetto et al [286] tested a novel method to locate victims of disaster by using their existing portable devices (such as simple mobile phones or laptops) and the existing cellular network. Their proposed technology is mediated by a satellite system and requires little action by victims. Although this new system has been tested in experimental settings, it is yet to be used in real-life scenarios. In northern Nigeria, polio vaccination teams were tracked with handheld GPS devices, and their movements were overlaid on catchment area maps [304]. This method allowed the identification of low vaccine coverage areas and was identified as a tool to improve microplanning of global health projects.

The use of GNSS to improve transportation for improved public health appears as one area in which there is a huge potential for growth, for instance by preventing road accidents. Guo et al [301], working under the current constraints of suboptimal space-time reference for vehicles, conducted research with the aim of locating vehicles with high precision, down to the lane in which the vehicle is moving. This has immense safety implications which, in addition to a safety notification system, can provide information on high-risk vehicles (eg, trucks carrying chemicals) or high-priority-of-way vehicles (eg, school busses) and can also track illegal or dangerous vehicle movements [301]. Other transportation-related GPS studies have focused on speeding [300], commute routes, and daily mobility [299,298], as well as emergency patient transportation [302].

Finally, in a category of its own, accurate timekeeping using GPS was a proposal brought forth by Aljewari et al, especially in settings where time is of utmost importance, such as in hospitals [307].

Domain C: Satellite Communication

Definition

Satellite communication is the ability of information to travel from one area to another via a communication satellite that is in orbit around the Earth. It is often performed with mobile satellite phones and is distinct from cellular phones that use earth-based towers that form a cellular network. "Wide area coverage, reliable data delivery, and robustness and broadcast or multicast are the unique features of satellite systems" [308].

How It Works

Satellite communication has two main components: the ground segment, which consists of fixed or mobile transmission, reception, and ancillary equipment, and the space segment, which primarily is the satellite itself. A typical satellite link involves the transmission (uplinking) of a signal from an Earth station to a satellite. The satellite then receives and amplifies the signal and retransmits it back to Earth (downlinking). Satellite receivers on the ground include direct-to-home satellite equipment, mobile reception equipment in aircraft, satellite telephones, and handheld devices [309].

Insights From the Literature Review

This domain was largely centered on telemedicine, often combined with tele-education. Health-on-the-go is defined below with several examples from the literature, and there are a handful of demonstrations of how satellite communication can be of importance in disaster situations. Main themes are presented in Table 4.

Main themes of global health applications in the satellite communication domain were identified, and the total number of resources per theme were counted as described in the Methods section.

Telemedicine is the application of communication technologies to the field of health in instances where medical expertise or resources are not available on site for different reasons. These reasons, nonexhaustively, include the geographical distance; physical barriers (mountains, space, desert, etc) and insufficient time or resources to transfer a patient. Often, the patient may be in the physical presence of a health care provider (HCP), but telemedicine could mean connecting the two parties to a third party at a distance, such as a medical specialist or a general practitioner (GP) if the HCP is a nonphysician. Telemedicine is possible via satellite and cellular network. This review is limited to telemedicine by means of satellite communications. More in-depth assessment of the definition and breadth of telemedicine can be found in several review and discussion references [390,386,337,393,441,383,352,353], some theoretical articles linking satellite communication with health [441,415,446,448,372,444,451,453,449,445,442,454,447,364,450], as well as country reports [330,387,322,378,343,355,359,325,321,399].

Table 4. Main themes in the satellite communication domain.

Theme category	Theme	Articles identified
Main themes-E		
E-1	Telemedicine	90 [310-399]
E-2	Tele-education	14 [400-413]
E-3	Health-on-the-go	14 [414-427]
E-4	Disaster prevention, early warning, and management	13 [428-440]
Secondary themes-F		
F-1	Miscellaneous	13 [441-453]

A first example is in Thailand, where the country's first communication satellite, THAICOM, was launched in 1993. HCPs in rural areas were connected with specialists in urban areas, and consultations became possible, with two main components: videoconferencing and exchange of medical images. Thailand's telemedicine network is housed in its Ministry of Public Health, with all hospitals that are in the telemedicine network also having a direct communication link with the government base. The Thai example illustrates a common model of telemedicine and teleconsultation: access to expert opinion by GPs, nurses, or paramedics via videoconferencing or textual exchange [373,368,315,318,316,397,351,314]. These are often accompanied by still images from radiography [332,361,311] or dermoscopy [356], but innovative advances have made possible the transfer of 3D images [381] and live ultrasound feed [331,374,317,326]. Use of telemedicine methods has been reported in various medical fields including dermatology [345], pediatrics [327], and surgery. Telesurgery [376] has been trialed on internal mammary artery dissection in pigs with robotic technology to determine feasibility and bandwidth requirements. The authors concluded that telesurgery via satellite communication is feasible and also identified the limit of satellite bandwidth below which it cannot be performed (3 Mb/s).

Telemedicine using satellite communication may also be useful for a country's defense system. By equipping more than 300 US Navy ships with telemedicine capabilities, researchers estimated that 17% of medical evacuations could be avoided, representing US \$4400 savings per single medical evacuation [384]. Similarly, German defense units have access to a telemedicine workstation, accompanied by a medical officer present on-board the ship or at the unit [371]. This station has the possibility of being equipped with various medical devices (X-ray film digitizer, dermatoscope, otoscope) and can also contain other imaging methods (eg, videocamera and ultrasound). The authors propose cooperation not only between civilian and military health service providers but also military-military cooperation between the medical services of allied armed forces.

As the field of telemedicine is both broadly defined and applied, as well as having fluid borders with tele-education and health-on-the-go, further sources were found in this search that do not fall under the broader categories discussed above [310,312,313,319,320,323,324,328,329,333-335,338,339,341,342,344,

347-350,354,357,358,360,362,363,365,369,370,375,377,379,380,382,385,388,389,412,391,392,394,452,395,398].

Medical tele-education, the practice of providing new or continuing medical education via distance learning, often uses the same networks and infrastructures as telemedicine does. It is especially useful for HCPs who are located far from teaching facilities [321,400,401,403-411]. The Réseau en Afrique Francophone pour la Télémédecine network is one such example of successful implementation of tele-education; a model that has expanded into multiple countries and continents [402]. Health educators, usually located in teaching universities of larger cities of the region, teach courses to HCPs in peripheral areas in real time. Two-way communication enables students to ask questions and collaborate with the lecturer. Exchanges in the same country or between neighboring countries are promoted as much as possible to build capacity and collaboration. Another application of tele-education is implemented in Japan, where 39 universities and institutes were connected by satellite for a joint radiology conference [413]. Participants engaged in discussions around various images, and the results of a survey to radiologists after the conference showed that while the technology used may not be good for diagnosing purposes, it is useful for discussion and educational purposes.

The third broadly studied area of satellite communication and global health is what we refer to as health-on-the-go. In this theme, which can be considered as subcategory of telemedicine, mobile medical units can provide treatment and can transmit health information (text, health parameters, images, laboratory exams) using satellite communication [426,427,421,419,422,414,418]. This gives the ability to provide health care services to individuals over a large area that may be deprived of traditional communication systems. The TraumaStation is one such device, a portable and lightweight suitcase that carries ultrasound, electrocardiogram, blood pressure, and oxygen meter apparatus [425]. The TraumaStation allows for telecommunication with instant messaging and real-time video stream through satellite and a variety of other gateways. Alternatively, the HOPEmobile provides biometric measurement (body mass index, cholesterol, glycosylated hemoglobin, and retinal screening) from a mobile unit [416]. The study reported a return on investment of US \$15 for every US \$1 spent and a significant reduction in overall cholesterol at the second screening of a patient. Finally, Guo and colleagues (2015) [417] describe a portable, robust, and low-power device

that performs all essential functions of enzyme-linked immunosorbent assay and can thus diagnose diseases in remote, mobile contexts. The results can then be sent via cell phone short message service (SMS) messaging or in email format via satellite. The authors describe how patient confidentiality is taken into account through the usage of this device. Another area of health-on-the-go is emergency patient transportation. The transmission of the patient's medical history, vital signs, and laboratory exams (for instance electrocardiogram) during the transport can allow a remotely based medical expert to guide the management of the patient. Nakajima et al [424] explain that 3G mobile networks tend to be sensitive to congestion in urban areas and that the satellite provides a good solution to counter this. One technical innovation in this area includes the Emergency Medical Video Multiplexing Transport System. This divides a patient's live video stream from a medical vehicle into four pieces, and these translate to high-quality videos that can be viewed by emergency doctors in a remote location [423].

Satellite communication is also valuable in emergency situations arising from natural disasters, man-made disasters (eg, terrorism and war), highly contagious diseases, or large-scale epidemics [431,440,436,437,433,432,430,429,435,439,434]. Satellites for Epidemiology (SAFE) is a system for early health warnings in a postdisaster period. It is a system that combines satellite, radio, wireless networks, and GIS to promptly identify and respond to a disease outbreak. SAFE's added value is reported to be its integration into already-existing national, regional, and international preparedness plans [428]. Existing cellular and telephone networks almost always become overloaded or disabled following disasters, so satellite communication methods are superior in these instances. For this reason, East Carolina University tested the time it would take to set up a fault-tolerant communications infrastructure from scratch; one component of several being the satellite connection. They concluded that the time it took to mount the network by technically trained personnel made it a feasible and valuable contribution to disaster response operations. Potential drawbacks of this are that technical experts of the system may need to be made a part of the team of emergency responders and that in case of loss of electrical power, alternative methods need to be used [438].

Domain D: Human Space Flight

Definition

We looked for evidence on how inhabited space flight-associated technologies and procedures may promote global health.

Literature Review

Two main themes and one secondary theme were identified and are represented in Table 5. Main themes of global health applications in the inhabited space flight domain were identified, and the total number of resources per theme were counted as described in the Methods section.

Telemedicine seems to be one of the dominant theme at the crossings of inhabited space flights and global health. Indeed, providing health care for an astronaut needing medical assistance onboard a space station, or an individual living far from medical expertise in a desolated rural area, may pose similar problems. In both cases, one must deal with the restriction of not being

able to quickly transfer the patient and limited medical resources and expertise in the patient's vicinity [366,462,463,465,467,466,468]. Telemedicine thus provides a possible solution in both cases. Interestingly, challenges for successful implementation are similar in space and on earth. They include dealing with low bandwidth connection, maintaining stable electrical power, assuring data storage, developing intelligent software, and training users.

Going further in the similarities between space and earth telemedicine, tele-ultrasound has been extensively designed and tested in space [460,458,459,457,455,456,461,480] but is also used on earth [484]. In addition, tele-surgery has been developed and practiced on earth [376] and is foreseen to be a requirement to medical support in extraplanetary human outposts [469]. Challenges for this particular implementation notably include the latency between the command and the robot movement, induced by the long distance [469]. Taken together, telemedicine in space and telemedicine on earth are facing similar yet complementary challenges that are potential synergies for researches in the development, implementation, and testing phases.

Among included articles, technology transfer of space technologies to earth appears to be an important topic [470,429,471-479,481]. An example is the successful reprogramming of neural networks initially trained to identify craters or incoming missiles in space toward the detection of cancer-associated breast microcalcifications on mammograms [478,479,474]. The potential use, on Earth, of miniature or implantable biometric sensors developed by the NASA sensors 2000! program (S2K!) is another example of technology transfer [472].

The space scientific community is actively conducting research on how to provide adequate life support for long extraterrestrial missions or on extraplanetary outposts. In addition to new technology transfers, outputs from this research should lead to development of new medical procedures that may be applicable on earth [483].

This review focuses on how inhabited space flight-associated technologies and procedures may promote global health. It is important to note that, in addition to global health, the space research community has also been very active in various domains of life sciences. These domains include microgravity physiology, microgravity microbiology, microgravity surgery, radiation medicine, and the study of the psychological effects because of space travel and isolation.

Insights From the Questionnaires

Respondents' insights were collated and are reported below.

Domain A: Remote Sensing

Stakeholders believe that incorporating environmental exposure data into clinical practice will improve the quality of care. Indeed, diagnostic accuracy may be improved via integration of remotely sensed parameters into decision support tools. For example, knowing that the allergens concentration was high over the last days will increase the probability of asthma when a patient consults for breathlessness.

Table 5. Main themes in the inhabited space flight domain.

Theme category	Theme	Articles Identified
Main themes-G		
G-1	Telemedicine	16
G1-1	Tele-ultrasound	7 [455-461]
G1-2	General telemedicine	8 [462-466,366,467,468]
G1-3	Tele-surgery	1 [469]
G-2	Technology transfer	11 [470,429,471-479]
Secondary themes-H		
H-1	Application of medical procedures	4 [480-483]

Domain B: Global Navigation Satellite Systems

No insights were provided by respondents.

Domain C: Satellite Communications

Stakeholders provided more examples about situations in which telemedicine is used through satellite networks. These situations include people onboard a plane, on a boat, working on an off-shore platform, or on construction sites. Satellite communication may also be necessary to provide telemedicine in remote areas of developed countries. Examples include communications between French overseas territories and the mainland, or locally between islands. In all these cases, satellite communications can be used to compensate for the unavailability of the cellular network.

Domain D: Human Space Flight

In this domain, in particular, stakeholder questionnaires provided insightful additions. Physical inactivity is a major determinant of NCDs such as cardiovascular diseases and osteoporosis. It is thus of particular interest to global health. Despite this fact, studies on the physiological effects of physical inactivity are lacking. In space, astronauts are exposed to microgravity, and accordingly, the space research and development area has been very active in studying the physiological effects of microgravity, notably by using ground-based bed rest analog. As microgravity partly mimics physical inactivity and aging, space-associated study results may help us to understand the deleterious physiological effects behind those processes. Future joint research programs should thus be encouraged.

As another example of space technology transfer, bone quality measurement tools were initially developed by the space industry. A NASA review of spin-offs of space research can be found on their website.

Long-term missions will require the development of “integrated countermeasures” to prevent the adverse effects of the space environment, including radiation. These countermeasures may find applications on earth, such as in radiation medicine.

Another challenge is to be able to personalize space medicine, which is a major trend in nowadays medicine. Moreover, when thinking about long flight duration, space health systems will need to achieve some level of autonomy, which imply the development of decision algorithms and consistent procedures

that may be of benefit to global health, especially in isolated settings.

Finally, another big challenge of human space exploration is to develop a closed-loop environmental system technology to maintain, at low cost, an environment suitable for human life. These technologies include monitoring and control of physical, chemical, and biological environments; waste recycling; and food production. Results from such research may contribute to the development of sustainable and green solutions of benefit to global health.

Discussion

Principal Findings

Using a scoping review methodology, including a literature review and questionnaires to stakeholders, we identified, described, and illustrated key areas in which space technology is, or may be, of benefit to global health. Remote sensing of environmental parameters allows the prediction of communicable and NCD evolution, often in association with GIS and GNSS. GIS and GNSS are also used to bring new insights to epidemiological research, to improve access to health care, to develop autonomy assistance for the disabled, and to assist in disaster response. For this last task, space communications are also used, as well as in telemedicine and tele-education. Finally, some technologies and procedures developed by the space industry for inhabited-space flights are applied on earth. Overall, our results strengthen the vision that space technologies and global health are two synergistic fields, and they help us to identify perspectives and issues for the coming years that will be discussed in this section.

Remote Sensing

Remote sensing brings new tools for monitoring diseases, investigating their association with multiple sensed parameters, and ultimately creating an intelligent alert system. The literature is particularly abundant on infectious diseases and air pollutants. One limit is that most studies do not link directly sensed parameters to health outcomes but rather to some disease determinants (disease vectors, air pollutants). This is an interesting first step as it gives insight to more than one disease. Yet, more studies investigating direct health outcomes are needed to allow the creation of relevant models that will guide public health decisions. Importantly, the limited presence of

environmental monitoring systems in low-income countries is an obstacle. Moreover, achieving high spatial and temporal resolution either by hardware improvement or through the development of numerical models is an important challenge in remote sensing. Finally, the sustainability of the developed alert systems, as well as their reproducibility across different geographical areas, must be evaluated. In addition to adding value at the population level, remote sensing used in combination with GNSS holds great potential to assist caregivers in their routine decision making for individual patients. This could be done, eg, by assessing relevant environmental data for each patient.

Global Navigation Satellite Systems

The last example illustrates the synergy between remote sensing, GIS, and GNSS applied to global health. Indeed, most of the epidemiological studies identified in this review and aiming at predicting disease evolution based on environmental parameters are using GIS and GNSS, in addition to remote sensing. GNSS and GIS are also used in innovative epidemiological methodologies for activity tracking (eg, movement or localization as an outcome or a determinant), randomization, or population estimation. Disaster response and autonomy improvement of disabled patients constitute two other fields in which these technologies are used. Requirements for the successful use of GNSSs are stable and easily accessible signals, as well as procedures preventing power failure. Combining space-, cyber-, and ground-data thus holds a great potential. The use of big data analytics and machine learning may lead to further applications that are not even suspected nowadays. Creating a platform warranting availability, interoperability, and quality of data issued from different sources is a requirement to go further in this direction.

Satellite Communications

Satellite communications are mainly used when standard telecommunications using landlines and antenna are not available, such as in disaster situations. Through these networks, telemedicine and health tele-education are possible. Bringing medical expertise at distance is useful in various places such as in isolated rural areas, areas affected by natural disaster, but also elderly homes, isolated places in high-income countries (northern Canada, Alps), ambulances, and remote work places (off-shore platforms, boats, airplanes). In addition to information exchanges, telemedicine encompasses laboratory exams and medical procedures at distance and sometimes in real time. Examples include tele-echography, tele-electrocardiogram, tele-dermatoscopy, and tele-surgery.

Human Space Flight

In parallel, research in outer space has been very active in developing telemedicine, including tele-echography and tele-surgery. In space and on earth, challenges for the development, implementation, and testing of telemedicine are similar and complementary. Strengthening existing collaborations in the field and creating new ones thus appear particularly relevant. In addition to telemedicine, we retrieved from the literature evidence of technology and medical procedure transfers from the space industry to the health sector.

However, the number of articles retrieved was small and is probably not a true reflection of all ongoing synergies. This was confirmed by the questionnaire results that identified additional examples and themes such as the use of microgravity to study the physiology of physical inactivity, which is a major and frequent determinant of NCDs. Overall, it seems that encouraging collaborations between the space and health sectors is of particular interest for this domain (inhabited space flights). Moreover, reinforcement of the scientific publishing and public communication is needed to strengthen the scientific community awareness of the existing synergies.

Value Added From Questionnaires

In all collected questionnaires, the potential of space applications to improve global health was reported to suffer from a lack of awareness among health workers and space researchers. Moreover, a deficit in space-associated skills and knowledge was also reported for health researchers. More interdisciplinary collaboration and an easier access for health researchers to space technologies was expressed. Finally, a gap in organizational level activities was identified. Accordingly, efforts are reported to be necessary to:

- Raise awareness on the potential global health applications of space technologies
- Train researchers interested in the field
- Promote interdisciplinary collaborations
- Improve the organizational-level governance

Results from the questionnaires suggested the reinforcement of public communication and the organization of dedicated conferences and training sessions as a first step toward a more comprehensive solution. Moreover, early involvement of end users and policy makers in the various projects has been suggested to improve their relevance.

Implications of the Research

By providing a thorough review of the published literature on space and global health, as well as the identification of key stakeholders, this work presents a solid base for improving mutual understanding between the two domains. This should lead to more synergies among the various actors, including the development of formal interagency coordination mechanisms. Comprehensive strategies to address sustainable development goals must indeed leverage the complementary competencies from UN agencies such as the WHO, the UNOOSA, UNOSAT, as well as other organizations such as the GEO.

Limitations

This review has several limitations. A scoping review is a methodology useful to gather as many insights as possible on a broad subject, such as this one, to achieve a better awareness of the question and its past and ongoing research, practices, and initiatives. We chose this methodology as it matches our objectives well. The searches that we ran gave us thousands of hits but only came from two search engines: PubMed and RERO Western Swiss database. Moreover, only 3 of the 16 questionnaires sent were answered despite two reminders. Accordingly, we can't exclude that eventual supplementary themes were missed. In addition, the low response rate to the

questionnaire may have introduced biases in the insights that were reported. Insights gathered from the questionnaires should thus be considered as expert opinions. Another limitation to our study is that we cannot draw definitive conclusions on precise subthemes and questions (eg, is remote sensing effective in predicting malaria outbreaks in Africa?). For this purpose, systematic reviews are needed. The different domains that guided the searches were suggested by the Expert Group on Space and Global Health. This group is mainly constituted by key stakeholders of various national space agencies and public health authorities. Accordingly, it is unlikely that an important domain was missed, but it constitutes a limitation to our study.

The language barrier is another one. Indeed, the space literature in Russian or Japanese is abundant and not always available in an English translation, save for the abstract. Accordingly, key concepts may have been missed.

As the paper has technology at its core, one must note that articles used in the review date back to 1981. Space technology and access to it has improved significantly since then, but to remain aligned with the goal of the review, we reference all relevant articles. Nevertheless, as ease of use and access to space technology has improved in recent times, as well as an increased human presence in outer space, the themes will be largely shaped by more recent articles, simply as there are more of them.

Acknowledgments

The authors would like to thank Dr Alban Duverdier, Centre National d'Etudes Spatiales; Prof Fazlay S Faruque, The University of Mississippi Medical Center; and Mr Hajime Shinomiya, Japan Aerospace Exploration Agency for their answers to the questionnaires and their insights for the review.

In addition, they thank all the members from the Expert Group on Space and Global Health for their valuable insights and comments to this review.

Authors' Contributions

DD and RD participated in the literature review, the stakeholder involvement, and wrote the manuscript. SD participated in the literature review and reviewed the manuscript. GF participated in the literature review. AG supervised the work and reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of stakeholder websites consulted for projects and meetings.

[PDF File (Adobe PDF File), 32KB - [jmir_v20i6e230_app1.pdf](#)]

Multimedia Appendix 2

Questionnaire sent to stakeholders.

[PDF File (Adobe PDF File), 15KB - [jmir_v20i6e230_app2.pdf](#)]

References

1. Pascal M. Second meeting of the expert group on space and global health, 18-19 February 2016. 2016 Feb 18 Presented at: 53rd session of the Scientific and Technical Subcommittee (STSC, or the Subcommittee) of the Committee on the Peaceful Uses of Outer Space (COPUOS, or the Committee); 15-26 February 2016; Vienna URL: http://www.unoosa.org/res/oosadoc/data/documents/2016/aac_105c_12016crp/aac_105c_12016crp_21_0.html/AC105_C1_2016_CRP21E.pdf
2. Mays NR, Roberts E, Popay J. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, editors. Methods for Studying the Delivery and Organisation of Health Services: research methods. London: Routledge; 2001:188-220.
3. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2007;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
4. Dijkers M. KT update. Vol 4. 2015. What is a scoping review URL: <http://ktdrr.org/products/update/v4n1/> [accessed 2018-06-05] [WebCite Cache ID 6zwt54EOc]
5. Pascal M. Report on the proposed mandate, workplan and initial considerations. 2015 Presented at: First meeting of the expert group on space and global health held on 5 February; February 2015; Vienna URL: http://www.unoosa.org/pdf/limited/c1/AC105_C1_2015_CRP29E.pdf
6. Graham S.. earthobservatory.nasa.gov. 1999 Sep 17. Remote Sensing : Feature Articles URL: https://earthobservatory.nasa.gov/Features/RemoteSensing/remote_08.php [accessed 2017-10-17] [WebCite Cache ID 6uHjQe24i]

7. Adde A, Roucou P, Mangeas M, Ardillon V, Desenclos JC, Rousset D, et al. Predicting Dengue Fever Outbreaks in French Guiana Using Climate Indicators. *PLoS Negl Trop Dis* 2016 Apr;10(4):e0004681 [FREE Full text] [doi: [10.1371/journal.pntd.0004681](https://doi.org/10.1371/journal.pntd.0004681)] [Medline: [27128312](https://pubmed.ncbi.nlm.nih.gov/27128312/)]
8. Adde A, Roux E, Mangeas M, Dessay N, Nacher M, Dusfour I, et al. Dynamical Mapping of Anopheles darlingi Densities in a Residual Malaria Transmission Area of French Guiana by Using Remote Sensing and Meteorological Data. *PLoS One* 2016;11(10):e0164685 [FREE Full text] [doi: [10.1371/journal.pone.0164685](https://doi.org/10.1371/journal.pone.0164685)] [Medline: [27749938](https://pubmed.ncbi.nlm.nih.gov/27749938/)]
9. Adimi F, Soebiyanto RP, Safi N, Kiang R. Towards malaria risk prediction in Afghanistan using remote sensing. *Malar J* 2010 May 13;9:125 [FREE Full text] [doi: [10.1186/1475-2875-9-125](https://doi.org/10.1186/1475-2875-9-125)] [Medline: [20465824](https://pubmed.ncbi.nlm.nih.gov/20465824/)]
10. Ageep TB, Cox J, Hassan MM, Knols BGJ, Benedict MQ, Malcolm CA, et al. Spatial and temporal distribution of the malaria mosquito Anopheles arabiensis in northern Sudan: influence of environmental factors and implications for vector control. *Malar J* 2009 Jun 07;8:123 [FREE Full text] [doi: [10.1186/1475-2875-8-123](https://doi.org/10.1186/1475-2875-8-123)] [Medline: [19500425](https://pubmed.ncbi.nlm.nih.gov/19500425/)]
11. Almeida AS, Werneck GL. Prediction of high-risk areas for visceral leishmaniasis using socioeconomic indicators and remote sensing data. *Int J Health Geogr* 2014 May 20;13:13 [FREE Full text] [doi: [10.1186/1476-072X-13-13](https://doi.org/10.1186/1476-072X-13-13)] [Medline: [24885128](https://pubmed.ncbi.nlm.nih.gov/24885128/)]
12. Anno S, Imaoka K, Tadono T, Igarashi T, Sivaganesh S, Kannathasan S, et al. Space-time clustering characteristics of dengue based on ecological, socio-economic and demographic factors in northern Sri Lanka. *Geospat Health* 2015 Nov 26;10(2):376 [FREE Full text] [doi: [10.4081/gh.2015.376](https://doi.org/10.4081/gh.2015.376)] [Medline: [26618322](https://pubmed.ncbi.nlm.nih.gov/26618322/)]
13. Arboleda S, Jaramillo- O, Peterson AT. Mapping environmental dimensions of dengue fever transmission risk in the Aburrá Valley, Colombia. *Int J Environ Res Public Health* 2009 Dec;6(12):3040-3055 [FREE Full text] [doi: [10.3390/ijerph6123040](https://doi.org/10.3390/ijerph6123040)] [Medline: [20049244](https://pubmed.ncbi.nlm.nih.gov/20049244/)]
14. Asadi SS, Vuppala P, Reddy MA. Remote sensing and GIS techniques for evaluation of groundwater quality in municipal corporation of Hyderabad (Zone-V), India. *Int J Environ Res Public Health* 2007 Mar;4(1):45-52 [FREE Full text] [Medline: [17431315](https://pubmed.ncbi.nlm.nih.gov/17431315/)]
15. Assaré RK, Lai Y, Yapi A, Tian-Bi YT, Ouattara M, Yao PK, et al. The spatial distribution of Schistosoma mansoni infection in four regions of western Côte d'Ivoire. *Geospat Health* 2015 Jun 03;10(1):345 [FREE Full text] [doi: [10.4081/gh.2015.345](https://doi.org/10.4081/gh.2015.345)] [Medline: [26054523](https://pubmed.ncbi.nlm.nih.gov/26054523/)]
16. Aziz S, Aidil RM, Nisfariza MN, Ngui R, Lim YAL, Yusoff WSW, et al. Spatial density of Aedes distribution in urban areas: a case study of breteau index in Kuala Lumpur, Malaysia. *J Vector Borne Dis* 2014 Jun;51(2):91-96 [FREE Full text] [Medline: [24947215](https://pubmed.ncbi.nlm.nih.gov/24947215/)]
17. Baeza A, Bouma MJ, Dhiman RC, Baskerville EB, Ceccato P, Yadav RS, et al. Long-lasting transition toward sustainable elimination of desert malaria under irrigation development. *Proc Natl Acad Sci U S A* 2013 Sep 10;110(37):15157-15162 [FREE Full text] [doi: [10.1073/pnas.1305728110](https://doi.org/10.1073/pnas.1305728110)] [Medline: [23942131](https://pubmed.ncbi.nlm.nih.gov/23942131/)]
18. Beck LR, Lobitz BM, Wood BL. Remote sensing and human health: new sensors and new opportunities. *Emerg Infect Dis* 2000;6(3):217-227 [FREE Full text] [doi: [10.3201/eid0603.000301](https://doi.org/10.3201/eid0603.000301)] [Medline: [10827111](https://pubmed.ncbi.nlm.nih.gov/10827111/)]
19. Bhunia GS, Kesari S, Jeyaram A, Kumar V, Das P. Influence of topography on the endemicity of Kala-azar: a study based on remote sensing and geographical information system. *Geospat Health* 2010 May;4(2):155-165. [doi: [10.4081/gh.2010.197](https://doi.org/10.4081/gh.2010.197)] [Medline: [20503185](https://pubmed.ncbi.nlm.nih.gov/20503185/)]
20. Bo Y, Song C, Wang J, Li X. Using an autologistic regression model to identify spatial risk factors and spatial risk patterns of hand, foot and mouth disease (HFMD) in Mainland China. *BMC Public Health* 2014 Apr 14;14:358 [FREE Full text] [doi: [10.1186/1471-2458-14-358](https://doi.org/10.1186/1471-2458-14-358)] [Medline: [24731248](https://pubmed.ncbi.nlm.nih.gov/24731248/)]
21. Brooker S, Beasley M, Ndinaromtan M, Madjiouroum EM, Baboguel M, Djenguinabe E, et al. Use of remote sensing and a geographical information system in a national helminth control programme in Chad. *Bull World Health Organ* 2002;80(10):783-789 [FREE Full text] [Medline: [12471398](https://pubmed.ncbi.nlm.nih.gov/12471398/)]
22. Brooker S, Michael E. The potential of geographical information systems and remote sensing in the epidemiology and control of human helminth infections. *Advances in Parasitology* 2000;47:245-288. [Medline: [10997209](https://pubmed.ncbi.nlm.nih.gov/10997209/)]
23. Buczak AL, Koshute PT, Babin SM, Feighner BH, Lewis SH. A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. *BMC Med Inform Decis Mak* 2012 Nov 05;12:124 [FREE Full text] [doi: [10.1186/1472-6947-12-124](https://doi.org/10.1186/1472-6947-12-124)] [Medline: [23126401](https://pubmed.ncbi.nlm.nih.gov/23126401/)]
24. Caldas de Castro M, Yamagata Y, Mtasiwa D, Tanner M, Utzinger J, Keiser J, et al. Integrated urban malaria control: a case study in dar es salaam, Tanzania. *Am J Trop Med Hyg* 2004 Aug;71(2 Suppl):103-117. [Medline: [15331826](https://pubmed.ncbi.nlm.nih.gov/15331826/)]
25. Cappelle J, Gaidet N, Iverson SA, Takekawa JY, Newman SH, Fofana B, et al. Characterizing the interface between wild ducks and poultry to evaluate the potential of transmission of avian pathogens. *Int J Health Geogr* 2011 Nov 15;10:60 [FREE Full text] [doi: [10.1186/1476-072X-10-60](https://doi.org/10.1186/1476-072X-10-60)] [Medline: [22085837](https://pubmed.ncbi.nlm.nih.gov/22085837/)]
26. Ceccato P, Connor SJ, Jeanne I, Thomson MC. Application of Geographical Information Systems and Remote Sensing technologies for assessing and monitoring malaria risk. *Parassitologia* 2005 Mar;47(1):81-96. [Medline: [16044677](https://pubmed.ncbi.nlm.nih.gov/16044677/)]
27. Chan L, Cheung GTY, Lauder IJ, Kumana CR, Lauder IJ. Screening for fever by remote-sensing infrared thermographic camera. *J Travel Med* 2004;11(5):273-279 [FREE Full text] [Medline: [15544710](https://pubmed.ncbi.nlm.nih.gov/15544710/)]

28. Cheng C, Wei Y, Sun X, Zhou Y. Estimation of chlorophyll-a concentration in Turbid Lake using spectral smoothing and derivative analysis. *Int J Environ Res Public Health* 2013 Jul 16;10(7):2979-2994 [[FREE Full text](#)] [doi: [10.3390/ijerph10072979](#)] [Medline: [23880727](#)]
29. Chiang M, Lin P, Lin L, Chiou H, Chien C, Chu S, et al. Mass screening of suspected febrile patients with remote-sensing infrared thermography: alarm temperature and optimal distance. *J Formos Med Assoc* 2008 Dec;107(12):937-944 [[FREE Full text](#)] [doi: [10.1016/S0929-6646\(09\)60017-6](#)] [Medline: [19129054](#)]
30. Clarke K, Osleeb J, Sherry J, Meert J, Larsson R. The use of remote sensing and geographic information systems in UNICEF's dracunculiasis (Guinea worm) eradication effort. *Preventive Veterinary Medicine* 1991 Dec;11(3-4):229-235. [doi: [10.1016/S0167-5877\(05\)80007-8](#)] [Medline: [15331826](#)]
31. Clennon JA, Kamanga A, Musapa M, Shiff C, Glass GE. Identifying malaria vector breeding habitats with remote sensing data and terrain-based landscape indices in Zambia. *Int J Health Geogr* 2010 Nov 05;9:58. [doi: [10.1186/1476-072X-9-58](#)] [Medline: [21050496](#)]
32. Corner RJ, Dewan AM, Hashizume M. Modelling typhoid risk in Dhaka metropolitan area of Bangladesh: the role of socio-economic and environmental factors. *Int J Health Geogr* 2013 Mar 16;12:13 [[FREE Full text](#)] [doi: [10.1186/1476-072X-12-13](#)] [Medline: [23497202](#)]
33. Dambach P, Machault V, Lacaux J, Vignolles C, Sié A, Sauerborn R. Utilization of combined remote sensing techniques to detect environmental variables influencing malaria vector densities in rural West Africa. *Int J Health Geogr* 2012 Mar 23;11:8 [[FREE Full text](#)] [doi: [10.1186/1476-072X-11-8](#)] [Medline: [22443452](#)]
34. Dambach P, Sié A, Lacaux JP, Vignolles C, Machault V, Sauerborn R. Using high spatial resolution remote sensing for risk mapping of malaria occurrence in the Nouna district, Burkina Faso. *Glob Health Action* 2009 Nov 11;2:1 [[FREE Full text](#)] [doi: [10.3402/gha.v2i0.2094](#)] [Medline: [20052428](#)]
35. de Oliveira ES, dos Santos ES, Zeilhofer P, Souza-Santos R, Atanaka-Santos M. Geographic information systems and logistic regression for high-resolution malaria risk mapping in a rural settlement of the southern Brazilian Amazon. *Malar J* 2013 Nov 15;12:420 [[FREE Full text](#)] [doi: [10.1186/1475-2875-12-420](#)] [Medline: [24237621](#)]
36. De Roek E, Van Coillie F, De Wulf R, Soenen K, Charlier J, Vercruysse J, et al. Fine-scale mapping of vector habitats using very high resolution satellite imagery: a liver fluke case-study. *Geospat Health* 2014 Dec 01;8(3):S671-S683 [[FREE Full text](#)] [doi: [10.4081/gh.2014.296](#)] [Medline: [25599638](#)]
37. de Souza Gomes E, Leal-Neto OB, Albuquerque J, Pereira da Silva H, Barbosa CS. Schistosomiasis transmission and environmental change: a spatio-temporal analysis in Porto de Galinhas, Pernambuco--Brazil. *Int J Health Geogr* 2012 Nov 20;11:51 [[FREE Full text](#)] [doi: [10.1186/1476-072X-11-51](#)] [Medline: [23164247](#)]
38. Diuk-Wasser MA, Dolo G, Bagayoko M, Sogoba N, Toure MB, Moghaddam M, et al. Patterns of irrigated rice growth and malaria vector breeding in Mali using multi-temporal ERS-2 synthetic aperture radar. *Int J Remote Sens* 2006 Feb;27(3):535-548 [[FREE Full text](#)] [doi: [10.1080/01431160500104350](#)] [Medline: [17710188](#)]
39. Dlamini SN, Franke J, Vounatsou P. Assessing the relationship between environmental factors and malaria vector breeding sites in Swaziland using multi-scale remotely sensed data. *Geospat Health* 2015 Jun 03;10(1):302 [[FREE Full text](#)] [doi: [10.4081/gh.2015.302](#)] [Medline: [26054511](#)]
40. Ebhuoma O, Gebreslasie M. Remote Sensing-Driven Climatic/Environmental Variables for Modelling Malaria Transmission in Sub-Saharan Africa. *Int J Environ Res Public Health* 2016 Dec 14;13(6):pii: E58 [[FREE Full text](#)] [doi: [10.3390/ijerph13060584](#)] [Medline: [27314369](#)]
41. Ehlkes L, Krefis AC, Kreuels B, Krumkamp R, Adjei O, Ayim-Akonor M, et al. Geographically weighted regression of land cover determinants of Plasmodium falciparum transmission in the Ashanti Region of Ghana. *Int J Health Geogr* 2014 Sep 30;13:35 [[FREE Full text](#)] [doi: [10.1186/1476-072X-13-35](#)] [Medline: [25270342](#)]
42. Eisele TP, Keating J, Swalm C, Mbogo CM, Githeko AK, Regens JL, et al. Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya. *Malar J* 2003 Dec 10;2(1):44 [[FREE Full text](#)] [doi: [10.1186/1475-2875-2-44](#)] [Medline: [14667243](#)]
43. Eisenberg MC, Kujbida G, Tuite AR, Fisman DN, Tien JH. Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches. *Epidemics* 2013 Dec;5(4):197-207 [[FREE Full text](#)] [doi: [10.1016/j.epidem.2013.09.004](#)] [Medline: [24267876](#)]
44. Espinosa M, Weinberg D, Rotela CH, Polop F, Abril M, Scavuzzo CM. Temporal Dynamics and Spatial Patterns of Aedes aegypti Breeding Sites, in the Context of a Dengue Control Program in Tartagal (Salta Province, Argentina). *PLoS Negl Trop Dis* 2016 Dec;10(5):e0004621 [[FREE Full text](#)] [doi: [10.1371/journal.pntd.0004621](#)] [Medline: [27223693](#)]
45. Estrada-Peña A. Increasing habitat suitability in the United States for the tick that transmits Lyme disease: a remote sensing approach. *Environ Health Perspect* 2002 Jul;110(7):635-640 [[FREE Full text](#)] [Medline: [12117639](#)]
46. Franke J, Gebreslasie M, Bauwens I, Deleu J, Siegert F. Earth observation in support of malaria control and epidemiology: MALAREO monitoring approaches. *Geospat Health* 2015 Jun 03;10(1):335 [[FREE Full text](#)] [doi: [10.4081/gh.2015.335](#)] [Medline: [26054520](#)]
47. Fuentes MV. Remote sensing and climate data as a key for understanding fasciolosis transmission in the Andes: review and update of an ongoing interdisciplinary project. *Geospat Health* 2006 Nov;1(1):59-70. [doi: [10.4081/gh.2006.281](#)] [Medline: [18686232](#)]

48. Gebreslasie MT. A review of spatial technologies with applications for malaria transmission modelling and control in Africa. *Geospat Health* 2015 Nov 26;10(2):328 [FREE Full text] [doi: [10.4081/gh.2015.328](https://doi.org/10.4081/gh.2015.328)] [Medline: [26618308](https://pubmed.ncbi.nlm.nih.gov/26618308/)]
49. Giardina F, Franke J, Vounatsou P. Geostatistical modelling of the malaria risk in Mozambique: effect of the spatial resolution when using remotely-sensed imagery. *Geospat Health* 2015 Nov 26;10(2):333 [FREE Full text] [doi: [10.4081/gh.2015.333](https://doi.org/10.4081/gh.2015.333)] [Medline: [26618310](https://pubmed.ncbi.nlm.nih.gov/26618310/)]
50. Gilbert M, Newman SH, Takekawa JY, Loth L, Biradar C, Prosser DJ, et al. Flying over an infected landscape: distribution of highly pathogenic avian influenza H5N1 risk in South Asia and satellite tracking of wild waterfowl. *Ecohealth* 2010 Dec;7(4):448-458 [FREE Full text] [doi: [10.1007/s10393-010-0672-8](https://doi.org/10.1007/s10393-010-0672-8)] [Medline: [21267626](https://pubmed.ncbi.nlm.nih.gov/21267626/)]
51. Goetz S, Prince SD, Small J. Advances in satellite remote sensing of environmental variables for epidemiological applications. *Adv Parasitol* 2000;47:289-307. [Medline: [10997210](https://pubmed.ncbi.nlm.nih.gov/10997210/)]
52. Grimes DJ, Ford TE, Colwell RR, Baker-Austin C, Martinez-Urtaza J, Subramaniam A, et al. Viewing marine bacteria, their activity and response to environmental drivers from orbit: satellite remote sensing of bacteria. *Microb Ecol* 2014 Apr;67(3):489-500 [FREE Full text] [doi: [10.1007/s00248-013-0363-4](https://doi.org/10.1007/s00248-013-0363-4)] [Medline: [24477922](https://pubmed.ncbi.nlm.nih.gov/24477922/)]
53. Guimarães RJDP, Freitas CC, Dutra LV, Scholte RGC, Martins-Bedé FT, Fonseca FR, et al. A geoprocessing approach for studying and controlling schistosomiasis in the state of Minas Gerais, Brazil. *Mem Inst Oswaldo Cruz* 2010 Jul;105(4):524-531 [FREE Full text] [Medline: [20721503](https://pubmed.ncbi.nlm.nih.gov/20721503/)]
54. Guo J, Vounatsou P, Cao C, Utzinger J, Zhu H, Anderegg D, et al. A geographic information and remote sensing based model for prediction of *Oncomelania hupensis* habitats in the Poyang Lake area, China. *Acta Trop* 2005;96(2-3):213-222. [doi: [10.1016/j.actatropica.2005.07.029](https://doi.org/10.1016/j.actatropica.2005.07.029)] [Medline: [16140246](https://pubmed.ncbi.nlm.nih.gov/16140246/)]
55. Hanafi-Bojd AA, Vatandoost H, Oshaghi MA, Charrahy Z, Haghdoost AA, Sedaghat MM, et al. Larval habitats and biodiversity of anopheline mosquitoes (Diptera: Culicidae) in a malarious area of southern Iran. *J Vector Borne Dis* 2012 Jun;49(2):91-100 [FREE Full text] [Medline: [22898481](https://pubmed.ncbi.nlm.nih.gov/22898481/)]
56. Haque U, Magalhães RJS, Reid HL, Clements ACA, Ahmed SM, Islam A, et al. Spatial prediction of malaria prevalence in an endemic area of Bangladesh. *Malar J* 2010 May 09;9:120 [FREE Full text] [doi: [10.1186/1475-2875-9-120](https://doi.org/10.1186/1475-2875-9-120)] [Medline: [20459690](https://pubmed.ncbi.nlm.nih.gov/20459690/)]
57. Hassan AN, Onsi HM. Remote sensing as a tool for mapping mosquito breeding habitats and associated health risk to assist control efforts and development plans: a case study in Wadi El Natroun, Egypt. *J Egypt Soc Parasitol* 2004 Aug;34(2):367-382. [Medline: [15287164](https://pubmed.ncbi.nlm.nih.gov/15287164/)]
58. Hay SI, Lennon JJ. Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Trop Med Int Health* 1999 Jan;4(1):58-71 [FREE Full text] [Medline: [10203175](https://pubmed.ncbi.nlm.nih.gov/10203175/)]
59. Hay S, Packer M, Rogers D. Review article The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. *International Journal of Remote Sensing* 1997 Sep;18(14):2899-2930. [doi: [10.1080/014311697217125](https://doi.org/10.1080/014311697217125)]
60. Hayes RO, Maxwell EL, Mitchell CJ, Woodzick TL. Detection, identification, and classification of mosquito larval habitats using remote sensing scanners in earth-orbiting satellites. *Bull World Health Organ* 1985;63(2):361-374 [FREE Full text] [Medline: [2861917](https://pubmed.ncbi.nlm.nih.gov/2861917/)]
61. Herbreteau V, Salem G, Souris M, Hugot JP, Gonzalez JP. Sizing up human health through remote sensing: uses and misuses. *Parassitologia* 2005 Mar;47(1):63-79. [Medline: [16044676](https://pubmed.ncbi.nlm.nih.gov/16044676/)]
62. Herbreteau V, Salem G, Souris M, Hugot J, Gonzalez J. Thirty years of use and improvement of remote sensing, applied to epidemiology: from early promises to lasting frustration. *Health Place* 2007 Jun;13(2):400-403. [doi: [10.1016/j.healthplace.2006.03.003](https://doi.org/10.1016/j.healthplace.2006.03.003)] [Medline: [16735137](https://pubmed.ncbi.nlm.nih.gov/16735137/)]
63. Hu Y, Li R, Bergquist R, Lynn H, Gao F, Wang Q, et al. Spatio-temporal transmission and environmental determinants of Schistosomiasis Japonica in Anhui Province, China. *PLoS Negl Trop Dis* 2015 Feb;9(2):e0003470 [FREE Full text] [doi: [10.1371/journal.pntd.0003470](https://doi.org/10.1371/journal.pntd.0003470)] [Medline: [25659112](https://pubmed.ncbi.nlm.nih.gov/25659112/)]
64. Hugh-Jones M. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitol Today* 1989 Aug;5(8):244-251. [Medline: [15463225](https://pubmed.ncbi.nlm.nih.gov/15463225/)]
65. Hunter P, Tyler A, Carvalho L, Codd G, Maberly S. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sensing of Environment* 2010 Nov;114(11):2705-2718. [doi: [10.1016/j.rse.2010.06.006](https://doi.org/10.1016/j.rse.2010.06.006)]
66. Hunter PD, Tyler AN, Gilvear DJ, Willby NJ. Using remote sensing to aid the assessment of human health risks from blooms of potentially toxic cyanobacteria. *Environ Sci Technol* 2009 Apr 01;43(7):2627-2633. [Medline: [19452927](https://pubmed.ncbi.nlm.nih.gov/19452927/)]
67. Ippoliti C, Gilbert M, Vanhuysse S, Goffredo M, Satta G, Wolff E, et al. Can landscape metrics help determine the *Culicoides* imicola distribution in Italy? *Geospat Health* 2013 Nov;8(1):267-277 [FREE Full text] [doi: [10.4081/gh.2013.72](https://doi.org/10.4081/gh.2013.72)] [Medline: [24258901](https://pubmed.ncbi.nlm.nih.gov/24258901/)]
68. Jacob BG, Novak RJ, Toe LD, Sanfo M, Griffith DA, Lakwo TL, et al. Validation of a remote sensing model to identify *Simulium damnosum* s.l. breeding sites in Sub-Saharan Africa. *PLoS Negl Trop Dis* 2013;7(7):e2342 [FREE Full text] [doi: [10.1371/journal.pntd.0002342](https://doi.org/10.1371/journal.pntd.0002342)] [Medline: [23936571](https://pubmed.ncbi.nlm.nih.gov/23936571/)]

69. Jagai J, Sarkar R, Castronovo D, Kattula D, McEntee J, Ward H, et al. Seasonality of rotavirus in South Asia: a meta-analysis approach assessing associations with temperature, precipitation, and vegetation index. *PLoS One* 2012;7(5):e38168 [FREE Full text] [doi: [10.1371/journal.pone.0038168](https://doi.org/10.1371/journal.pone.0038168)] [Medline: [22693594](https://pubmed.ncbi.nlm.nih.gov/22693594/)]
70. Jia P, Joyner A. Human brucellosis occurrences in inner mongolia, China: a spatio-temporal distribution and ecological niche modeling approach. *BMC Infect Dis* 2015 Feb 03;15:36 [FREE Full text] [doi: [10.1186/s12879-015-0763-9](https://doi.org/10.1186/s12879-015-0763-9)] [Medline: [25644986](https://pubmed.ncbi.nlm.nih.gov/25644986/)]
71. Jutla A, Akanda AS, Huq A, Faruque ASG, Colwell R, Islam S. A water marker monitored by satellites to predict seasonal endemic cholera. *Remote Sens Lett* 2013;4(8):822-831 [FREE Full text] [doi: [10.1080/2150704X.2013.802097](https://doi.org/10.1080/2150704X.2013.802097)] [Medline: [23878762](https://pubmed.ncbi.nlm.nih.gov/23878762/)]
72. Jutla A, Aldaach H, Billian H, Akanda A, Huq A, Colwell R. Satellite Based Assessment of Hydroclimatic Conditions Related to Cholera in Zimbabwe. *PLoS One* 2015;10(9):e0137828 [FREE Full text] [doi: [10.1371/journal.pone.0137828](https://doi.org/10.1371/journal.pone.0137828)] [Medline: [26417994](https://pubmed.ncbi.nlm.nih.gov/26417994/)]
73. Jutla A, Huq A, Colwell RR. Diagnostic approach for monitoring hydroclimatic conditions related to emergence of west Nile virus in west virginia. *Front Public Health* 2015;3:10 [FREE Full text] [doi: [10.3389/fpubh.2015.00010](https://doi.org/10.3389/fpubh.2015.00010)] [Medline: [25729746](https://pubmed.ncbi.nlm.nih.gov/25729746/)]
74. Jutla A, Akanda AS, Islam S. Tracking Cholera in Coastal Regions using Satellite Observations. *J Am Water Resour Assoc* 2010 Aug;46(4):651-662 [FREE Full text] [doi: [10.1111/j.1752-1688.2010.00448.x](https://doi.org/10.1111/j.1752-1688.2010.00448.x)] [Medline: [21072249](https://pubmed.ncbi.nlm.nih.gov/21072249/)]
75. Jutla A, Akanda AS, Islam S. Satellite Remote Sensing of Space-Time Plankton Variability in the Bay of Bengal: Connections to Cholera Outbreaks. *Remote Sens Environ* 2012 Aug;123:196-206 [FREE Full text] [doi: [10.1016/j.rse.2012.03.005](https://doi.org/10.1016/j.rse.2012.03.005)] [Medline: [22544976](https://pubmed.ncbi.nlm.nih.gov/22544976/)]
76. Kabaria C, Molteni F, Mandike R, Chacky F, Noor AM, Snow RW, et al. Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam. *Int J Health Geogr* 2016 Dec 30;15(1):26 [FREE Full text] [doi: [10.1186/s12942-016-0051-y](https://doi.org/10.1186/s12942-016-0051-y)] [Medline: [27473186](https://pubmed.ncbi.nlm.nih.gov/27473186/)]
77. Kalluri S, Gilruth P, Rogers D, Szczur M. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: a review. *PLoS Pathog* 2007 Oct 26;3(10):1361-1371 [FREE Full text] [doi: [10.1371/journal.ppat.0030116](https://doi.org/10.1371/journal.ppat.0030116)] [Medline: [17967056](https://pubmed.ncbi.nlm.nih.gov/17967056/)]
78. Kesari S, Bhunia G, Chatterjee N, Kumar V, Mandal R, Das P. Appraisal of Phlebotomus argentipes habitat suitability using a remotely sensed index in the kala-azar endemic focus of Bihar, India. *Mem Inst Oswaldo Cruz* 2013 Apr;108(2):197-204 [FREE Full text] [doi: [10.1590/0074-0276108022013012](https://doi.org/10.1590/0074-0276108022013012)] [Medline: [23579800](https://pubmed.ncbi.nlm.nih.gov/23579800/)]
79. Kiewra D, Lonc E. Epidemiological consequences of host specificity of ticks (Ixodida). *Ann Parasitol* 2012;58(4):181-187 [FREE Full text] [Medline: [23914612](https://pubmed.ncbi.nlm.nih.gov/23914612/)]
80. Kirpich A, Weppelmann TA, Yang Y, Ali A, Morris JG, Longini IM. Cholera Transmission in Ouest Department of Haiti: Dynamic Modeling and the Future of the Epidemic. *PLoS Negl Trop Dis* 2015;9(10):e0004153 [FREE Full text] [doi: [10.1371/journal.pntd.0004153](https://doi.org/10.1371/journal.pntd.0004153)] [Medline: [26488620](https://pubmed.ncbi.nlm.nih.gov/26488620/)]
81. Kutser T. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnol. Oceanogr* 2004 Nov 12;49(6):2179-2189. [doi: [10.4319/lo.2004.49.6.2179](https://doi.org/10.4319/lo.2004.49.6.2179)]
82. Kutser T, Metsamaa L, Strömbeck N, Vahtmäe E. Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine, Coastal and Shelf Science* 2006 Mar;67(1-2):303-312. [doi: [10.1016/j.ecss.2005.11.024](https://doi.org/10.1016/j.ecss.2005.11.024)]
83. Kzar A, Mat JMZ, Mutter KN, Syahreza S. A Modified Hopfield Neural Network Algorithm (MHNNA) Using ALOS Image for Water Quality Mapping. *Int J Environ Res Public Health* 2015 Dec 30;13(1):pii-E92 [FREE Full text] [doi: [10.3390/ijerph13010092](https://doi.org/10.3390/ijerph13010092)] [Medline: [26729148](https://pubmed.ncbi.nlm.nih.gov/26729148/)]
84. Lafaye M, Sall B, Ndiaye Y, Vignolles C, Tourre YM, Borchio FO, et al. Rift Valley fever dynamics in Senegal: a project for pro-active adaptation and improvement of livestock raising management. *Geospat Health* 2013 Nov;8(1):279-288 [FREE Full text] [doi: [10.4081/gh.2013.73](https://doi.org/10.4081/gh.2013.73)] [Medline: [24258902](https://pubmed.ncbi.nlm.nih.gov/24258902/)]
85. Lai Y, Zhou X, Utzinger J, Vounatsou P. Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. *Parasit Vectors* 2013 Dec 18;6:359 [FREE Full text] [doi: [10.1186/1756-3305-6-359](https://doi.org/10.1186/1756-3305-6-359)] [Medline: [24350825](https://pubmed.ncbi.nlm.nih.gov/24350825/)]
86. Lauer A, Talamantes J, Castañón Olivares L, Medina L, Baal J, Casimiro K, et al. Combining Forces - The Use of Landsat TM Satellite Imagery, Soil Parameter Information, and Multiplex PCR to Detect Coccidioides immitis Growth Sites in Kern County, California. *PLoS ONE* 2014 Nov 7;9(11):e111921. [doi: [10.1371/journal.pone.0111921](https://doi.org/10.1371/journal.pone.0111921)]
87. Li Z, Roux E, Dessay N, Girod R, Stefani A, Nacher M, et al. Mapping a Knowledge-Based Malaria Hazard Index Related to Landscape Using Remote Sensing: Application to the Cross-Border Area between French Guiana and Brazil. *Remote Sensing* 2016 Apr 11;8(4):319. [doi: [10.3390/rs8040319](https://doi.org/10.3390/rs8040319)]
88. Lobitz B, Beck L, Huq A, Wood B, Fuchs G, Faruque AS, et al. Climate and infectious disease: use of remote sensing for detection of Vibrio cholerae by indirect measurement. *Proc Natl Acad Sci U S A* 2000 Feb 15;97(4):1438-1443 [FREE Full text] [Medline: [10677480](https://pubmed.ncbi.nlm.nih.gov/10677480/)]
89. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, et al. Modeling tools for dengue risk mapping - a systematic review. *Int J Health Geogr* 2014 Dec 09;13:50 [FREE Full text] [doi: [10.1186/1476-072X-13-50](https://doi.org/10.1186/1476-072X-13-50)] [Medline: [25487167](https://pubmed.ncbi.nlm.nih.gov/25487167/)]

90. Machault V, Orlandi-Pradines E, Michel R, Pagès F, Texier G, Pradines B, et al. Remote sensing and malaria risk for military personnel in Africa. *J Travel Med* 2008;15(4):216-220 [[FREE Full text](#)] [doi: [10.1111/j.1708-8305.2008.00202.x](https://doi.org/10.1111/j.1708-8305.2008.00202.x)] [Medline: [18666920](#)]
91. Machault V, Vignolles C, Borchì F, Vounatsou P, Pages F, Briolant S, et al. The use of remotely sensed environmental data in the study of malaria. *Geospat Health* 2011 May;5(2):151-168. [doi: [10.4081/gh.2011.167](https://doi.org/10.4081/gh.2011.167)] [Medline: [21590665](#)]
92. Machault V, Vignolles C, Pagès F, Gadiaga L, Gaye A, Sokhna C, et al. Spatial heterogeneity and temporal evolution of malaria transmission risk in Dakar, Senegal, according to remotely sensed environmental data. *Malar J* 2010 Sep 03;9:252 [[FREE Full text](#)] [doi: [10.1186/1475-2875-9-252](https://doi.org/10.1186/1475-2875-9-252)] [Medline: [20815867](#)]
93. Machault V, Vignolles C, Pagès F, Gadiaga L, Tourre YM, Gaye A, et al. Risk mapping of *Anopheles gambiae* s.l. densities using remotely-sensed environmental and meteorological data in an urban area: Dakar, Senegal. *PLoS One* 2012;7(11):e50674 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0050674](https://doi.org/10.1371/journal.pone.0050674)] [Medline: [23226351](#)]
94. Machault V, Yébakima A, Etienne M, Vignolles C, Palany P, Tourre Y, et al. Mapping Entomological Dengue Risk Levels in Martinique Using High-Resolution Remote-Sensing Environmental Data. *ISPRS International Journal of Geo-Information* 2014;3(4):1352.
95. Manyangadze T, Chimbari MJ, Gebreslasie M, Mukaratirwa S. Application of geo-spatial technology in schistosomiasis modelling in Africa: a review. *Geospat Health* 2015 Nov 04;10(2):326 [[FREE Full text](#)] [doi: [10.4081/gh.2015.326](https://doi.org/10.4081/gh.2015.326)] [Medline: [26618307](#)]
96. Marcantonio M, Rizzoli A, Metz M, Rosà R, Marini G, Chadwick E, et al. Identifying the environmental conditions favouring West Nile Virus outbreaks in Europe. *PLoS One* 2015;10(3):e0121158 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0121158](https://doi.org/10.1371/journal.pone.0121158)] [Medline: [25803814](#)]
97. Mattikalli N, Richards K. Estimation of Surface Water Quality Changes in Response to Land Use Change: Application of The Export Coefficient Model Using Remote Sensing and Geographical Information System. *Journal of Environmental Management* 1996 Nov;48(3):263-282. [doi: [10.1006/jema.1996.0077](https://doi.org/10.1006/jema.1996.0077)]
98. Méndez-Lázaro P, Muller-Karger FE, Otis D, McCarthy MJ, Peña-Orellana M. Assessing climate variability effects on dengue incidence in San Juan, Puerto Rico. *Int J Environ Res Public Health* 2014 Sep 11;11(9):9409-9428 [[FREE Full text](#)] [doi: [10.3390/ijerph110909409](https://doi.org/10.3390/ijerph110909409)] [Medline: [25216253](#)]
99. Midekisa A, Senay G, Henebry GM, Semuniguse P, Wimberly MC. Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. *Malar J* 2012 May 14;11:165 [[FREE Full text](#)] [doi: [10.1186/1475-2875-11-165](https://doi.org/10.1186/1475-2875-11-165)] [Medline: [22583705](#)]
100. Midekisa A, Senay GB, Wimberly MC. Multisensor earth observations to characterize wetlands and malaria epidemiology in Ethiopia. *Water Resour Res* 2014 Nov;50(11):8791-8806 [[FREE Full text](#)] [doi: [10.1002/2014WR015634](https://doi.org/10.1002/2014WR015634)] [Medline: [25653462](#)]
101. Moloney JM, Skelly C, Weinstein P, Maguire M, Ritchie S. Domestic *Aedes aegypti* breeding site surveillance: limitations of remote sensing as a predictive surveillance tool. *Am J Trop Med Hyg* 1998 Aug;59(2):261-264. [Medline: [9715943](#)]
102. Moss WJ, Hamapumbu H, Kobayashi T, Shields T, Kamanga A, Clennon J, et al. Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: a cross-sectional and longitudinal community survey. *Malar J* 2011 Jun 10;10:163 [[FREE Full text](#)] [doi: [10.1186/1475-2875-10-163](https://doi.org/10.1186/1475-2875-10-163)] [Medline: [21663661](#)]
103. Mushinzimana E, Munga S, Minakawa N, Li L, Feng C, Bian L, et al. Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. *Malar J* 2006 Feb 16;5:13 [[FREE Full text](#)] [doi: [10.1186/1475-2875-5-13](https://doi.org/10.1186/1475-2875-5-13)] [Medline: [16480523](#)]
104. Myers M, Rogers DJ, Cox J, Flahault A, Hay SI. Forecasting disease risk for increased epidemic preparedness in public health. *Adv Parasitol* 2000;47:309-330 [[FREE Full text](#)] [Medline: [10997211](#)]
105. Nygren D, Stoyanov C, Lewold C, Månsson F, Miller J, Kamanga A, et al. Remotely-sensed, nocturnal, dew point correlates with malaria transmission in Southern Province, Zambia: a time-series study. *Malar J* 2014 Jun 13;13:231 [[FREE Full text](#)] [doi: [10.1186/1475-2875-13-231](https://doi.org/10.1186/1475-2875-13-231)] [Medline: [24927747](#)]
106. Oliveira EFD, Silva EAE, Fernandes CEDS, Paranhos FAC, Gamarra RM, Ribeiro AA, et al. Biotic factors and occurrence of *Lutzomyia longipalpis* in endemic area of visceral leishmaniasis, Mato Grosso do Sul, Brazil. *Mem Inst Oswaldo Cruz* 2012 May;107(3):396-401 [[FREE Full text](#)] [Medline: [22510836](#)]
107. Omumbo JA, Hay SI, Goetz SJ, Snow RW, Rogers DJ. Updating Historical Maps of Malaria Transmission Intensity in East Africa Using Remote Sensing. *Photogramm Eng Remote Sensing* 2002 Feb;68(2):161-166 [[FREE Full text](#)] [Medline: [23814324](#)]
108. Ozdenerol E. GIS and Remote Sensing Use in the Exploration of Lyme Disease Epidemiology. *Int J Environ Res Public Health* 2015 Dec 01;12(12):15182-15203 [[FREE Full text](#)] [doi: [10.3390/ijerph121214971](https://doi.org/10.3390/ijerph121214971)] [Medline: [26633445](#)]
109. Masimalai P. Remote sensing and Geographic Information Systems (GIS) as the applied public health and environmental epidemiology. *Int J Med Sci Public Health* 2014;3(12):1430. [doi: [10.5455/ijmsph.2014.081020141](https://doi.org/10.5455/ijmsph.2014.081020141)]
110. Polo G, Labruna MB, Ferreira F. Satellite Hyperspectral Imagery to Support Tick-Borne Infectious Diseases Surveillance. *PLoS One* 2015;10(11):e0143736 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0143736](https://doi.org/10.1371/journal.pone.0143736)] [Medline: [26599337](#)]
111. Pope K, Rejmankova E, Savage HM, Arredondo-Jimenez JI, Rodriguez MH, Roberts DR. Remote sensing of tropical wetlands for malaria control in Chiapas, Mexico. *Ecol Appl* 1994 Feb;4(1):81-90. [Medline: [11539870](#)]

112. Rahman A, Kogan F, Roytman L. Analysis of malaria cases in Bangladesh with remote sensing data. *Am J Trop Med Hyg* 2006 Jan;74(1):17-19. [Medline: [16407340](#)]
113. Rahman A, Krakauer N, Roytman L, Goldberg M, Kogan F. Application of advanced very high resolution radiometer (AVHRR)-based vegetation health indices for estimation of malaria cases. *Am J Trop Med Hyg* 2010 Jun;82(6):1004-1009 [FREE Full text] [doi: [10.4269/ajtmh.2010.09-0201](#)] [Medline: [20519592](#)]
114. Rai PK, Nathawat MS, Rai S. Using the information value method in a geographic information system and remote sensing for malaria mapping: a case study from India. *Inform Prim Care* 2013;21(1):43-52 [FREE Full text] [doi: [10.14236/jhi.v21i1.38](#)] [Medline: [24629656](#)]
115. Rakotomanana F, Ratovonjato J, Randremanana RV, Randrianasolo L, Raheerinjafy R, Rudant J, et al. Geographical and environmental approaches to urban malaria in Antananarivo (Madagascar). *BMC Infect Dis* 2010 Jun 16;10:173 [FREE Full text] [doi: [10.1186/1471-2334-10-173](#)] [Medline: [20553598](#)]
116. Samson DM, Archer RS, Alimi TO, Arheart KL, Impoinvil DE, Oscar R, et al. New baseline environmental assessment of mosquito ecology in northern Haiti during increased urbanization. *J Vector Ecol* 2015 Jun;40(1):46-58 [FREE Full text] [doi: [10.1111/jvec.12131](#)] [Medline: [26047183](#)]
117. Sarfraz M, Tripathi NK, Faruque FS, Bajwa UI, Kitamoto A, Souris M. Mapping urban and peri-urban breeding habitats of Aedes mosquitoes using a fuzzy analytical hierarchical process based on climatic and physical parameters. *Geospat Health* 2014 Dec 01;8(3):S685-S697 [FREE Full text] [doi: [10.4081/gh.2014.297](#)] [Medline: [25599639](#)]
118. Saxena R, Das MK, Nagpal BN, Srivastava A, Gupta SK, Kumar A, et al. Identification of risk factors for malaria control by focused interventions in Ranchi district, Jharkhand, India. *J Vector Borne Dis* 2014 Dec;51(4):276-281 [FREE Full text] [Medline: [25540958](#)]
119. Saxena R, Nagpal BN, Singh VP, Srivastava A, Dev V, Sharma MC, et al. Impact of deforestation on known malaria vectors in Sonitpur district of Assam, India. *J Vector Borne Dis* 2014 Sep;51(3):211-215 [FREE Full text] [Medline: [25253214](#)]
120. Scholte RGC, Schur N, Bavia ME, Carvalho EM, Chammartin F, Utzinger J, et al. Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models. *Geospat Health* 2013 Nov;8(1):97-110 [FREE Full text] [doi: [10.4081/gh.2013.58](#)] [Medline: [24258887](#)]
121. Schuster G, Ebert EE, Stevenson MA, Corner RJ, Johansen CA. Application of satellite precipitation data to analyse and model arbovirus activity in the tropics. *Int J Health Geogr* 2011 Jan 22;10:8 [FREE Full text] [doi: [10.1186/1476-072X-10-8](#)] [Medline: [21255449](#)]
122. Sewe M, Ahlm C, Rocklöv J. Remotely Sensed Environmental Conditions and Malaria Mortality in Three Malaria Endemic Regions in Western Kenya. *PLoS One* 2016;11(4):e0154204 [FREE Full text] [doi: [10.1371/journal.pone.0154204](#)] [Medline: [27115874](#)]
123. Simis S, Ruiz-Verdú A, Domínguez-Gómez J, Peña-Martínez R, Peters S, Gons H. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. *Remote Sensing of Environment* 2007 Feb;106(4):414-427. [doi: [10.1016/j.rse.2006.09.008](#)]
124. Simoonga C, Utzinger J, Brooker S, Vounatsou P, Appleton CC, Stensgaard AS, et al. Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. *Parasitology* 2009 Nov;136(13):1683-1693 [FREE Full text] [doi: [10.1017/S0031182009006222](#)] [Medline: [19627627](#)]
125. Soti V, Chevalier V, Maura J, Bégué A, Lelong C, Lancelot R, et al. Identifying landscape features associated with Rift Valley fever virus transmission, Ferlo region, Senegal, using very high spatial resolution satellite imagery. *Int J Health Geogr* 2013 Mar 01;12:10 [FREE Full text] [doi: [10.1186/1476-072X-12-10](#)] [Medline: [23452759](#)]
126. Spear RC, Hubbard A, Liang S, Seto E. Disease transmission models for public health decision making: toward an approach for designing intervention strategies for Schistosomiasis japonica. *Environ Health Perspect* 2002 Sep;110(9):907-915 [FREE Full text] [Medline: [12204826](#)]
127. Spear RC, Seto E, Liang S, Birkner M, Hubbard A, Qiu D, et al. Factors influencing the transmission of Schistosoma japonicum in the mountains of Sichuan Province of China. *Am J Trop Med Hyg* 2004 Jan;70(1):48-56. [Medline: [14971698](#)]
128. Stadler H, Skritek P. Remote water quality monitoring "on-line" using LEO satellites. *Water Sci Technol* 2003;47(2):197-204. [Medline: [12636080](#)]
129. Stefani A, Dusfour I, Corrêa APSA, Cruz MCB, Dessay N, Galardo AKR, et al. Land cover, land use and malaria in the Amazon: a systematic literature review of studies using remotely sensed data. *Malar J* 2013 Jun 08;12:192 [FREE Full text] [doi: [10.1186/1475-2875-12-192](#)] [Medline: [23758827](#)]
130. Stefani A, Roux E, Fotsing J, Carme B. Studying relationships between environment and malaria incidence in Camopi (French Guiana) through the objective selection of buffer-based landscape characterisations. *Int J Health Geogr* 2011 Dec 13;10:65 [FREE Full text] [doi: [10.1186/1476-072X-10-65](#)] [Medline: [22151738](#)]
131. Tatem AJ, Hay SI. Measuring urbanization pattern and extent for malaria research: a review of remote sensing approaches. *J Urban Health* 2004 Sep;81(3):363-376 [FREE Full text] [doi: [10.1093/jurban/jth124](#)] [Medline: [15273262](#)]
132. Texier G, Machault V, Barragti M, Boutin JP, Rogier C. Environmental determinant of malaria cases among travellers. *Malar J* 2013 Mar 04;12:87 [FREE Full text] [doi: [10.1186/1475-2875-12-87](#)] [Medline: [23496931](#)]

133. Thanapongtharm W, Linard C, Wiriyarat W, Chinsorn P, Kanchanasaka B, Xiao X, et al. Spatial characterization of colonies of the flying fox bat, a carrier of Nipah virus in Thailand. *BMC Vet Res* 2015 Mar 28;11:81 [[FREE Full text](#)] [doi: [10.1186/s12917-015-0390-0](#)] [Medline: [25880385](#)]
134. Thanapongtharm W, Van BTP, Biradar C, Xiao X, Gilbert M. Rivers and flooded areas identified by medium-resolution remote sensing improve risk prediction of the highly pathogenic avian influenza H5N1 in Thailand. *Geospat Health* 2013 Nov;8(1):193-201 [[FREE Full text](#)] [doi: [10.4081/gh.2013.66](#)] [Medline: [24258895](#)]
135. Thomas CJ, Lindsay SW. Local-scale variation in malaria infection amongst rural Gambian children estimated by satellite remote sensing. *Trans R Soc Trop Med Hyg* 2000;94(2):159-163. [Medline: [10897355](#)]
136. Tian H, Cui Y, Dong L, Zhou S, Li X, Huang S, et al. Spatial, temporal and genetic dynamics of highly pathogenic avian influenza A (H5N1) virus in China. *BMC Infect Dis* 2015 Feb 13;15:54 [[FREE Full text](#)] [doi: [10.1186/s12879-015-0770-x](#)] [Medline: [25887370](#)]
137. Torbick N, Corbiere M. A Multiscale Mapping Assessment of Lake Champlain Cyanobacterial Harmful Algal Blooms. *Int J Environ Res Public Health* 2015 Sep 15;12(9):11560-11578 [[FREE Full text](#)] [doi: [10.3390/ijerph120911560](#)] [Medline: [26389930](#)]
138. Torbick N, Hession S, Stommel E, Caller T. Mapping amyotrophic lateral sclerosis lake risk factors across northern New England. *Int J Health Geogr* 2014 Jan 02;13:1 [[FREE Full text](#)] [doi: [10.1186/1476-072X-13-1](#)] [Medline: [24383521](#)]
139. Tourre Y, Lacaux JP, Vignolles C, Lafaye M. Climate impacts on environmental risks evaluated from space: a conceptual approach to the case of Rift Valley Fever in Senegal. *Glob Health Action* 2009 Nov 11;2:e [[FREE Full text](#)] [doi: [10.3402/gha.v2i0.2053](#)] [Medline: [20052381](#)]
140. Tran A, L'Ambert G, Lacour G, Benoît R, Demarchi M, Cros M, et al. A rainfall- and temperature-driven abundance model for *Aedes albopictus* populations. *Int J Environ Res Public Health* 2013 Apr 26;10(5):1698-1719 [[FREE Full text](#)] [doi: [10.3390/ijerph10051698](#)] [Medline: [23624579](#)]
141. Tran A, Ponçon N, Toty C, Linard C, Guis H, Ferré J, et al. Using remote sensing to map larval and adult populations of *Anopheles hyrcanus* (Diptera: Culicidae) a potential malaria vector in Southern France. *Int J Health Geogr* 2008 Feb 26;7:9 [[FREE Full text](#)] [doi: [10.1186/1476-072X-7-9](#)] [Medline: [18302749](#)]
142. Walz Y, Wegmann M, Dech S, Raso G, Utzinger J. Risk profiling of schistosomiasis using remote sensing: approaches, challenges and outlook. *Parasit Vectors* 2015 Mar 17;8:163 [[FREE Full text](#)] [doi: [10.1186/s13071-015-0732-6](#)] [Medline: [25890278](#)]
143. Walz Y, Wegmann M, Dech S, Vounatsou P, Poda JN, N'Goran EK, et al. Modeling and Validation of Environmental Suitability for Schistosomiasis Transmission Using Remote Sensing. *PLoS Negl Trop Dis* 2015 Nov;9(11):e0004217 [[FREE Full text](#)] [doi: [10.1371/journal.pntd.0004217](#)] [Medline: [26587839](#)]
144. Walz Y, Wegmann M, Leutner B, Dech S, Vounatsou P, N'Goran EK, et al. Use of an ecologically relevant modelling approach to improve remote sensing-based schistosomiasis risk profiling. *Geospat Health* 2015 Nov 30;10(2):398 [[FREE Full text](#)] [doi: [10.4081/gh.2015.398](#)] [Medline: [26618326](#)]
145. Wang Y, Zhuang D. A Rapid Monitoring and Evaluation Method of Schistosomiasis Based on Spatial Information Technology. *Int J Environ Res Public Health* 2015 Dec 12;12(12):15843-15859 [[FREE Full text](#)] [doi: [10.3390/ijerph121215025](#)] [Medline: [26703635](#)]
146. Wang Y, Feng C, Sithithaworn P. Environmental determinants of *Opisthorchis viverrini* prevalence in northeast Thailand. *Geospat Health* 2013 Nov;8(1):111-123 [[FREE Full text](#)] [doi: [10.4081/gh.2013.59](#)] [Medline: [24258888](#)]
147. Washino RK, Wood BL. Application of remote sensing to arthropod vector surveillance and control. *Am J Trop Med Hyg* 1994;50(6 Suppl):134-144. [Medline: [8024079](#)]
148. Watanabe F, Alcântara E, Rodrigues TWP, Imai NN, Barbosa CCF, Rotta LHDS. Estimation of Chlorophyll-a Concentration and the Trophic State of the Barra Bonita Hydroelectric Reservoir Using OLI/Landsat-8 Images. *Int J Environ Res Public Health* 2015 Aug 26;12(9):10391-10417 [[FREE Full text](#)] [doi: [10.3390/ijerph120910391](#)] [Medline: [26322489](#)]
149. Wei L, Qian Q, Wang Z, Glass GE, Song S, Zhang W, et al. Using geographic information system-based ecologic niche models to forecast the risk of hantavirus infection in Shandong Province, China. *Am J Trop Med Hyg* 2011 Mar;84(3):497-503 [[FREE Full text](#)] [doi: [10.4269/ajtmh.2011.10-0314](#)] [Medline: [21363991](#)]
150. Wynne TT, Stumpf RP. Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in Western Lake Erie from 2002-2014. *Toxins (Basel)* 2015 May 12;7(5):1649-1663 [[FREE Full text](#)] [doi: [10.3390/toxins7051649](#)] [Medline: [25985390](#)]
151. Xu M, Cao C, Wang D, Kan B. Identifying environmental risk factors of cholera in a coastal area with geospatial technologies. *Int J Environ Res Public Health* 2014 Dec 29;12(1):354-370 [[FREE Full text](#)] [doi: [10.3390/ijerph120100354](#)] [Medline: [25551518](#)]
152. Yan L, Fang L, Huang H, Zhang L, Feng D, Zhao W, et al. Landscape elements and Hantaan virus-related hemorrhagic fever with renal syndrome, People's Republic of China. *Emerg Infect Dis* 2007 Sep;13(9):1301-1306 [[FREE Full text](#)] [doi: [10.3201/eid1309.061481](#)] [Medline: [18252099](#)]
153. Yang G, Vounatsou P, Tanner M, Zhou X, Utzinger J. Remote sensing for predicting potential habitats of *Oncomelania hupensis* in Hongze, Baima and Gaoyou lakes in Jiangsu province, China. *Geospat Health* 2006 Nov;1(1):85-92. [doi: [10.4081/gh.2006.283](#)] [Medline: [18686234](#)]

154. Yang G, Vounatsou P, Zhou X, Utzinger J, Tanner M. A review of geographic information system and remote sensing with applications to the epidemiology and control of schistosomiasis in China. *Acta Trop* 2005;96(2-3):117-129. [doi: [10.1016/j.actatropica.2005.07.006](https://doi.org/10.1016/j.actatropica.2005.07.006)] [Medline: [16112638](#)]
155. Zhang F, Lee J, Liang S, Shum CK. Cyanobacteria blooms and non-alcoholic liver disease: evidence from a county level ecological study in the United States. *Environ Health* 2015 May 07;14:41 [FREE Full text] [doi: [10.1186/s12940-015-0026-7](https://doi.org/10.1186/s12940-015-0026-7)] [Medline: [25948281](#)]
156. Zhang Z, Bergquist R, Chen D, Yao B, Wang Z, Gao J, et al. Identification of parasite-host habitats in Anxiang county, Hunan Province, China based on multi-temporal China-Brazil earth resources satellite (CBERS) images. *PLoS One* 2013;8(7):e69447 [FREE Full text] [doi: [10.1371/journal.pone.0069447](https://doi.org/10.1371/journal.pone.0069447)] [Medline: [23922712](#)]
157. Zhang Z, Ward M, Gao J, Wang Z, Yao B, Zhang T, et al. Remote sensing and disease control in China: past, present and future. *Parasit Vectors* 2013 Jan 11;6:11 [FREE Full text] [doi: [10.1186/1756-3305-6-11](https://doi.org/10.1186/1756-3305-6-11)] [Medline: [23311958](#)]
158. Zhu H, Liu L, Zhou XN, Yang GJ. Ecological Model to Predict Potential Habitats of *Oncomelania hupensis*, the Intermediate Host of *Schistosoma japonicum* in the Mountainous Regions, China. *PLoS Negl Trop Dis* 2015;9(8):e0004028 [FREE Full text] [doi: [10.1371/journal.pntd.0004028](https://doi.org/10.1371/journal.pntd.0004028)] [Medline: [26305881](#)]
159. Zou L, Miller SN, Schmidtmann ET. Mosquito larval habitat mapping using remote sensing and GIS: implications of coalbed methane development and West Nile virus. *J Med Entomol* 2006 Sep;43(5):1034-1041. [Medline: [17017244](#)]
160. Wang Z, Liu Y, Hu M, Pan X, Shi J, Chen F, et al. Acute health impacts of airborne particles estimated from satellite remote sensing. *Environ Int* 2013 Jan;51:150-159 [FREE Full text] [doi: [10.1016/j.envint.2012.10.011](https://doi.org/10.1016/j.envint.2012.10.011)] [Medline: [23220016](#)]
161. Higgs G, Sterling DA, Aryal S, Vemulapalli A, Priftis KN, Sifakis NI. Aerosol optical depth as a measure of particulate exposure using imputed censored data, and relationship with childhood asthma hospital admissions for 2004 in Athens, Greece. *Environ Health Insights* 2015;9(Suppl 1):27-33 [FREE Full text] [doi: [10.4137/EHL.S15665](https://doi.org/10.4137/EHL.S15665)] [Medline: [25987842](#)]
162. Andrade FV, Artaxo P, Hacon S, Carmo CND, Cirino G. Aerosols from biomass burning and respiratory diseases in children, Manaus, Northern Brazil. *Rev Saude Publica* 2013 Apr;47(2):239-247 [FREE Full text] [doi: [10.1590/S0034-8910.2013047004011](https://doi.org/10.1590/S0034-8910.2013047004011)] [Medline: [24037350](#)]
163. Shi W, Wong MS, Wang J, Zhao Y. Analysis of airborne particulate matter (PM_{2.5}) over Hong Kong using remote sensing and GIS. *Sensors (Basel)* 2012;12(6):6825-6836 [FREE Full text] [doi: [10.3390/s120606825](https://doi.org/10.3390/s120606825)] [Medline: [22969323](#)]
164. Lee H, Kang CM, Coull BA, Bell ML, Koutrakis P. Assessment of primary and secondary ambient particle trends using satellite aerosol optical depth and ground speciation data in the New England region, United States. *Environ Res* 2014 Aug;133:103-110 [FREE Full text] [doi: [10.1016/j.envres.2014.04.006](https://doi.org/10.1016/j.envres.2014.04.006)] [Medline: [24906074](#)]
165. McGuinn L, Ward-Caviness CK, Neas LM, Schneider A, Diaz-Sanchez D, Cascio WE, et al. Association between satellite-based estimates of long-term PM_{2.5} exposure and coronary artery disease. *Environ Res* 2016 Feb;145:9-17 [FREE Full text] [doi: [10.1016/j.envres.2015.10.026](https://doi.org/10.1016/j.envres.2015.10.026)] [Medline: [26613345](#)]
166. Jerrett M, Turner MC, Beckerman BS, Pope CA, van DA, Martin RV, et al. Comparing the Health Effects of Ambient Particulate Matter Estimated Using Ground-Based versus Remote Sensing Exposure Estimates. *Environ Health Perspect* 2017 Dec;125(4):552-559 [FREE Full text] [doi: [10.1289/EHP575](https://doi.org/10.1289/EHP575)] [Medline: [27611476](#)]
167. Lee SJ, Serre ML, van DA, Martin RV, Burnett RT, Jerrett M. Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM_{2.5} concentrations across the continental United States. *Environ Health Perspect* 2012 Dec;120(12):1727-1732 [FREE Full text] [doi: [10.1289/ehp.1205006](https://doi.org/10.1289/ehp.1205006)] [Medline: [23033456](#)]
168. Yao J, Henderson SB. An empirical model to estimate daily forest fire smoke exposure over a large geographic area using air quality, meteorological, and remote sensing data. *J Expo Sci Environ Epidemiol* 2014;24(3):328-335 [FREE Full text] [doi: [10.1038/jes.2013.87](https://doi.org/10.1038/jes.2013.87)] [Medline: [24301352](#)]
169. Al-Hamdan MZ, Crosson WL, Economou SA, Estes MG, Estes SM, Hemmings SN, et al. Environmental Public Health Applications Using Remotely Sensed Data. *Geocarto Int* 2014 Jan 01;29(1):85-98 [FREE Full text] [doi: [10.1080/10106049.2012.715209](https://doi.org/10.1080/10106049.2012.715209)] [Medline: [24910505](#)]
170. Liu Y, Sarnat JA, Kilaru V, Jacob DJ, Koutrakis P. Estimating ground-level PM_{2.5} in the eastern United States using satellite remote sensing. *Environ Sci Technol* 2005 May 01;39(9):3269-3278. [Medline: [15926578](#)]
171. Sotoudeheian S, Arhami M. Estimating ground-level PM₁₀ using satellite remote sensing and ground-based meteorological measurements over Tehran. *J Environ Health Sci Eng* 2014;12(1):122 [FREE Full text] [doi: [10.1186/s40201-014-0122-6](https://doi.org/10.1186/s40201-014-0122-6)] [Medline: [25343043](#)]
172. van Donkelaar A, Martin RV, Park RJ. Estimating ground-level PM using aerosol optical depth determined from satellite remote sensing. *J. Geophys. Res* 2006 Nov 02;111(D21):e. [doi: [10.1029/2005JD006996](https://doi.org/10.1029/2005JD006996)]
173. Song Y, Yang HL, Peng JH, Song YR, Sun Q, Li Y. Estimating PM_{2.5} Concentrations in Xi'an City Using a Generalized Additive Model with Multi-Source Monitoring Data. *PLoS One* 2015;10(11):e0142149 [FREE Full text] [doi: [10.1371/journal.pone.0142149](https://doi.org/10.1371/journal.pone.0142149)] [Medline: [26540446](#)]
174. Liu Y, Paciorek CJ, Koutrakis P. Estimating regional spatial and temporal variability of PM(2.5) concentrations using satellite data, meteorology, and land use information. *Environ Health Perspect* 2009 Jun;117(6):886-892 [FREE Full text] [doi: [10.1289/ehp.0800123](https://doi.org/10.1289/ehp.0800123)] [Medline: [19590678](#)]

175. Lary D, Faruque FS, Malakar N, Moore A, Roscoe B, Adams ZL, et al. Estimating the global abundance of ground level presence of particulate matter (PM_{2.5}). *Geospat Health* 2014 Dec 01;8(3):S611-S630 [[FREE Full text](#)] [doi: [10.4081/gh.2014.292](#)] [Medline: [25599634](#)]
176. Yao J, Brauer M, Henderson SB. Evaluation of a wildfire smoke forecasting system as a tool for public health protection. *Environ Health Perspect* 2013 Oct;121(10):1142-1147 [[FREE Full text](#)] [doi: [10.1289/ehp.1306768](#)] [Medline: [23906969](#)]
177. Brauer M, Amann M, Burnett RT, Cohen A, Dentener F, Ezzati M, et al. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environ Sci Technol* 2012 Jan 17;46(2):652-660 [[FREE Full text](#)] [doi: [10.1021/es2025752](#)] [Medline: [22148428](#)]
178. Xu L, Yin H, Xie XD. Health risk assessment of inhalable particulate matter in Beijing based on the thermal environment. *Int J Environ Res Public Health* 2014 Nov 28;11(12):12368-12388 [[FREE Full text](#)] [doi: [10.3390/ijerph111212368](#)] [Medline: [25464132](#)]
179. Kloog I, Ridgway B, Koutrakis P, Coull BA, Schwartz JD. Long- and short-term exposure to PM_{2.5} and mortality: using novel exposure models. *Epidemiology* 2013 Jul;24(4):555-561 [[FREE Full text](#)] [doi: [10.1097/EDE.0b013e318294beaa](#)] [Medline: [23676266](#)]
180. Geddes JA, Martin RV, Boys BL, van Donkelaar A. Long-Term Trends Worldwide in Ambient NO₂ Concentrations Inferred from Satellite Observations. *Environ Health Perspect* 2016 Mar;124(3):281-289 [[FREE Full text](#)] [doi: [10.1289/ehp.1409567](#)] [Medline: [26241114](#)]
181. Coker E, Ghosh J, Jerrett M, Gomez-Rubio V, Beckerman B, Cockburn M, et al. Modeling spatial effects of PM_{2.5} on term low birth weight in Los Angeles County. *Environ Res* 2015 Oct;142:354-364 [[FREE Full text](#)] [doi: [10.1016/j.envres.2015.06.044](#)] [Medline: [26196780](#)]
182. Olaguer EP. Overview of the Benzene and Other Toxics Exposure (BEE-TEX) Field Study. *Environ Health Insights* 2015;9(Suppl 4):1-6 [[FREE Full text](#)] [doi: [10.4137/EHI.S15654](#)] [Medline: [26549972](#)]
183. Yao L, Lu N. Particulate matter pollution and population exposure assessment over mainland China in 2010 with remote sensing. *Int J Environ Res Public Health* 2014 May 14;11(5):5241-5250 [[FREE Full text](#)] [doi: [10.3390/ijerph110505241](#)] [Medline: [24830453](#)]
184. Mackie S, Watson M. Probabilistic detection of volcanic ash using a Bayesian approach. *J Geophys Res Atmos* 2014 Mar 16;119(5):2409-2428 [[FREE Full text](#)] [doi: [10.1002/2013JD021077](#)] [Medline: [25844278](#)]
185. Rashidi M, Ramesht MH, Zohary M, Poursafa P, Kelishadi R, Rashidi Z, et al. Relation of air pollution with epidemiology of respiratory diseases in isfahan, Iran from 2005 to 2009. *J Res Med Sci* 2013 Dec;18(12):1074-1079 [[FREE Full text](#)] [Medline: [24523799](#)]
186. Hoff R, Christopher SA. Remote sensing of particulate pollution from space: have we reached the promised land? *J Air Waste Manag Assoc* 2009 Jun;59(6):645-75; discussion 642. [Medline: [19603734](#)]
187. Gupta P, Christopher S, Wang J, Gehrig R, Lee Y, Kumar N. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment* 2006 Sep;40(30):5880-5892. [doi: [10.1016/j.atmosenv.2006.03.016](#)]
188. Anderson HR, Butland BK, van Donkelaar DA, Brauer M, Strachan DP, Clayton T, et al. Satellite-based estimates of ambient air pollution and global variations in childhood asthma prevalence. *Environ Health Perspect* 2012 Sep;120(9):1333-1339 [[FREE Full text](#)] [doi: [10.1289/ehp.1104724](#)] [Medline: [22548921](#)]
189. Lin G, Fu J, Jiang D, Wang J, Wang Q, Dong D. Spatial Variation of the Relationship between PM_{2.5} Concentrations and Meteorological Parameters in China. *Biomed Res Int* 2015;2015:684618 [[FREE Full text](#)] [doi: [10.1155/2015/684618](#)] [Medline: [26295046](#)]
190. Liu JC, Pereira G, Uhl SA, Bravo MA, Bell ML. A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environ Res* 2015 Jan;136:120-132 [[FREE Full text](#)] [doi: [10.1016/j.envres.2014.10.015](#)] [Medline: [25460628](#)]
191. Elliott CT, Henderson SB, Wan V. Time series analysis of fine particulate matter and asthma reliever dispensations in populations affected by forest fires. *Environ Health* 2013 Jan 28;12:11 [[FREE Full text](#)] [doi: [10.1186/1476-069X-12-11](#)] [Medline: [23356966](#)]
192. van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ Health Perspect* 2015 Feb;123(2):135-143 [[FREE Full text](#)] [doi: [10.1289/ehp.1408646](#)] [Medline: [25343779](#)]
193. Lee HJ, Coull BA, Bell ML, Koutrakis P. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM_{2.5} concentrations. *Environ Res* 2012 Oct;118:8-15 [[FREE Full text](#)] [doi: [10.1016/j.envres.2012.06.011](#)] [Medline: [22841416](#)]
194. Lary DJ, Lary T, Sattler B. Using Machine Learning to Estimate Global PM_{2.5} for Environmental Health Studies. *Environ Health Insights* 2015;9(Suppl 1):41-52 [[FREE Full text](#)] [doi: [10.4137/EHI.S15664](#)] [Medline: [26005352](#)]
195. Kloog I, Melly SJ, Ridgway WL, Coull BA, Schwartz J. Using new satellite based exposure methods to study the association between pregnancy PM_{2.5} exposure, premature birth and birth weight in Massachusetts. *Environ Health* 2012 Jun 18;11:40 [[FREE Full text](#)] [doi: [10.1186/1476-069X-11-40](#)] [Medline: [22709681](#)]
196. Konkel L. The view from afar: satellite-derived estimates of global PM_{2.5}. *Environ Health Perspect* 2015 Feb;123(2):A43 [[FREE Full text](#)] [doi: [10.1289/ehp.123-A43](#)] [Medline: [25643426](#)]

197. Witheetrirong Y, Tripathi NK, Tipdecho T, Parkpian P. Estimation of the effect of soil texture on nitrate-nitrogen content in groundwater using optical remote sensing. *Int J Environ Res Public Health* 2011 Dec;8(8):3416-3436 [[FREE Full text](#)] [doi: [10.3390/ijerph8083416](#)] [Medline: [21909315](#)]
198. Bandaru V, Daughtry CS, Codling EE, Hansen DJ, White-Hansen S, Green CE. Evaluating Leaf and Canopy Reflectance of Stressed Rice Plants to Monitor Arsenic Contamination. *Int J Environ Res Public Health* 2016 Dec 18;13(6):pii:-E606 [[FREE Full text](#)] [doi: [10.3390/ijerph13060606](#)] [Medline: [27322304](#)]
199. Kamińska IA, Oldak A, Turski WA. Geographical Information System (GIS) as a tool for monitoring and analysing pesticide pollution and its impact on public health. *Ann Agric Environ Med* 2004;11(2):181-184 [[FREE Full text](#)] [Medline: [15627322](#)]
200. Ward MH, Nuckols JR, Weigel SJ, Maxwell SK, Cantor KP, Miller RS. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System. *Environ Health Perspect* 2000 Jan;108(1):5-12 [[FREE Full text](#)] [Medline: [10622770](#)]
201. Yang K, Zhou X, Yan W, Hang D, Steinmann P. Landfills in Jiangsu province, China, and potential threats for public health: leachate appraisal and spatial analysis using geographic information system and remote sensing. *Waste Manag* 2008 Dec;28(12):2750-2757. [doi: [10.1016/j.wasman.2008.01.021](#)] [Medline: [18396395](#)]
202. Frassy F, Candiani G, Rusmini M, Maiani P, Marchesi A, Rota NF, et al. Mapping asbestos-cement roofing with hyperspectral remote sensing over a large mountain region of the Italian Western Alps. *Sensors (Basel)* 2014 Aug 27;14(9):15900-15913 [[FREE Full text](#)] [doi: [10.3390/s140915900](#)] [Medline: [25166502](#)]
203. Chen Y, Liu Y, Liu Y, Lin A, Kong X, Liu D, et al. Mapping of Cu and Pb contaminations in soil using combined geochemistry, topography, and remote sensing: a case study in the Le'an River floodplain, China. *Int J Environ Res Public Health* 2012 Dec;9(5):1874-1886 [[FREE Full text](#)] [doi: [10.3390/ijerph9051874](#)] [Medline: [22754479](#)]
204. Maxwell SK, Airola M, Nuckols JR. Using Landsat satellite data to support pesticide exposure assessment in California. *Int J Health Geogr* 2010 Sep 16;9:46 [[FREE Full text](#)] [doi: [10.1186/1476-072X-9-46](#)] [Medline: [20846438](#)]
205. Xu Z, Liu Y, Ma Z, Sam TG, Hu W, Tong S. Assessment of the temperature effect on childhood diarrhea using satellite imagery. *Sci Rep* 2014 Jun 23;4:5389 [[FREE Full text](#)] [doi: [10.1038/srep05389](#)] [Medline: [24953087](#)]
206. Pereira G, Foster S, Martin K, Christian H, Boruff BJ, Knuiman M, et al. The association between neighborhood greenness and cardiovascular disease: an observational study. *BMC Public Health* 2012 Jun 21;12:466 [[FREE Full text](#)] [doi: [10.1186/1471-2458-12-466](#)] [Medline: [22720780](#)]
207. Bennie J, Davies TW, Duffy JP, Inger R, Gaston KJ. Contrasting trends in light pollution across Europe based on satellite observed night time lights. *Sci Rep* 2014 Jan 21;4:3789 [[FREE Full text](#)] [doi: [10.1038/srep03789](#)] [Medline: [24445659](#)]
208. Smith LT, Aragão LEOC, Sabel CE, Nakaya T. Drought impacts on children's respiratory health in the Brazilian Amazon. *Sci Rep* 2014 Jan 16;4:3726 [[FREE Full text](#)] [doi: [10.1038/srep03726](#)] [Medline: [24430803](#)]
209. Hjort J, Hugg TT, Antikainen H, Rusanen J, Sofiev M, Kukkonen J, et al. Fine-Scale Exposure to Allergenic Pollen in the Urban Environment: Evaluation of Land Use Regression Approach. *Environ Health Perspect* 2016 Dec;124(5):619-626 [[FREE Full text](#)] [doi: [10.1289/ehp.1509761](#)] [Medline: [26452296](#)]
210. Liss A, Koch M, Naumova EN. Redefining climate regions in the United States of America using satellite remote sensing and machine learning for public health applications. *Geospat Health* 2014 Dec 01;8(3):S647-S659 [[FREE Full text](#)] [doi: [10.4081/gh.2014.294](#)] [Medline: [25599636](#)]
211. Johnson D, Vijay L, Stanforth A, Webber J. Remote Sensing of Heat-Related Health Risks: The Trend Toward Coupling Socioeconomic and Remotely Sensed Data. *Geography Compass* 2011;5(10):767-780. [doi: [10.1111/j.1749-8198.2011.00442](#)]
212. Hystad P, Davies HW, Frank L, Van Loon J, Gehring U, Tamburic L, et al. Residential greenness and birth outcomes: evaluating the influence of spatially correlated built-environment factors. *Environ Health Perspect* 2014 Oct;122(10):1095-1102 [[FREE Full text](#)] [doi: [10.1289/ehp.1308049](#)] [Medline: [25014041](#)]
213. Johnson D, Wilson JS, Lubert GC. Socioeconomic indicators of heat-related health risk supplemented with remotely sensed data. *Int J Health Geogr* 2009 Oct 16;8:57 [[FREE Full text](#)] [doi: [10.1186/1476-072X-8-57](#)] [Medline: [19835578](#)]
214. Song D, Jiang D, Wang Y, Chen W, Huang Y, Zhuang D. Study on association between spatial distribution of metal mines and disease mortality: a case study in Suxian District, South China. *Int J Environ Res Public Health* 2013 Oct 16;10(10):5163-5177 [[FREE Full text](#)] [doi: [10.3390/ijerph10105163](#)] [Medline: [24135822](#)]
215. Morabito M, Crisci A, Gioli B, Gualtieri G, Toscano P, Di SV, et al. Urban-hazard risk analysis: mapping of heat-related risks in the elderly in major Italian cities. *PLoS One* 2015;10(5):e0127277 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0127277](#)] [Medline: [25985204](#)]
216. Brown ME, Grace K, Shively G, Johnson KB, Carroll M. Using satellite remote sensing and household survey data to assess human health and nutrition response to environmental change. *Popul Environ* 2014;36:48-72 [[FREE Full text](#)] [doi: [10.1007/s11111-013-0201-0](#)] [Medline: [25132700](#)]
217. Hay SI. An overview of remote sensing and geodesy for epidemiology and public health application. *Advances in Parasitology* 2000;47:1-35. [doi: [10.1016/S0065-308X\(00\)47005-3](#)]
218. Igarashi T, Kuze A, Sobue S, Yamamoto A, Yamamoto K, Oyoshi K, et al. Japan's efforts to promote global health using satellite remote sensing data from the Japan Aerospace Exploration Agency for prediction of infectious diseases and air quality. *Geospat Health* 2014 Dec 01;8(3):S603-S610 [[FREE Full text](#)] [doi: [10.4081/gh.2014.299](#)] [Medline: [25599641](#)]

219. Jovanović P. Satellite remote sensing imagery in public health. *Acta Astronautica* 1987 Nov;15(11):951-953. [doi: [10.1016/0094-5765\(87\)90056-7](https://doi.org/10.1016/0094-5765(87)90056-7)]
220. Boulos MN, Roudsari AV, Carson ER. Health geomatics: an enabling suite of technologies in health and healthcare. *J Biomed Inform* 2001 Jun;34(3):195-219 [FREE Full text] [Medline: [11723701](https://pubmed.ncbi.nlm.nih.gov/11723701/)]
221. Kistemann T, Dangendorf F, Schweikart J. New perspectives on the use of Geographical Information Systems (GIS) in environmental health sciences. *Int J Hyg Environ Health* 2002 Apr;205(3):169-181. [doi: [10.1078/1438-4639-00145](https://doi.org/10.1078/1438-4639-00145)] [Medline: [12040915](https://pubmed.ncbi.nlm.nih.gov/12040915/)]
222. Li Z, Xu D, Guo X. Remote sensing of ecosystem health: opportunities, challenges, and future perspectives. *Sensors (Basel)* 2014 Nov 07;14(11):21117-21139 [FREE Full text] [doi: [10.3390/s141121117](https://doi.org/10.3390/s141121117)] [Medline: [25386759](https://pubmed.ncbi.nlm.nih.gov/25386759/)]
223. Patz J. Satellite remote sensing can improve chances of achieving sustainable health. *Environ Health Perspect* 2005 Feb;113(2):A84-A85 [FREE Full text] [Medline: [15687039](https://pubmed.ncbi.nlm.nih.gov/15687039/)]
224. Robinson T. Spatial statistics and geographical information systems in epidemiology and public health. *Adv Parasitol* 2000;47:81-128. [Medline: [10997205](https://pubmed.ncbi.nlm.nih.gov/10997205/)]
225. Seltenrich N. Remote-sensing applications for environmental health research. *Environ Health Perspect* 2014 Oct;122(10):A268-A275 [FREE Full text] [doi: [10.1289/ehp.122-A268](https://doi.org/10.1289/ehp.122-A268)] [Medline: [25272250](https://pubmed.ncbi.nlm.nih.gov/25272250/)]
226. Sprigg W. Public-health applications in remote sensing. *SPIE Newsroom* 2009;1-3. [doi: [10.1117/2.1200902.1488](https://doi.org/10.1117/2.1200902.1488)]
227. Wood B, Beck LR, Lobitz BM, Bobo MR. Education, outreach and the future of remote sensing in human health. *Adv Parasitol* 2000;47:331-344. [Medline: [10997212](https://pubmed.ncbi.nlm.nih.gov/10997212/)]
228. What is the difference between GPS and GNSS?. 2013 Sep. Retrieved November 20, 2016, from URL: <http://kb.unavco.org/kb/article/what-is-the-difference-between-gps-and-gnss-167.html> [accessed 2017-11-19] [WebCite Cache ID 6v643Jx97]
229. Rob. How GNSS Works. (October 10, 2011) URL: http://gnss.be/how_tutorial [accessed 2017-11-19] [WebCite Cache ID 6v647RaLX]
230. Almanza E, Jerrett M, Dunton G, Seto E, Pentz MA. A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. *Health Place* 2012 Jan;18(1):46-54 [FREE Full text] [doi: [10.1016/j.healthplace.2011.09.003](https://doi.org/10.1016/j.healthplace.2011.09.003)] [Medline: [22243906](https://pubmed.ncbi.nlm.nih.gov/22243906/)]
231. Bürgi R, Tomatis L, Murer K, de Bruin ED. Localization of Physical Activity in Primary School Children Using Accelerometry and Global Positioning System. *PLoS One* 2015;10(11):e0142223 [FREE Full text] [doi: [10.1371/journal.pone.0142223](https://doi.org/10.1371/journal.pone.0142223)] [Medline: [26536595](https://pubmed.ncbi.nlm.nih.gov/26536595/)]
232. Carvalho LFR, de Melo CB, McManus C, Haddad JPA. Use of satellite images for geographical localization of livestock holdings in Brazil. *Prev Vet Med* 2012 Jan 01;103(1):74-77 [FREE Full text] [doi: [10.1016/j.prevetmed.2011.08.006](https://doi.org/10.1016/j.prevetmed.2011.08.006)] [Medline: [21917345](https://pubmed.ncbi.nlm.nih.gov/21917345/)]
233. Chaix B, Kestens Y, Duncan S, Merrien C, Thierry B, Pannier B, et al. Active transportation and public transportation use to achieve physical activity recommendations? A combined GPS, accelerometer, and mobility survey study. *Int J Behav Nutr Phys Act* 2014 Sep 27;11:124 [FREE Full text] [doi: [10.1186/s12966-014-0124-x](https://doi.org/10.1186/s12966-014-0124-x)] [Medline: [25260793](https://pubmed.ncbi.nlm.nih.gov/25260793/)]
234. Edwards N, Hooper P, Knuijan M, Foster S, Giles-Corti B. Associations between park features and adolescent park use for physical activity. *Int J Behav Nutr Phys Act* 2015 Feb 18;12:21 [FREE Full text] [doi: [10.1186/s12966-015-0178-4](https://doi.org/10.1186/s12966-015-0178-4)] [Medline: [25879200](https://pubmed.ncbi.nlm.nih.gov/25879200/)]
235. James P, Banay RF, Hart JE, Laden F. A Review of the Health Benefits of Greenness. *Curr Epidemiol Rep* 2015 Jun;2(2):131-142 [FREE Full text] [doi: [10.1007/s40471-015-0043-7](https://doi.org/10.1007/s40471-015-0043-7)] [Medline: [26185745](https://pubmed.ncbi.nlm.nih.gov/26185745/)]
236. Larson K. A new way to detect volcanic plumes. *Geophys. Res. Lett* 2013 Jun 13;40(11):2657-2660. [doi: [10.1002/grl.50556](https://doi.org/10.1002/grl.50556)]
237. Li R, Kling SR, Salata MJ, Cupp SA, Sheehan J, Voos JE. Wearable Performance Devices in Sports Medicine. *Sports Health* 2016;8(1):74-78 [FREE Full text] [doi: [10.1177/1941738115616917](https://doi.org/10.1177/1941738115616917)] [Medline: [26733594](https://pubmed.ncbi.nlm.nih.gov/26733594/)]
238. MacKerron G, Mourato S. Happiness is greater in natural environments. *Global Environmental Change* 2013 Oct;23(5):992-1000. [doi: [10.1016/j.gloenvcha.2013.03.010](https://doi.org/10.1016/j.gloenvcha.2013.03.010)]
239. Marani R, Gelao G, Perri AG. High quality heart and lung auscultation system for diagnostic use on remote patients in real time. *Open Biomed Eng J* 2010 Nov 03;4:250-256 [FREE Full text] [doi: [10.2174/1874120701004010250](https://doi.org/10.2174/1874120701004010250)] [Medline: [21379392](https://pubmed.ncbi.nlm.nih.gov/21379392/)]
240. McCrorie PRW, Fenton C, Ellaway A. Combining GPS, GIS, and accelerometry to explore the physical activity and environment relationship in children and young people - a review. *Int J Behav Nutr Phys Act* 2014 Sep 13;11:93 [FREE Full text] [doi: [10.1186/s12966-014-0093-0](https://doi.org/10.1186/s12966-014-0093-0)] [Medline: [25356782](https://pubmed.ncbi.nlm.nih.gov/25356782/)]
241. Babu PM, Sankar GJ, Sreenivasulu V, Harikrishna DK. Hydrochemical Analysis and Evaluation of Groundwater Quality in Part of Krishna District, Andhra Pradesh – Using Remotesensing and GIS Techniques. *IJER* 2014 Aug 1;3(8):476-481. [doi: [10.17950/ijer/v3s8/801](https://doi.org/10.17950/ijer/v3s8/801)]
242. O'Connor TM, Cerin E, Lee RE, Parker N, Chen TA, Hughes SO, et al. Environmental and cultural correlates of physical activity parenting practices among Latino parents with preschool-aged children: Niños Activos. *BMC Public Health* 2014 Jul 10;14:707 [FREE Full text] [doi: [10.1186/1471-2458-14-707](https://doi.org/10.1186/1471-2458-14-707)] [Medline: [25011669](https://pubmed.ncbi.nlm.nih.gov/25011669/)]
243. Oliver M, Badland H, Mavoa S, Duncan MJ, Duncan S. Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. *J Phys Act Health* 2010 Jan;7(1):102-108. [Medline: [20231761](https://pubmed.ncbi.nlm.nih.gov/20231761/)]

244. Oreskovic NM, Perrin JM, Robinson AI, Locascio JJ, Blossom J, Chen ML, et al. Adolescents' use of the built environment for physical activity. *BMC Public Health* 2015 Mar 15;15:251 [FREE Full text] [doi: [10.1186/s12889-015-1596-6](https://doi.org/10.1186/s12889-015-1596-6)] [Medline: [25880654](https://pubmed.ncbi.nlm.nih.gov/25880654/)]
245. Ramesh MNR, Sathish CMR, Undi M, Aravind M, Puthusseri YP. Global positioning system: a new tool to measure the distribution of anemia and nutritional status of children (5-10 years) in a rural area in south India. *Indian J Med Sci* 2012;66(1-2):13-22. [doi: [10.4103/0019-5359.110854](https://doi.org/10.4103/0019-5359.110854)] [Medline: [23603568](https://pubmed.ncbi.nlm.nih.gov/23603568/)]
246. Roelfsema C, Phinn S, Dennison W, Dekker A, Brando V. Monitoring toxic cyanobacteria *Lyngbya majuscula* (Gomont) in Moreton Bay, Australia by integrating satellite image data and field mapping. *Harmful Algae* 2006 Jan;5(1):45-56. [doi: [10.1016/j.hal.2005.05.001](https://doi.org/10.1016/j.hal.2005.05.001)]
247. Townshend AD, Worringham CJ, Stewart IB. Assessment of speed and position during human locomotion using nondifferential GPS. *Med Sci Sports Exerc* 2008 Jan;40(1):124-132. [doi: [10.1249/mss.0b013e3181590bc2](https://doi.org/10.1249/mss.0b013e3181590bc2)] [Medline: [18091013](https://pubmed.ncbi.nlm.nih.gov/18091013/)]
248. Tymms P, Curtis SE, Routen AC, Thomson KH, Bolden DS, Bock S, et al. Clustered randomised controlled trial of two education interventions designed to increase physical activity and well-being of secondary school students: the MOVE Project. *BMJ Open* 2016 Jan 06;6(1):e009318 [FREE Full text] [doi: [10.1136/bmjopen-2015-009318](https://doi.org/10.1136/bmjopen-2015-009318)] [Medline: [26739729](https://pubmed.ncbi.nlm.nih.gov/26739729/)]
249. Walsh S, McCleary AL, Heumann BW, Brewington L, Raczkowski EJ, Mena CF. Community Expansion and Infrastructure Development: Implications for Human Health and Environmental Quality in the Galápagos Islands of Ecuador. *Journal of Latin American Geography* 2010;9(3):137-159.
250. Wieters KM, Kim J, Lee C. Assessment of wearable global positioning system units for physical activity research. *J Phys Act Health* 2012 Sep;9(7):913-923. [Medline: [21975729](https://pubmed.ncbi.nlm.nih.gov/21975729/)]
251. Wu J, Jiang C, Liu Z, Houston D, Jaimes G, McConnell R. Performances of different global positioning system devices for time-location tracking in air pollution epidemiological studies. *Environ Health Insights* 2010 Nov 23;4:93-108 [FREE Full text] [doi: [10.4137/EHI.S6246](https://doi.org/10.4137/EHI.S6246)] [Medline: [21151593](https://pubmed.ncbi.nlm.nih.gov/21151593/)]
252. Achee NL, Grieco JP, Masuoka P, Andre RG, Roberts DR, Thomas J, et al. Use of remote sensing and geographic information systems to predict locations of *Anopheles darlingi*-positive breeding sites within the Sibun River in Belize, Central America. *J Med Entomol* 2006 Mar;43(2):382-392. [Medline: [16619625](https://pubmed.ncbi.nlm.nih.gov/16619625/)]
253. Barau I, Zubairu M, Mwanza MN, Seaman VY. Improving polio vaccination coverage in Nigeria through the use of geographic information system technology. *J Infect Dis* 2014 Nov 01;210 Suppl 1:S102-S110. [doi: [10.1093/infdis/jiu010](https://doi.org/10.1093/infdis/jiu010)] [Medline: [25316823](https://pubmed.ncbi.nlm.nih.gov/25316823/)]
254. Cui P, Hou Y, Tang M, Zhang H, Zhou Y, Yin Z, et al. Movement patterns of Bar-headed Geese *Anser indicus* during breeding and post-breeding periods at Qinghai Lake, China. *J Ornithol* 2010 Jul 8;152(1):83-92. [doi: [10.1007/s10336-010-0552-6](https://doi.org/10.1007/s10336-010-0552-6)]
255. De Souza Dias M, Dias GH, Nobre ML. The use of Geographical Information System (GIS) to improve active leprosy case finding campaigns in the municipality of Mossoró, Rio Grande do Norte State, Brazil. *Lepr Rev* 2007 Sep;78(3):261-269. [Medline: [18035777](https://pubmed.ncbi.nlm.nih.gov/18035777/)]
256. Hightower AW, Ombok M, Otieno R, Odhiambo R, Oloo AJ, Lal AA, et al. A geographic information system applied to a malaria field study in western Kenya. *Am J Trop Med Hyg* 1998 Mar;58(3):266-272. [Medline: [9546401](https://pubmed.ncbi.nlm.nih.gov/9546401/)]
257. Kaewwaen W, Bhumiratana A. Landscape ecology and epidemiology of malaria associated with rubber plantations in Thailand: integrated approaches to malaria ecotoping. *Interdiscip Perspect Infect Dis* 2015;2015:909106 [FREE Full text] [doi: [10.1155/2015/909106](https://doi.org/10.1155/2015/909106)] [Medline: [25838822](https://pubmed.ncbi.nlm.nih.gov/25838822/)]
258. Krech T. TBE foci in Switzerland. *Int J Med Microbiol* 2002 Jun;291 Suppl 33:30-33. [Medline: [12141754](https://pubmed.ncbi.nlm.nih.gov/12141754/)]
259. Maynard NG, Conway GA. A view from above: use of satellite imagery to enhance our understanding of potential impacts of climate change on human health in the Arctic. *Alaska Med* 2007;49(3):78-85. [Medline: [18062618](https://pubmed.ncbi.nlm.nih.gov/18062618/)]
260. Newman S, Hill NJ, Spragens KA, Janies D, Voronkin IO, Prosser DJ, et al. Eco-virological approach for assessing the role of wild birds in the spread of avian influenza H5N1 along the Central Asian Flyway. *PLoS One* 2012;7(2):e30636 [FREE Full text] [doi: [10.1371/journal.pone.0030636](https://doi.org/10.1371/journal.pone.0030636)] [Medline: [22347393](https://pubmed.ncbi.nlm.nih.gov/22347393/)]
261. Prosser D, Takekawa JY, Newman SH, Yan B, Douglas DC, Hou Y, et al. Satellite-marked waterfowl reveal migratory connection between H5N1 outbreak areas in China and Mongolia. *Ibis* 2009;151(3):568-576. [doi: [10.1111/j.1474-919X.2009.00932.x](https://doi.org/10.1111/j.1474-919X.2009.00932.x)]
262. Shin J, Lee K, Kim S, Hwang J, Woo C, Kim J, et al. Tracking Mallards (*Anas platyrhynchos*) with GPS Satellite Transmitters Along Their Migration Route Through Northeast Asia. *Avian Dis* 2016 May;60(1 Suppl):311-315. [doi: [10.1637/11096-042015-Reg](https://doi.org/10.1637/11096-042015-Reg)] [Medline: [27309072](https://pubmed.ncbi.nlm.nih.gov/27309072/)]
263. Shirayama Y, Phompida S, Shibuya K. Geographic information system (GIS) maps and malaria control monitoring: intervention coverage and health outcome in distal villages of Khammouane province, Laos. *Malar J* 2009 Sep 22;8:217 [FREE Full text] [doi: [10.1186/1475-2875-8-217](https://doi.org/10.1186/1475-2875-8-217)] [Medline: [19772628](https://pubmed.ncbi.nlm.nih.gov/19772628/)]
264. Sithiprasasna R, Ugsang DM, Honda K, Jones JW, Singhasivanon P. Ikonos-derived malaria transmission risk in northwestern Thailand. *Southeast Asian J Trop Med Public Health* 2005 Jan;36(1):14-22. [Medline: [15906637](https://pubmed.ncbi.nlm.nih.gov/15906637/)]

265. Silué KD, Raso G, Yapi A, Vounatsou P, Tanner M, N'goran EK, et al. Spatially-explicit risk profiling of *Plasmodium falciparum* infections at a small scale: a geostatistical modelling approach. *Malar J* 2008 Jun 23;7:111 [FREE Full text] [doi: [10.1186/1475-2875-7-111](https://doi.org/10.1186/1475-2875-7-111)] [Medline: [18570685](https://pubmed.ncbi.nlm.nih.gov/18570685/)]
266. Ali M, Rasool S, Park J, Saeed S, Ochiai RL, Nizami Q, et al. Use of satellite imagery in constructing a household GIS database for health studies in Karachi, Pakistan. *Int J Health Geogr* 2004 Sep 28;3(1):20 [FREE Full text] [doi: [10.1186/1476-072X-3-20](https://doi.org/10.1186/1476-072X-3-20)] [Medline: [15450121](https://pubmed.ncbi.nlm.nih.gov/15450121/)]
267. Bonner M, Han D, Nie J, Rogerson P, Vena JE, Freudenheim JL. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology* 2003 Jul;14(4):408-412. [doi: [10.1097/01.EDE.0000073121.63254.c5](https://doi.org/10.1097/01.EDE.0000073121.63254.c5)] [Medline: [12843763](https://pubmed.ncbi.nlm.nih.gov/12843763/)]
268. Boruff BJ, Nathan A, Nijënstein S. Using GPS technology to (re)-examine operational definitions of 'neighbourhood' in place-based health research. *Int J Health Geogr* 2012 Jun 27;11:22 [FREE Full text] [doi: [10.1186/1476-072X-11-22](https://doi.org/10.1186/1476-072X-11-22)] [Medline: [22738807](https://pubmed.ncbi.nlm.nih.gov/22738807/)]
269. Croner C, Sperling J, Broome FR. Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships. *Stat Med* 1996;15(17-18):1961-1977. [Medline: [8888488](https://pubmed.ncbi.nlm.nih.gov/8888488/)]
270. Elgethun K, Fenske RA, Yost MG, Palcisko GJ. Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. *Environ Health Perspect* 2003 Jan;111(1):115-122 [FREE Full text] [Medline: [12515689](https://pubmed.ncbi.nlm.nih.gov/12515689/)]
271. Graham D, Hipp JA. Emerging technologies to promote and evaluate physical activity: cutting-edge research and future directions. *Front Public Health* 2014;2:66 [FREE Full text] [doi: [10.3389/fpubh.2014.00066](https://doi.org/10.3389/fpubh.2014.00066)] [Medline: [25019066](https://pubmed.ncbi.nlm.nih.gov/25019066/)]
272. Haenssge MJ. Satellite-aided survey sampling and implementation in low- and middle-income contexts: a low-cost/low-tech alternative. *Emerg Themes Epidemiol* 2015;12:20 [FREE Full text] [doi: [10.1186/s12982-015-0041-8](https://doi.org/10.1186/s12982-015-0041-8)] [Medline: [26705402](https://pubmed.ncbi.nlm.nih.gov/26705402/)]
273. Hillson R, Alejandro JD, Jacobsen KH, Ansumana R, Bockarie AS, Bangura U, et al. Methods for determining the uncertainty of population estimates derived from satellite imagery and limited survey data: a case study of Bo city, Sierra Leone. *PLoS One* 2014;9(11):e112241 [FREE Full text] [doi: [10.1371/journal.pone.0112241](https://doi.org/10.1371/journal.pone.0112241)] [Medline: [25398101](https://pubmed.ncbi.nlm.nih.gov/25398101/)]
274. Kerr J, Duncan S, Schipperijn J, Schipperijn J. Using global positioning systems in health research: a practical approach to data collection and processing. *Am J Prev Med* 2011 Nov;41(5):532-540. [doi: [10.1016/j.amepre.2011.07.017](https://doi.org/10.1016/j.amepre.2011.07.017)] [Medline: [22011426](https://pubmed.ncbi.nlm.nih.gov/22011426/)]
275. Kleiner K. They can find you: GPS implants will make it easy to pinpoint people. *New Sci* 2000 Jan 08;165(2220):7. [Medline: [11885636](https://pubmed.ncbi.nlm.nih.gov/11885636/)]
276. Kondo MC, Bream KDW, Barg FK, Branas CC. A random spatial sampling method in a rural developing nation. *BMC Public Health* 2014 Apr 10;14:338 [FREE Full text] [doi: [10.1186/1471-2458-14-338](https://doi.org/10.1186/1471-2458-14-338)] [Medline: [24716473](https://pubmed.ncbi.nlm.nih.gov/24716473/)]
277. Krishnamurthy R, Frolov A, Wolkon A, Vanden EJ, Hightower A. Application of pre-programmed PDA devices equipped with global GPS to conduct paperless household surveys in rural Mozambique. *AMIA Annu Symp Proc* 2006:991 [FREE Full text] [Medline: [17238610](https://pubmed.ncbi.nlm.nih.gov/17238610/)]
278. Lin Y, Kuwayama DP. Using satellite imagery and GPS technology to create random sampling frames in high risk environments. *Int J Surg* 2016 Aug;32:123-128. [doi: [10.1016/j.ijsu.2016.06.044](https://doi.org/10.1016/j.ijsu.2016.06.044)] [Medline: [27392719](https://pubmed.ncbi.nlm.nih.gov/27392719/)]
279. Liu H, Skjetne E, Kobernus M. Mobile phone tracking: in support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment. *Environ Health* 2013 Nov 04;12:93 [FREE Full text] [doi: [10.1186/1476-069X-12-93](https://doi.org/10.1186/1476-069X-12-93)] [Medline: [24188173](https://pubmed.ncbi.nlm.nih.gov/24188173/)]
280. Lowther S, Curriero FC, Shields T, Ahmed S, Monze M, Moss WJ. Feasibility of satellite image-based sampling for a health survey among urban townships of Lusaka, Zambia. *Trop Med Int Health* 2009 Jan;14(1):70-78 [FREE Full text] [doi: [10.1111/j.1365-3156.2008.02185.x](https://doi.org/10.1111/j.1365-3156.2008.02185.x)] [Medline: [19121149](https://pubmed.ncbi.nlm.nih.gov/19121149/)]
281. Noury-Desvaux B, Abraham P, Mahé G, Sauvaet T, Leftheriotis G, Le FA. The accuracy of a simple, low-cost GPS data logger/receiver to study outdoor human walking in view of health and clinical studies. *PLoS One* 2011;6(9):e23027 [FREE Full text] [doi: [10.1371/journal.pone.0023027](https://doi.org/10.1371/journal.pone.0023027)] [Medline: [21931593](https://pubmed.ncbi.nlm.nih.gov/21931593/)]
282. Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect* 2004 Jun;112(9):1007-1015 [FREE Full text] [Medline: [15198921](https://pubmed.ncbi.nlm.nih.gov/15198921/)]
283. Pager D. Impacts for medicine of global monitoring. *Biomed Sci Instrum* 2002;38:283-287. [Medline: [12085617](https://pubmed.ncbi.nlm.nih.gov/12085617/)]
284. Richardson D, Volkow ND, Kwan MP, Kaplan RM, Goodchild MF, Croyle RT. Medicine. Spatial turn in health research. *Science* 2013 Mar 22;339(6126):1390-1392. [doi: [10.1126/science.1232257](https://doi.org/10.1126/science.1232257)] [Medline: [23520099](https://pubmed.ncbi.nlm.nih.gov/23520099/)]
285. Stothard JR, Sousa-Figueiredo JC, Betson M, Seto EYW, Kabatereine NB. Investigating the spatial micro-epidemiology of diseases within a point-prevalence sample: a field applicable method for rapid mapping of households using low-cost GPS-dataloggers. *Trans R Soc Trop Med Hyg* 2011 Sep;105(9):500-506 [FREE Full text] [doi: [10.1016/j.trstmh.2011.05.007](https://doi.org/10.1016/j.trstmh.2011.05.007)] [Medline: [21714979](https://pubmed.ncbi.nlm.nih.gov/21714979/)]
286. Tassetto D, Fazli E, Werner M. A novel hybrid algorithm for passive localization of victims in emergency situations. *Int. J. Satell. Commun. Network* 2010 Oct 17;29(5):461-478. [doi: [10.1002/sat.984](https://doi.org/10.1002/sat.984)]
287. Tatem AJ, Noor AM, Hay SI. Defining approaches to settlement mapping for public health management in Kenya using medium spatial resolution satellite imagery. *Remote Sens Environ* 2004 Oct 30;93(1-2):42-52 [FREE Full text] [doi: [10.1016/j.rse.2004.06.014](https://doi.org/10.1016/j.rse.2004.06.014)] [Medline: [22581984](https://pubmed.ncbi.nlm.nih.gov/22581984/)]

288. Thierry B, Chaix B, Kestens Y. Detecting activity locations from raw GPS data: a novel kernel-based algorithm. *Int J Health Geogr* 2013 Mar 16;12:14 [FREE Full text] [doi: [10.1186/1476-072X-12-14](https://doi.org/10.1186/1476-072X-12-14)] [Medline: [23497213](https://pubmed.ncbi.nlm.nih.gov/23497213/)]
289. Thomson D, Shitole S, Shitole T, Sawant K, Subbaraman R, Bloom DE, et al. A system for household enumeration and re-identification in densely populated slums to facilitate community research, education, and advocacy. *PLoS One* 2014;9(4):e93925 [FREE Full text] [doi: [10.1371/journal.pone.0093925](https://doi.org/10.1371/journal.pone.0093925)] [Medline: [24722369](https://pubmed.ncbi.nlm.nih.gov/24722369/)]
290. Ward M, Nuckols JR, Giglierano J, Bonner MR, Wolter C, Airola M, et al. Positional accuracy of two methods of geocoding. *Epidemiology* 2005 Jul;16(4):542-547. [Medline: [15951673](https://pubmed.ncbi.nlm.nih.gov/15951673/)]
291. Wu J, Jiang C, Jaimes G, Bartell S, Dang A, Baker D, et al. Travel patterns during pregnancy: comparison between Global Positioning System (GPS) tracking and questionnaire data. *Environ Health* 2013 Oct 09;12(1):86 [FREE Full text] [doi: [10.1186/1476-069X-12-86](https://doi.org/10.1186/1476-069X-12-86)] [Medline: [24107241](https://pubmed.ncbi.nlm.nih.gov/24107241/)]
292. Tenbrink T, Wiener JM, Claramunt C, Gallay M, Denis M, Auvray M. Navigation assistance for blind pedestrians: guidelines for the design of devices and implications for spatial cognition. In: *Representing space in cognition : interrelations of behaviour, language, and formal models*. Oxford: Oxford University Press; 2014.
293. Alisky JM. Integrated electronic monitoring systems could revolutionize care for patients with cognitive impairment. *Med Hypotheses* 2006;66(6):1161-1164. [doi: [10.1016/j.mehy.2005.11.045](https://doi.org/10.1016/j.mehy.2005.11.045)] [Medline: [16488550](https://pubmed.ncbi.nlm.nih.gov/16488550/)]
294. Lin C, Chiu MJ, Hsiao CC, Lee RG, Tsai YS. Wireless health care service system for elderly with dementia. *IEEE Trans Inf Technol Biomed* 2006 Oct;10(4):696-704. [Medline: [17044403](https://pubmed.ncbi.nlm.nih.gov/17044403/)]
295. Milne H, van DPM, McCloughan L, Hanley J, Mead G, Starr J, et al. The use of global positional satellite location in dementia: a feasibility study for a randomised controlled trial. *BMC Psychiatry* 2014 May 30;14:160 [FREE Full text] [doi: [10.1186/1471-244X-14-160](https://doi.org/10.1186/1471-244X-14-160)] [Medline: [24885489](https://pubmed.ncbi.nlm.nih.gov/24885489/)]
296. Tenbrink T, Wiener JM, Claramunt C. *Representing space in cognition : interrelations of behaviour, language, and formal models*. Oxford: Oxford University Press; 2014.
297. Varshney U. Pervasive Healthcare and Wireless Health Monitoring. *Mobile Netw Appl* 2007 Jul 12;12(2-3):113-127. [doi: [10.1007/s11036-007-0017-1](https://doi.org/10.1007/s11036-007-0017-1)]
298. Badland H, Duncan MJ, Oliver M, Duncan JS, Mavoa S. Examining commute routes: applications of GIS and GPS technology. *Environ Health Prev Med* 2010 Sep;15(5):327-330 [FREE Full text] [doi: [10.1007/s12199-010-0138-1](https://doi.org/10.1007/s12199-010-0138-1)] [Medline: [21432562](https://pubmed.ncbi.nlm.nih.gov/21432562/)]
299. Chaix B, Méline J, Duncan S, Merrien C, Karusisi N, Perchoux C, et al. GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference? *Health Place* 2013 May;21:46-51. [doi: [10.1016/j.healthplace.2013.01.003](https://doi.org/10.1016/j.healthplace.2013.01.003)] [Medline: [23425661](https://pubmed.ncbi.nlm.nih.gov/23425661/)]
300. Chevalier A, Chevalier AJ, Clarke E, Wall J, Coxon K, Brown J, et al. Naturalistic speeding data: Drivers aged 75 years and older. *Data Brief* 2016 Sep;8:136-141 [FREE Full text] [doi: [10.1016/j.dib.2016.05.016](https://doi.org/10.1016/j.dib.2016.05.016)] [Medline: [27294182](https://pubmed.ncbi.nlm.nih.gov/27294182/)]
301. Guo C, Guo W, Cao G, Dong H. A lane-level LBS system for vehicle network with high-precision BDS/GPS positioning. *Comput Intell Neurosci* 2015;2015:531321 [FREE Full text] [doi: [10.1155/2015/531321](https://doi.org/10.1155/2015/531321)] [Medline: [25755665](https://pubmed.ncbi.nlm.nih.gov/25755665/)]
302. Satyanarayana K, Sarma AD, Sravan J, Malini M, Venkateswarlu G. GPS and GPRS Based Telemonitoring System for Emergency Patient Transportation. *J Med Eng* 2013;2013:363508 [FREE Full text] [doi: [10.1155/2013/363508](https://doi.org/10.1155/2013/363508)] [Medline: [27019844](https://pubmed.ncbi.nlm.nih.gov/27019844/)]
303. de Savigny D, Mayombana C, Mwageni E, Masanja H, Minhaj A, Mkilindi Y, et al. Care-seeking patterns for fatal malaria in Tanzania. *Malar J* 2004 Jul 28;3:27 [FREE Full text] [doi: [10.1186/1475-2875-3-27](https://doi.org/10.1186/1475-2875-3-27)] [Medline: [15282029](https://pubmed.ncbi.nlm.nih.gov/15282029/)]
304. Gammino V, Nuhu A, Chenoweth P, Manneh F, Young RR, Sugerman DE, et al. Using geographic information systems to track polio vaccination team performance: pilot project report. *J Infect Dis* 2014 Nov 01;210 Suppl 1:S98-101. [doi: [10.1093/infdis/jit285](https://doi.org/10.1093/infdis/jit285)] [Medline: [25316882](https://pubmed.ncbi.nlm.nih.gov/25316882/)]
305. Perry B, Gesler W. Physical access to primary health care in Andean Bolivia. *Soc Sci Med* 2000 May;50(9):1177-1188. [Medline: [10728839](https://pubmed.ncbi.nlm.nih.gov/10728839/)]
306. Satava R, Angood PB, Harnett B, Macedonia C, Merrell R. The physiologic cipher at altitude: telemedicine and real-time monitoring of climbers on Mount Everest. *Telemed J E Health* 2000;6(3):303-313. [doi: [10.1089/153056200750040165](https://doi.org/10.1089/153056200750040165)] [Medline: [11110634](https://pubmed.ncbi.nlm.nih.gov/11110634/)]
307. Karim Aljewari Y, Ahmad R, Ahmed A. Providing Complete Precision Timing Solution for Hospitals by GPS Time Synchronized with MCS. *JNHC* 2015 Mar 01;2(1):1. [doi: [10.5176/2345-718x_2.1.67](https://doi.org/10.5176/2345-718x_2.1.67)]
308. Khan M, Beg S. Transference and retrieval of voice message over low signal strength in satellite communication. *Innovations Syst Softw Eng* 2012 Aug 22;8(4):293-299. [doi: [10.1007/s11334-012-0191-6](https://doi.org/10.1007/s11334-012-0191-6)]
309. Labrador V. www.britannica.com. Satellite Communication URL: <https://www.britannica.com/technology/satellite-communication> [accessed 2017-11-19] [WebCite Cache ID 6v63y6MG8]
310. www.inmarsat.com. Inmarsat and Global eHealth Foundation join forces to bring healthcare to the world's poorest communities URL: <http://www.inmarsat.com/press-release/inmarsat-global-ehealth-foundation-join-forces-bring-healthcare-worlds-poorest-communities/> [accessed 2016-12-20] [WebCite Cache ID 6zyfIE4aV]
311. Telemammographyinformation M. Telemammography and information management. *Academic Radiology* 1998 Nov;5:S484-S487. [doi: [10.1016/S1076-6332\(98\)80316-4](https://doi.org/10.1016/S1076-6332(98)80316-4)]

312. Agroyannis B, Fourtounas C, Romagnoli G, Skiadas M, Tsavdaris C, Chassomeris C, et al. Telemedicine technology and applications for home hemodialysis. *Int J Artif Organs* 1999 Oct;22(10):679-683. [Medline: [10585132](#)]
313. Anogianakis G, Maglavera S. Transeuropean network for the provision of added-VALUE Services in Telemedicine--(VAST-Net). *Stud Health Technol Inform* 1997;39:298-306. [Medline: [10168925](#)]
314. Anogianakis G, Maglavera S. MERMAID 1996--report on the implementation of a European Project on "medical emergency aid through telematics". *Stud Health Technol Inform* 1997;39:264-270. [Medline: [10168920](#)]
315. Anogianakis G, Maglavera S, Pomportsis A. Relief for maritime medical emergencies through telematics. *IEEE Trans Inf Technol Biomed* 1998 Dec;2(4):254-260. [Medline: [10719535](#)]
316. Anscombe DL. Healthcare delivery for oil rig workers: telemedicine plays a vital role. *Telemed J E Health* 2010;16(6):659-663. [doi: [10.1089/tmj.2010.9957](#)] [Medline: [20690836](#)]
317. Arbeille P, Ayoub J, Kieffer V, Combes B, Coitrieux A, Herve P, et al. Abdominal and fetal echography tele-operated in several medical centres sites, from an expert center, using a robotic arm & telephone or satellite link. *J Gravit Physiol* 2007 Jul;14(1):P139-P140. [Medline: [18372738](#)]
318. Armstrong IJ, Haston WS. Medical decision support for remote general practitioners using telemedicine. *J Telemed Telecare* 1997;3(1):27-34. [doi: [10.1258/1357633971930166](#)] [Medline: [9139758](#)]
319. Ausseresses AD. Telecommunications requirements for telemedicine. *J Med Syst* 1995 Apr;19(2):143-151. [Medline: [7602246](#)]
320. Bagshaw M. Telemedicine in British Airways. *J Telemed Telecare* 1996;2 Suppl 1:36-38. [Medline: [9375086](#)]
321. Bhaskaranarayana A, Satyamurthy LS, Remilla MLN. Indian Space Research Organization and telemedicine in India. *Telemed J E Health* 2009;15(6):586-591. [doi: [10.1089/tmj.2009.0060](#)] [Medline: [19659415](#)]
322. Bhaskaranarayana A, Satyamurthy L, Remilla ML, Sethuraman K, Rayappa H. Bridging Health Divide Between Rural and Urban Areas --Satellite Based Telemedicine Networks in India. In: Olla P, editor. *Space Technologies for the Benefit of Human Society and Earth*. Dordrecht: Springer; 2009.
323. Chonin A. Telehealth: important concepts for future nursing practice in space environments. *Life Support Biosph Sci* 1998;5(4):433-435. [Medline: [11871451](#)]
324. Cova G, Xiong H, Gao Q, Guerrero E, Ricardo R, Estevez J. A perspective of state-of-the-art wireless technologies for e-health applications. 2009 Presented at: International Symposium on IT in Medicine Education; 14-16 August 2009; Jinan, Shandong, China p. 76-81 Vol 1. [doi: [10.1109/ITIME.2009.5236457](#)]
325. Dasgupta A, Deb S. Telemedicine: a new horizon in public health in India. *Indian J Community Med* 2008 Jan;33(1):3-8 [FREE Full text] [doi: [10.4103/0970-0218.39234](#)] [Medline: [19966987](#)]
326. Delgorge C, Courrèges F, Al BL, Novales C, Rosenberger C, Smith-Guerin N, et al. A tele-operated mobile ultrasound scanner using a light-weight robot. *IEEE Trans Inf Technol Biomed* 2005 Mar;9(1):50-58. [Medline: [15787007](#)]
327. Ehrlich A, Kobrinsky BA, Petlakh VI, Rozinov VM, Shabanov VE. Telemedicine for a Children's Field Hospital in Chechnya. *J Telemed Telecare* 2007;13(1):4-6. [doi: [10.1258/135763307779701211](#)] [Medline: [17288651](#)]
328. Ernst RD, Kawashima A, Shepherd W, Tamm EP, Sandler CM. Distributing digital imaging and communications in medicine data and optimizing access over satellite networks. *J Digit Imaging* 1999 May;12(2 Suppl 1):195-196 [FREE Full text] [Medline: [10342210](#)]
329. Fiadjoe J, Gurnaney H, Muralidhar K, Mohanty S, Kumar J, Viswanath R, et al. Telemedicine consultation and monitoring for pediatric liver transplant. *Anesth Analg* 2009 Apr;108(4):1212-1214. [doi: [10.1213/ane.0b013e318198f786](#)] [Medline: [19299789](#)]
330. Ganapathy K. Telemedicine in the Indian context: an overview. *Stud Health Technol Inform* 2004;104:178-181. [Medline: [15747977](#)]
331. Garawi SA, Courreges F, Istepanian RSH, Zisimopoulos H, Gosset P. Performance analysis of a compact robotic tele-echography E-health system over terrestrial and mobile communication links. 2004 Presented at: Fifth IEEE International Conference on 3G Mobile Communication Technologies; 18-20 October, 2004; London, England p. 118-122. [doi: [10.1049/cp:20040647](#)]
332. Gortzis L, Papadopoulos H, Roelofs TA, Rakowsky S, Karnabatidis D, Siablis D, et al. Collaborative work during interventional radiological procedures based on a multicast satellite-terrestrial network. *IEEE Trans Inf Technol Biomed* 2007 Sep;11(5):597-599. [Medline: [17912978](#)]
333. Grant IC. Telemedicine in the British Antarctic survey. *Int J Circumpolar Health* 2004 Dec;63(4):356-364. [Medline: [15709312](#)]
334. Graschew G, Roelofs TA, Rakowsky S, Schlag PM. Interactive telemedical applications in OP 2000 via satellite. *Biomed Tech (Berl)* 2002;47 Suppl 1 Pt 1:330-333. [Medline: [12451856](#)]
335. Graschew G, Roelofs TA, Rakowsky S, Schlag PM. Interactive telemedicine as a tool to avoid a digital divide in the world. *Stud Health Technol Inform* 2004;103:150-156. [Medline: [15747917](#)]
336. Graschew G, Roelofs TA, Rakowsky S, Schlag PM, Heinzlreiter P, Kranzlmüller D, et al. New trends in the virtualization of hospitals--tools for global e-Health. *Stud Health Technol Inform* 2006;121:168-175. [Medline: [17095814](#)]

337. Hartvigsen G, Johansen MA, Hasvold P, Bellika JG, Arsand E, Arild E, et al. Challenges in telemedicine and eHealth: lessons learned from 20 years with telemedicine in Tromsø. *Stud Health Technol Inform* 2007;129(Pt 1):82-86. [Medline: [17911683](#)]
338. Hudson H. Rural telemedicine: lessons from Alaska for developing regions. *Telemed J E Health* 2005 Aug;11(4):460-467. [doi: [10.1089/tmj.2005.11.460](#)] [Medline: [16149892](#)]
339. Hwang S, Lee J, Kim H, Lee M. Development of a web-based picture archiving and communication system using satellite data communication. *J Telemed Telecare* 2000;6(2):91-96. [doi: [10.1258/1357633001935103](#)] [Medline: [10824376](#)]
340. Hwang S, Lee MH. A WEB-based telePACS using an asymmetric satellite system. *IEEE Trans Inf Technol Biomed* 2000 Sep;4(3):212-215. [Medline: [11026591](#)]
341. Jarvis-Selinger S, Chan E, Payne R, Plohman K, Ho K. Clinical telehealth across the disciplines: lessons learned. *Telemed J E Health* 2008 Sep;14(7):720-725. [doi: [10.1089/tmj.2007.0108](#)] [Medline: [18817503](#)]
342. Jones R, Clarke M, Kanellopoulos N, Lioupis D, Fowles R. The AIDMAN project – a telemedicine approach to cardiology investigation, referral and outpatient care. *J Telemed Telecare* 2016 Dec 02;6(1_suppl):32-34. [doi: [10.1258/1357633001933961](#)] [Medline: [10793965](#)]
343. Junejo ZA. Suparco Telemedicine Pilot Project. 2007 Presented at: 3rd International Conference on Recent Advances in Space Technologies; 14-16 June 2007; Istanbul, Turkey. [doi: [10.1109/RAST.2007.4284074](#)]
344. Kanthraj GR. Classification and design of teledermatology practice: what dermatoses? Which technology to apply? *J Eur Acad Dermatol Venereol* 2009 Aug;23(8):865-875. [doi: [10.1111/j.1468-3083.2009.03136.x](#)] [Medline: [19250331](#)]
345. Kanthraj GR, Srinivas CR. Store and forward teledermatology. *Indian J Dermatol Venereol Leprol* 2007;73(1):5-12 [FREE Full text] [Medline: [17314439](#)]
346. Kasitipradith N. The Ministry of Public Health telemedicine network of Thailand. *Int J Med Inform* 2001 May;61(2-3):113-116. [Medline: [11311665](#)]
347. Kayser K. Telepathology in Europe. Its practical use. *Arch Anat Cytol Pathol* 1995;43(4):196-199. [Medline: [8526551](#)]
348. Kayser K. [Telemedicine]. *Wien Klin Wochenschr* 1996;108(17):532-540. [Medline: [8992786](#)]
349. Koike K, Takizawa M, Kamiya S, Kamata M, Oleinikova OV, Bogatchenko M, et al. [Medical support for Belarus after Chernobyl accident using a telemedicine system]. *Igaku Butsuri* 2003;23(1):44-50. [Medline: [12832864](#)]
350. Komnakos D, Constantinou P, Vouyioukas D, Maglogiannis I. QoS Study Performance of an Integrated Satellite Telemedicine Platform. 2007 Presented at: 18th International Symposium on Personal, Indoor and Mobile Radio Communications; 3-7 Sept. 2007; Athens, Greece.
351. Kyriacou E, Pavlopoulos S, Berler A, Neophytou M, Bourka A, Georgoulas A, et al. Multi-purpose HealthCare Telemedicine Systems with mobile communication link support. *Biomed Eng Online* 2003 Mar 24;2:7 [FREE Full text] [Medline: [12694629](#)]
352. Latifi R, editor. *Current Principles and Practices of Telemedicine and E-health*. USA: IOS Press; 2008.
353. Lin CF, Lee HW. Wireless Multimedia Communication toward Mobile Telemedicine. 2009 Presented at: 9th WSEAS International Conference on APPLIED INFORMATICS AND COMMUNICATIONS (AIC '09); August 20 - 22, 2009; Moscow, Russia.
354. Maheu M, Whitten P, Allen A. *E-Health, Telehealth, and Telemedicine: A Guide to Startup and Success*. New Jersey: John Wiley and Sons; 2002.
355. Malik AZ. Telemedicine Country Report-Pakistan. 2007 Presented at: 9th International Conference on e-Health Networking, Application and Services; 19-22 June 2007; Taipei, Taiwan p. 90-94.
356. Massone C, Brunasso AMG, Campbell TM, Soyer HP. Mobile teledermoscopy--melanoma diagnosis by one click? *Semin Cutan Med Surg* 2009 Sep;28(3):203-205. [doi: [10.1016/j.sder.2009.06.002](#)] [Medline: [19782945](#)]
357. Meystre S. The current state of telemonitoring: a comment on the literature. *Telemed J E Health* 2005 Feb;11(1):63-69. [doi: [10.1089/tmj.2005.11.63](#)] [Medline: [15785222](#)]
358. Min Z. Applications of the satellite communication telemedicine system in our hospital. *CNKI Journal* 2006 Mar:7th section [FREE Full text]
359. Mishra SK, Kapoor L, Singh IP. Telemedicine in India: current scenario and the future. *Telemed J E Health* 2009;15(6):568-575. [doi: [10.1089/tmj.2009.0059](#)] [Medline: [19659413](#)]
360. Mitka M. Physicians find teleactivity hot near the North Pole. *JAMA* 1998 Oct 21;280(15):1296. [Medline: [9794295](#)]
361. Miyashita T, Takizawa M, Nakai K, Okura H, Kanda H, Murase S, et al. Telemedicine of the heart: real-time telescreening of echocardiography using satellite telecommunication. *Circ J* 2003 Jun;67(6):562-564 [FREE Full text] [Medline: [12808280](#)]
362. Mohr M. Telemedicine in oncology: European perspective. *Stud Health Technol Inform* 2008;131:255-261. [Medline: [18305335](#)]
363. Mougiakakou SG, Kyriacou E, Perakis K, Papadopoulos H, Androulidakis A, Konnis G, et al. A feasibility study for the provision of electronic healthcare tools and services in areas of Greece, Cyprus and Italy. *Biomed Eng Online* 2011 Jun 07;10:49 [FREE Full text] [doi: [10.1186/1475-925X-10-49](#)] [Medline: [21649924](#)]
364. Nakajima I, Sastrokusumo U, Mishra SK, Komiya R, Malik AZ, Tanuma T. The Asia Pacific telecommunity's telemedicine activities. 2006 Presented at: 8th International Conference on e-Health Networking, Applications and Services; 17-19 Aug. 2006; New Delhi, India.

365. Nakajima I, Sawada Y, Ashihara T, Takashima Y. Problems and our solutions for implementing telemedicine systems. *J Med Syst* 1999 Dec;23(6):425-435. [Medline: [10763162](#)]
366. Nicogossian AE, Poher DF, Roy SA. Evolution of telemedicine in the space program and earth applications. *Telemed J E Health* 2001;7(1):1-15. [doi: [10.1089/153056201300093813](#)] [Medline: [11321704](#)]
367. Niyato D, Hossain E, Diamond J. IEEE 802.16/WiMAX-based broadband wireless access and its application for telemedicine/e-health services. *IEEE Wireless Commun* 2007 Feb;14(1):72-83. [doi: [10.1109/MWC.2007.314553](#)]
368. Ohno G, Watanabe K, Okada Y, Higuchi K. Practical experience of telehealth between an Antarctic station and Japan. *J Telemed Telecare* 2012 Dec;18(8):473-475. [doi: [10.1258/jtt.2012.gth111](#)] [Medline: [23209273](#)]
369. Okada Y, Haruki Y, Ogushi Y. Disaster drills and continuous medical education using satellite-based Internet. *Methods Inf Med* 2000 Dec;39(4-5):343-347. [Medline: [11191704](#)]
370. O'Sullivan UM, Somers J. Southern Health Board--advanced telematic/telemedicine in healthcare services in the south west of Ireland. *Stud Health Technol Inform* 1999;64:230-240. [Medline: [10747543](#)]
371. Otto C, Bolte M, Linneweber F, Weber T. Functional characteristics of the telemedical network for the medical service of the Bundeswehr for support of operations outside Germany and civil-military co-operation. *Stud Health Technol Inform* 1999;64:270-274. [Medline: [10747548](#)]
372. Pak HS, Brown-Connolly NE, Bloch C, Clarke M, Clyburn C, Doarn CR, et al. Global forum on telemedicine: connecting the world through partnerships. *Telemed J E Health* 2008 May;14(4):389-395. [doi: [10.1089/tmj.2008.0030](#)] [Medline: [18570571](#)]
373. Paul NL. Telepsychiatry, the satellite system and family consultation. *J Telemed Telecare* 1997;3 Suppl 1:52-53. [doi: [10.1258/1357633971930364](#)] [Medline: [9218383](#)]
374. Pyke J, Hart M, Popov V, Harris RD, McGrath S. A tele-ultrasound system for real-time medical imaging in resource-limited settings. In: *Conf Proc IEEE Eng Med Biol Soc. 2007 Presented at: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 22-26 Aug. 2007; Lyon, France p. 3094-3097.* [doi: [10.1109/IEMBS.2007.4352982](#)]
375. Rao UR. Keynote address: global connectivity through telemedicine. *J Med Syst* 1995 Jun;19(3):295-304. [Medline: [7643026](#)]
376. Rayman R, Croome K, Galbraith N, McClure R, Morady R, Peterson S, et al. Robotic telesurgery: a real-world comparison of ground- and satellite-based internet performance. *Int J Med Robot* 2007 Jun;3(2):111-116. [doi: [10.1002/rcs.133](#)] [Medline: [17554810](#)]
377. Rayman R, Primak S, Patel R, Moallem M, Morady R, Tavakoli M, et al. Effects of latency on telesurgery: an experimental study. *Med Image Comput Comput Assist Interv* 2005;8(Pt 2):57-64. [Medline: [16685943](#)]
378. Reith A, Olsen DR. Teleradiology with satellite units - six years experience at the norwegian radium hospital. *Stud Health Technol Inform* 2008;134:209-216. [Medline: [18376048](#)]
379. Rieke J, Kleinholz L, Hosten N, Zendel W, Lemke A, Wielgus W, et al. Telemedicine in rural areas. Experience with medical desktop-conferencing via satellite. *J Telemed Telecare* 1995;1(4):224-228. [doi: [10.1177/1357633X9500100406](#)] [Medline: [9375147](#)]
380. Ronga L, Jayousi S, Del Re E, Colitta L, Iannone G, Scorpiniti A, et al. TESHEALTH: An integrated satellite/terrestrial system for e-health services. In: *International Conference on Communications (ICC). 2012 Presented at: An integrated satellite/terrestrial system for e-health services. Paper presented at the IEEE International Conference on Communications (ICC); 10-15 June 2012; Ottawa, ON, Canada p. 3286-3289.*
381. Samothrakakis S, Arvanitis TN, Plataniotis A, McNeill MD, Lister PF. WWW creates new interactive 3D graphics and collaborative environments for medical research and education. *Int J Med Inform* 1997 Nov;47(1-2):69-73. [Medline: [9506396](#)]
382. Sawai T, Uzuki M, Watanabe M. [Telepathology at presence and in the future]. *Rinsho Byori* 2000 May;48(5):458-462. [Medline: [10892295](#)]
383. 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](#)]
384. Stoloff P, Garcia FE, Thomason JE, Shia DS. A cost-effectiveness analysis of shipboard telemedicine. *Telemed J* 1998;4(4):293-304. [doi: [10.1089/tmj.1.1998.4.293](#)] [Medline: [10220469](#)]
385. Sudhamony S, Nandakumar K, Binu P, Issac Niwas S. Telemedicine and tele-health services for cancer-care delivery in India. *IET Commun* 2008;2(2):231. [doi: [10.1049/iet-com:20060701](#)]
386. Tachakra S, Wang XH, Istepanian RSH, Song YH. Mobile e-health: the unwired evolution of telemedicine. *Telemed J E Health* 2003;9(3):247-257. [doi: [10.1089/153056203322502632](#)] [Medline: [14611692](#)]
387. Takahashi T. The present and future of telemedicine in Japan. *Int J Med Inform* 2001 May;61(2-3):131-137. [Medline: [11311667](#)]
388. Takizawa M, Miyashita T, Murase S, Kanda H, Karaki Y, Yagi K, et al. [Mobile hospital -real time mobile telehealthcare system with ultrasound and CT van using high-speed satellite communication-]. *Igaku Butsuri* 2003;23(1):51-58. [Medline: [12832865](#)]

389. Takizawa M, Sone S, Takashima S, Feng L, Maruyama Y, Hasegawa M, et al. The mobile hospital--an experimental telemedicine system for the early detection of disease. *J Telemed Telecare* 1998;4(3):146-151. [doi: [10.1258/1357633981932127](https://doi.org/10.1258/1357633981932127)] [Medline: [10321042](https://pubmed.ncbi.nlm.nih.gov/10321042/)]
390. Tan J, Cheng W, Rogers WJ. From Telemedicine to E-Health: Uncovering New Frontiers of Biomedical Research, Clinical Applications and Public Health Services Delivery. *Journal of Computer Information Systems* 2016;42(5):7-18. [doi: [10.1080/08874417.2002.11647605](https://doi.org/10.1080/08874417.2002.11647605)]
391. Tyrer HW. Specifications for a satellite based wide area network. *Biomed Sci Instrum* 1999;35:111-116. [Medline: [11143331](https://pubmed.ncbi.nlm.nih.gov/11143331/)]
392. Tyrer HW, Wiedemeier PD, Cattlet RW. Rural telemedicine: satellites and fiber optics. *Biomed Sci Instrum* 2001;37:417-422. [Medline: [11347427](https://pubmed.ncbi.nlm.nih.gov/11347427/)]
393. VandenBos G, Williams S. The Internet versus the telephone: What is telehealth anyway? *Professional Psychology: Research and Practice* 2000;31(5):490-492. [doi: [10.1037/0735-7028.31.5.490](https://doi.org/10.1037/0735-7028.31.5.490)]
394. Vargas A, Ugalde M, Vargas R, Narvaez R, Geissbuhler A. [Telemedicine in Bolivia: RAFT-Altiplano project, experiences, future prospects, and recommendations]. *Rev Panam Salud Publica* 2014;35(5-6):359-364 [FREE Full text] [Medline: [25211562](https://pubmed.ncbi.nlm.nih.gov/25211562/)]
395. Watson D. Telemedicine. *Med J Aust* 1989 Jul 17;151(2):62-6, 68, 71. [Medline: [2739609](https://pubmed.ncbi.nlm.nih.gov/2739609/)]
396. Yamaguchi T. Performance tests of a satellite-based asymmetric communication network for the 'hyper hospital'. *J Telemed Telecare* 1997;3(2):78-82. [doi: [10.1258/1357633971930904](https://doi.org/10.1258/1357633971930904)] [Medline: [9206277](https://pubmed.ncbi.nlm.nih.gov/9206277/)]
397. Yamashita S, Shibata Y, Takamura N, Ashizawa K, Sera N, Eguchi K. Satellite communication and medical assistance for thyroid disease diagnosis from Nagasaki to Chernobyl. *Thyroid* 1999 Sep;9(9):969. [doi: [10.1089/thy.1999.9.969](https://doi.org/10.1089/thy.1999.9.969)] [Medline: [10524578](https://pubmed.ncbi.nlm.nih.gov/10524578/)]
398. Yokota K, Takamura N, Shibata Y, Yamashita S, Mine M, Tomonaga M. Evaluation of a telemedicine system for supporting thyroid disease diagnosis. *Stud Health Technol Inform* 2001;84(Pt 1):866-869. [Medline: [11604859](https://pubmed.ncbi.nlm.nih.gov/11604859/)]
399. Zhao J, Zhang Z, Guo H, Ren L, Chen S. Development and recent achievements of telemedicine in china. *Telemed J E Health* 2010 Jun;16(5):634-638. [doi: [10.1089/tmj.2009.0159](https://doi.org/10.1089/tmj.2009.0159)] [Medline: [20575733](https://pubmed.ncbi.nlm.nih.gov/20575733/)]
400. Casalino NAD, Alessandro and Fadda C. Organizational Impact and Exploitation of the Results of an Italian Research Project for E-Health and Medical Training. 2005 Presented at: European Conference on Information Systems; May 1, 2005; Regensburg, Germania.
401. Dańda J, Juszkievicz K, Leszczuk M, Loziak K, Papir Z, Sikora M, et al. Medical video server construction. *Pol J Pathol* 2003;54(3):197-204. [Medline: [14703288](https://pubmed.ncbi.nlm.nih.gov/14703288/)]
402. Geissbuhler A, Bagayoko CO, Ly O. The RAFT network: 5 years of distance continuing medical education and tele-consultations over the Internet in French-speaking Africa. *Int J Med Inform* 2007;76(5-6):351-356. [doi: [10.1016/j.ijmedinf.2007.01.012](https://doi.org/10.1016/j.ijmedinf.2007.01.012)] [Medline: [17331799](https://pubmed.ncbi.nlm.nih.gov/17331799/)]
403. Grasczew G, Roelofs TA, Rakowsky S, Schlag PM. The Virtual Hospital as a Digital Tool for e-Health. 2006 Presented at: World Congress on Medical Physics and Biomedical Engineering; September 1, 2006; Seoul, Korea p. 358-361.
404. Groves T. SatelLife: getting relevant information to the developing world. *BMJ* 1996;313(7072):1606-1609 [FREE Full text] [Medline: [8991004](https://pubmed.ncbi.nlm.nih.gov/8991004/)]
405. Jones R, Skirton H, McMullan M. Feasibility of combining e-health for patients with e-learning for students using synchronous technologies. *J Adv Nurs* 2006 Oct;56(1):99-109. [doi: [10.1111/j.1365-2648.2006.03984.x](https://doi.org/10.1111/j.1365-2648.2006.03984.x)] [Medline: [16972923](https://pubmed.ncbi.nlm.nih.gov/16972923/)]
406. Jossif A, Pattichis CS, Kyriakides M, Pitsillides A, Kyriacou E, Dikaiakos M. Selected eHealth applications in Cyprus from the training perspective. *Methods Inf Med* 2007;46(1):84-89. [Medline: [17224988](https://pubmed.ncbi.nlm.nih.gov/17224988/)]
407. Kiuchi T, Takahashi T. High speed digital circuits for medical communication; the MINCS-UH project. *Methods Inf Med* 2000 Dec;39(4-5):353-355. [Medline: [11191706](https://pubmed.ncbi.nlm.nih.gov/11191706/)]
408. Mahapatra AK, Mishra SK, Kapoor L, Singh IP. Critical issues in medical education and the implications for telemedicine technology. *Telemed J E Health* 2009;15(6):592-596. [doi: [10.1089/tmj.2009.0057](https://doi.org/10.1089/tmj.2009.0057)] [Medline: [19659416](https://pubmed.ncbi.nlm.nih.gov/19659416/)]
409. Masic Z, Novo A, Masic I, Kudumovic M, Toromanovic S, Rama A, et al. Distance learning at biomedical faculties in bosnia & herzegovina. *Stud Health Technol Inform* 2005;116:267-272. [Medline: [16160270](https://pubmed.ncbi.nlm.nih.gov/16160270/)]
410. Moahi KH. Health Information Networks for Telehealth in Africa ? Challenges and Prospects: a Review of the Literature. *Libri* 2009;49(1):43-58. [doi: [10.1515/libr.1999.49.1.43](https://doi.org/10.1515/libr.1999.49.1.43)]
411. Singh I, Kapoor L, Daman R, Mishra SK. Comparative study of connectivity in telemedicine. *Telemed J E Health* 2008 Oct;14(8):846-850. [doi: [10.1089/tmj.2008.0095](https://doi.org/10.1089/tmj.2008.0095)] [Medline: [18954256](https://pubmed.ncbi.nlm.nih.gov/18954256/)]
412. Tangalos E, McGee R, Bigbee AW. Use of the new media for medical education. *J Telemed Telecare* 1997;3(1):40-47. [doi: [10.1258/1357633971930184](https://doi.org/10.1258/1357633971930184)] [Medline: [9139760](https://pubmed.ncbi.nlm.nih.gov/9139760/)]
413. Yamauchi K, Ikeda M, Ota Y, Yang S, Ishigaki T, Itouji E, et al. Evaluation of the Space Collaboration System: its history, image quality and effectiveness for joint case conference. *Nagoya J Med Sci* 2000 May;63(1-2):19-24. [Medline: [10911717](https://pubmed.ncbi.nlm.nih.gov/10911717/)]
414. Angood PB, Satava R, Doarn C, Merrell R, E3 Group. Telemedicine at the top of the world: the 1998 and 1999 Everest extreme expeditions. *Telemed J E Health* 2000;6(3):315-325. [doi: [10.1089/153056200750040174](https://doi.org/10.1089/153056200750040174)] [Medline: [11110635](https://pubmed.ncbi.nlm.nih.gov/11110635/)]
415. Berek B, Canna M. Telemedicine on the move: health care heads down the information superhighway. *Hosp Technol Ser* 1994;13(6):1-65. [Medline: [10134202](https://pubmed.ncbi.nlm.nih.gov/10134202/)]

416. Brown-Connolly N, Concha JB, English J. Mobile health is worth it! Economic benefit and impact on health of a population-based mobile screening program in new Mexico. *Telemed J E Health* 2014 Jan;20(1):18-23 [[FREE Full text](#)] [doi: [10.1089/tmj.2013.0080](#)] [Medline: [24237397](#)]
417. Guo T, Laksanasopin T, Sridhara AA, Nayak S, Sia SK. Mobile device for disease diagnosis and data tracking in resource-limited settings. *Methods Mol Biol* 2015;1256:3-14. [doi: [10.1007/978-1-4939-2172-0_1](#)] [Medline: [25626528](#)]
418. Kishi Y. [The current situation of i-STAT (portable clinical analyzer) and its future view]. *Rinsho Byori* 2007 May;55(5):460-464. [Medline: [17593692](#)]
419. Latifi R, Ong C, Peck K, Porter J, Williams M. Telepresence and telemedicine in trauma and emergency care management. *Eur Surg* 2005 Oct;37(5):293-297. [doi: [10.1007/s10353-005-0180-1](#)]
420. Lin C. Mobile telemedicine: a survey study. *J Med Syst* 2012 Apr;36(2):511-520. [doi: [10.1007/s10916-010-9496-x](#)] [Medline: [20703699](#)]
421. Maglaveras N, Chouvarda I, Koutkias V, Lekka I, Tsakali M, Tsetoglou S, et al. Citizen centered health and lifestyle management via interactive TV: The PANACEIA-ITV health system. *AMIA Annu Symp Proc* 2003:415-419 [[FREE Full text](#)] [Medline: [14728206](#)]
422. Mupela EN, Mustarde P, Jones HLC. Telemedicine in primary health: the virtual doctor project Zambia. *Philos Ethics Humanit Med* 2011 May 13;6:9 [[FREE Full text](#)] [doi: [10.1186/1747-5341-6-9](#)] [Medline: [21569490](#)]
423. Nagatsuma H. Development of an Emergency Medical Video Multiplexing Transport System (EMTS): aiming at the nation-wide prehospital care in ambulance. *J Med Syst* 2003 Jun;27(3):225-232. [Medline: [12705454](#)]
424. Nakajima I. Japanese telemedical concept of ambulatory application. *J Med Syst* 2011 Apr;35(2):215-220. [doi: [10.1007/s10916-009-9358-6](#)] [Medline: [20703568](#)]
425. Rizou D, Sachpazidis I, Salvatore L, Sakas G. TraumaStation: a portable telemedicine station. In: *Conf Proc IEEE Eng Med Biol Soc*. 2009 Presented at: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 3-6 Sept. 2009; Minneapolis, MN, USA p. 1254-1257. [doi: [10.1109/IEMBS.2009.5333908](#)]
426. Shimizu K. Telemedicine by mobile communication. *IEEE Eng Med Biol Mag* 1999;18(4):32-44. [Medline: [10429900](#)]
427. Sibert K, Ricci MA, Caputo M, Callas PW, Rogers FB, Charash W, et al. The feasibility of using ultrasound and video laryngoscopy in a mobile telemedicine consult. *Telemed J E Health* 2008 Apr;14(3):266-272. [doi: [10.1089/tmj.2007.0050](#)] [Medline: [18570551](#)]
428. Chronaki CE, Berthier A, Lleo MM, Esterle L, Lenglet A, Simon F, et al. A satellite infrastructure for health early warning in post-disaster health management. *Stud Health Technol Inform* 2007;129(Pt 1):87-91. [Medline: [17911684](#)]
429. Garshnek V. Applications of space communications technology to critical human needs: rescue, disaster relief, and remote medical assistance. *Space Commun* 1991 Jul;8(3-4):311-317. [Medline: [11540154](#)]
430. Garshnek V, Burkle FM. Telecommunications systems in support of disaster medicine: applications of basic information pathways. *Ann Emerg Med* 1999 Aug;34(2):213-218. [Medline: [10424923](#)]
431. Gomez E, Poropatich R, Karinch M, Zajchuk J. Tertiary Telemedicine Support During Global Military Humanitarian Missions. *Telemedicine Journal* 1996 Jan;2(3):201-210. [doi: [10.1089/tmj.1.1996.2.201](#)] [Medline: [10165543](#)]
432. Meade K, Lam DM. A deployable telemedicine capability in support of humanitarian operations. *Telemed J E Health* 2007 Jun;13(3):331-340. [doi: [10.1089/tmj.2006.0040](#)] [Medline: [17603836](#)]
433. Merrell R, Cone SW, Rafiq A. Telemedicine in extreme conditions: disasters, war, remote sites. *Stud Health Technol Inform* 2008;131:99-116. [Medline: [18305326](#)]
434. Merrell RC, Doarn CR. Disasters--how can telemedicine help? *Telemed J E Health* 2005 Oct;11(5):511-512. [doi: [10.1089/tmj.2005.11.511](#)] [Medline: [16250812](#)]
435. Otsu Y, Choh T, Yamazaki I, Kosaka K, Iguchi M, Nakajima I. Experiments on the quick-relief medical communications via the Japan's domestic communication satellite CS-2 for the case of disasters and emergencies. *Acta Astronaut* 1986;13(6-7):459-466. [Medline: [11542833](#)]
436. Sadiq MA, Nagami K, Nakajima I, Juzoji H, Igarashi K, Tanaka K. Mobile telemedicine package for disasters. 2006 Presented at: HEALTHCOM 2006 8th International Conference on e-Health Networking, Applications and Services; 17-19 Aug. 2006; New Delhi, India. [doi: [10.1109/HEALTH.2006.246414](#)]
437. Simmons S, Alverson D, Poropatich R, D'Iorio J, DeVany M, Doarn CR. Applying telehealth in natural and anthropogenic disasters. *Telemed J E Health* 2008 Nov;14(9):968-971. [doi: [10.1089/tmj.2008.0117](#)] [Medline: [19035809](#)]
438. Simmons SC, Murphy TA, Blannarovich A, Workman FT, Rosenthal DA, Carbone M. Telehealth technologies and applications for terrorism response: a report of the 2002 coastal North Carolina domestic preparedness training exercise. *J Am Med Inform Assoc* 2003;10(2):166-176 [[FREE Full text](#)] [Medline: [12595406](#)]
439. Woodall J. Official versus unofficial outbreak reporting through the Internet. *Int J Med Inform* 1997 Nov;47(1-2):31-34. [Medline: [9506388](#)]
440. Ziadlou D, Eslami A, Hassani HR. Telecommunication Methods for Implementation of Telemedicine Systems in Crisis. 2008 Presented at: 2008 Third International Conference on Broadband Communications, Information Technology & Biomedical Applications; 23-26 Nov. 2008; Gauteng, South Africa.
441. Algaet M, Noh Z, Shibghatullah A, Milad A, Mustapha A. Provisioning Quality of Service of Wireless Telemedicine for E-Health Services: A Review. *Wireless Pers Commun* 2014 Apr 10;78(1):375-406. [doi: [10.1007/s11277-014-1758-3](#)]

442. Crump A. Satellite teleHealth: good for the cutting edge and in the bush? *Geospat Health* 2006 Nov;1(1):7-10. [doi: [10.4081/gh.2006.276](https://doi.org/10.4081/gh.2006.276)] [Medline: [18686228](https://pubmed.ncbi.nlm.nih.gov/18686228/)]
443. DeBakey M. Telemedicine Has Now Come of Age. *Telemedicine Journal* 1995 Jan;1(1):3-4. [doi: [10.1089/tmj.1.1995.1.3](https://doi.org/10.1089/tmj.1.1995.1.3)]
444. Duplaga M, Zieliński K. Evolution of IT-Enhanced Healthcare: From Telemedicine to e-Health. In: Zieliński K, Duplaga M, Ingram D, editors. *Information Technology Solutions for Healthcare*. London: Springer; 2006:1-21.
445. Durrani H, Khoja S. A systematic review of the use of telehealth in Asian countries. *J Telemed Telecare* 2009;15(4):175-181. [doi: [10.1258/jtt.2009.080605](https://doi.org/10.1258/jtt.2009.080605)] [Medline: [19471028](https://pubmed.ncbi.nlm.nih.gov/19471028/)]
446. Graszew G, Roelofs TA, Rakowsky S, Schlag PM. Network design for telemedicine--e-health using satellite technology. *Stud Health Technol Inform* 2008;131:67-82. [Medline: [18305324](https://pubmed.ncbi.nlm.nih.gov/18305324/)]
447. Kocian A, De Sanctis M, Rossi T. Hybrid satellite/terrestrial telemedicine services: Network requirements and architecture. 2011 Presented at: Aerospace Conference, 2011 IEEE; 5-12 March 2011; Big Sky, MT, USA. [doi: [10.1109/AERO.2011.5747335](https://doi.org/10.1109/AERO.2011.5747335)]
448. Lamminen H. Mobile satellite systems. *J Telemed Telecare* 1999;5(2):71-83. [doi: [10.1258/1357633991933323](https://doi.org/10.1258/1357633991933323)] [Medline: [10628016](https://pubmed.ncbi.nlm.nih.gov/10628016/)]
449. Merrell RC, Lee A, Kwankam SY, Mwape B, Chinyama C, Latifi R, et al. Satellite applications for telehealth in the developing world. *J Telemed Telecare* 2006;12(6):321-324. [doi: [10.1258/135763306778558105](https://doi.org/10.1258/135763306778558105)] [Medline: [17022843](https://pubmed.ncbi.nlm.nih.gov/17022843/)]
450. Srivastava S, Pant M, Abraham A, Agrawal N. The Technological Growth in eHealth Services. *Comput Math Methods Med* 2015;2015:894171 [FREE Full text] [doi: [10.1155/2015/894171](https://doi.org/10.1155/2015/894171)] [Medline: [26146515](https://pubmed.ncbi.nlm.nih.gov/26146515/)]
451. Tulu B, Chatterjee S, Maheshwari M. Telemedicine taxonomy: a classification tool. *Telemed J E Health* 2007 Jun;13(3):349-358. [doi: [10.1089/tmj.2006.0055](https://doi.org/10.1089/tmj.2006.0055)] [Medline: [17603838](https://pubmed.ncbi.nlm.nih.gov/17603838/)]
452. Vouyioukas D, Maglogiannis I, Pasiadis V. Pervasive E-health services using the DVB-RCS communication technology. *J Med Syst* 2007 Aug;31(4):237-246. [Medline: [17685147](https://pubmed.ncbi.nlm.nih.gov/17685147/)]
453. Zaidan B, Zaidan A, Mat Kiah M. Impact of Data Privacy and Confidentiality on Developing Telemedicine Applications: A Review Participates Opinion and Expert Concerns. *International J. of Pharmacology* 2011 Mar 1;7(3):382-387. [doi: [10.3923/jip.2011.382.387](https://doi.org/10.3923/jip.2011.382.387)]
454. Graszew G, Roelofs TA, Rakowsky S, Schlag PM. Design of Satellite-Based Networks for u-Health - GALENOS, DELTASS, MEDASHIP, EMISPHER. 2007 Presented at: 2007 9th International Conference on e-Health Networking, Application and Services; 19-22 June 2007; Taipei, Taiwan p. 19-22.
455. Arbeille P, Herault S, Roumy J, Porcher M, Besnard S, Vieyres P. 3D realtime echography and echography assisted by a robotic arm for investigating astronauts in the ISS from the ground. *J Gravit Physiol* 2001 Jul;8(1):P143-P144. [Medline: [12650206](https://pubmed.ncbi.nlm.nih.gov/12650206/)]
456. Arbeille P, Poisson G, Vieyres P, Ayoub J, Porcher M, Boulay JL. Echographic examination in isolated sites controlled from an expert center using a 2-D echograph guided by a teleoperated robotic arm. *Ultrasound Med Biol* 2003 Jul;29(7):993-1000. [Medline: [12878245](https://pubmed.ncbi.nlm.nih.gov/12878245/)]
457. Arbeille P, Ruiz J, Ayoub J, Vieyres P, Porcher M, Boulay J, et al. The robot and the satellite for tele-operating echographic examination in Earth isolated sites, or onboard ISS. *J Gravit Physiol* 2004 Jul;11(2):P233-P234. [Medline: [16240525](https://pubmed.ncbi.nlm.nih.gov/16240525/)]
458. Chiao L, Sharipov S, Sargsyan AE, Melton S, Hamilton DR, McFarlin K, et al. Ocular examination for trauma; clinical ultrasound aboard the International Space Station. *J Trauma* 2005 May;58(5):885-889. [Medline: [15920397](https://pubmed.ncbi.nlm.nih.gov/15920397/)]
459. Jones JA, Sargsyan AE, Barr YR, Melton S, Hamilton DR, Dulchavsky SA, et al. Diagnostic ultrasound at MACH 20: retroperitoneal and pelvic imaging in space. *Ultrasound Med Biol* 2009 Jul;35(7):1059-1067. [doi: [10.1016/j.ultrasmedbio.2009.01.002](https://doi.org/10.1016/j.ultrasmedbio.2009.01.002)] [Medline: [19427106](https://pubmed.ncbi.nlm.nih.gov/19427106/)]
460. Sargsyan AE, Hamilton DR, Jones JA, Melton S, Whitson PA, Kirkpatrick AW, et al. FAST at MACH 20: clinical ultrasound aboard the International Space Station. *J Trauma* 2005 Jan;58(1):35-39. [Medline: [15674147](https://pubmed.ncbi.nlm.nih.gov/15674147/)]
461. Vieyres P, Poisson G, Courreges F, Merigeaux O, Arbeille P. The TERESA project: from space research to ground tele-echography. *Ind Rob* 2003;30(1):77-82. [Medline: [12807138](https://pubmed.ncbi.nlm.nih.gov/12807138/)]
462. Cermack M. Monitoring and telemedicine support in remote environments and in human space flight. *Br J Anaesth* 2006 Jul;97(1):107-114 [FREE Full text] [doi: [10.1093/bja/ael132](https://doi.org/10.1093/bja/ael132)] [Medline: [16731572](https://pubmed.ncbi.nlm.nih.gov/16731572/)]
463. Doarn CR. Telemedicine in extreme environments: analogs for space flight. *Stud Health Technol Inform* 2003;97:35-41. [Medline: [15537228](https://pubmed.ncbi.nlm.nih.gov/15537228/)]
464. Doarn CR, Nicogossian AE, Merrell RC. Applications of telemedicine in the United States space program. *Telemed J* 1998;4(1):19-30. [doi: [10.1089/tmj.1.1998.4.19](https://doi.org/10.1089/tmj.1.1998.4.19)] [Medline: [9599070](https://pubmed.ncbi.nlm.nih.gov/9599070/)]
465. Feliciani F. Medical care from space: Telemedicine. *ESA Bull* 2003 May;114:54-59. [Medline: [14503492](https://pubmed.ncbi.nlm.nih.gov/14503492/)]
466. Grigoriev AI, Orlov OI. Telemedicine and spaceflight. *Aviat Space Environ Med* 2002 Jul;73(7):688-693. [Medline: [12137107](https://pubmed.ncbi.nlm.nih.gov/12137107/)]
467. Ross MD. Medicine in long duration space exploration: the role of virtual reality and broad bandwidth telecommunications networks. *Acta Astronaut* 2001;49(3-10):441-445. [Medline: [11669129](https://pubmed.ncbi.nlm.nih.gov/11669129/)]
468. Wilke D, Padeken D, Weber TH, Gerzer R. Telemedicine for the International Space Station. *Acta Astronaut* 1999;44(7-12):579-581. [Medline: [11542521](https://pubmed.ncbi.nlm.nih.gov/11542521/)]

469. Haidegger T, Sándor J, Benyó Z. Surgery in space: the future of robotic telesurgery. *Surg Endosc* 2011 Mar;25(3):681-690. [doi: [10.1007/s00464-010-1243-3](https://doi.org/10.1007/s00464-010-1243-3)] [Medline: [20652320](#)]
470. Friedman DS. Biomedical applications of NASA technology. *Med Des Mater* 1991 Feb;1(2):54-57. [Medline: [10171593](#)]
471. Hertzfeld HR. Measuring the economic returns from successful NASA life sciences technology transfers. *J Technol Transf* 2002 Dec;27(4):311-320. [Medline: [14983842](#)]
472. Hines JW. Medical and surgical applications of space biosensor technology. *Acta Astronaut* 1996;38(4-8):261-267. [Medline: [11541303](#)]
473. Hughes S. Opportunities for the transfer of astronomical technology to medicine. *Australas Phys Eng Sci Med* 2007 Dec;30(4):292-294. [Medline: [18274070](#)]
474. Kukkonen CA. NASA high performance computing, communications, image processing, and data visualization-potential applications to medicine. *J Med Syst* 1995 Jun;19(3):263-273. [Medline: [7643022](#)]
475. Lathers CM, Charles JB, Bungo MW. Humans in space: NASA's contributions to medical technology and health care. *J Clin Pharmacol* 1990 Mar;30(3):223-225. [Medline: [2138169](#)]
476. Rouse DJ, Brown JN, Whitten RP. Methodology for NASA technology transfer in medicine. *Med Instrum* 1981;15(4):234-236. [Medline: [7197751](#)]
477. Rouse DJ, Winfield DL, Canada SC. NASA spinoffs to bioengineering and medicine. *Acta Astronaut* 1991 Feb;25(2):103-110. [Medline: [11540738](#)]
478. Winfield D, Silbiger M, Brown GS, Clarke L, Dwyer S, Yaffe M, et al. Technology transfer in digital mammography. Report of the Joint National Cancer Institute-National Aeronautics and Space Administration workshop of May 19-20, 1993. *Invest Radiol* 1994 Apr;29(4):507-515. [Medline: [7913457](#)]
479. Winfield DL. Aerospace technology transfer to breast cancer imaging. *Acta Astronaut* 1997;41(4-10):515-523. [Medline: [11541150](#)]
480. Foale CM, Kaleri AY, Sargsyan AE, Hamilton DR, Melton S, Martin D, et al. Diagnostic instrumentation aboard ISS: just-in-time training for non-physician crewmembers. *Aviat Space Environ Med* 2005 Jun;76(6):594-598. [Medline: [15945407](#)]
481. Garshnek V. Space medicine comes down to Earth. *Space Policy* 1989 Nov;5(4):330-332. [Medline: [11540958](#)]
482. Grigoriev A, Agadjanyan AA, Baranov VM, Polyakov VV. On the contribution of space medicine in the public health care. *Acta Astronaut* 1997;41(4-10):531-536. [Medline: [11541151](#)]
483. Kozlovskaya IB, Egorov AD. Some approaches to medical support for Martian expedition. *Acta Astronaut* 2003;53(4-10):269-275. [Medline: [14649256](#)]
484. Bediang G, Perrin C, Ruiz DCR, Kamga Y, Sawadogo A, Bagayoko CO, et al. The RAFT Telemedicine Network: Lessons Learnt and Perspectives from a Decade of Educational and Clinical Services in Low- and Middle-Incomes Countries. *Front Public Health* 2014;2:180 [FREE Full text] [doi: [10.3389/fpubh.2014.00180](https://doi.org/10.3389/fpubh.2014.00180)] [Medline: [25340048](#)]

Abbreviations

COPUOS: Committee on the Peaceful Uses of Outer Space

GEO: Group on Earth Observations

GIS: geographic information system

GLONASS: global navigation satellite systems

GNSS: global navigation satellite systems

GP: general practitioner

GPS: global positioning system

HCP: health care professional

MeSH: Medical Subject Headings

NASA: National Aeronautics and Space Administration

NCD: noncommunicable disease

PA: physical activity

PM: particulate matter

SAFE: Satellites for Epidemiology

UNOSAT: United Nations Operational Satellite Applications Program

UN-SPIDER: United Nations platform for Space-based Information for Disaster Management and Emergency Response

UNOOSA: United Nations Office for Outer Space Affairs

WHO: World Health Organization

Edited by G Eysenbach; submitted 20.11.17; peer-reviewed by D Davies, T Mackey, C Paton; comments to author 24.01.18; revised version received 21.03.18; accepted 22.04.18; published 27.06.18.

Please cite as:

Dietrich D, Dekova R, Davy S, Fahrni G, Geissbühler A

Applications of Space Technologies to Global Health: Scoping Review

J Med Internet Res 2018;20(6):e230

URL: <http://www.jmir.org/2018/6/e230/>

doi: [10.2196/jmir.9458](https://doi.org/10.2196/jmir.9458)

PMID: [29950289](https://pubmed.ncbi.nlm.nih.gov/29950289/)

©Damien Dietrich, Ralitza Dekova, Stephan Davy, Guillaume Fahrni, Antoine Geissbühler. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 27.06.2018. 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.

Original Paper

Patient Acceptance of Remote Scribing Powered by Google Glass in Outpatient Dermatology: Cross-Sectional Study

Sandra Odenheimer^{1,2,3}, DNP, RN, FNP-BC; Deepika Goyal^{2,3}, MS, PhD, FNP; Veena Goel Jones^{4,5}, MD, FAAP; Ruth Rosenblum^{2,3}, DNP, RN, PNP-BC, CNS; Lam Ho⁴, MHSA; Albert S Chan^{4,5}, MS, MD, FAAFP

¹Palo Alto Medical Foundation, Sunnyvale, CA, United States

²The Valley Foundation School of Nursing, San Jose State University, San Jose, CA, United States

³Department of Nursing, California State University Fresno, Fresno, CA, United States

⁴Office of Patient Experience, Sutter Health, Sacramento, CA, United States

⁵Palo Alto Medical Foundation Research Institute, Palo Alto, CA, United States

Corresponding Author:

Sandra Odenheimer, DNP, RN, FNP-BC

Palo Alto Medical Foundation

401 Old San Francisco Road

Sunnyvale, CA, 94086

United States

Phone: 1 4086214196

Fax: 1 4087357297

Email: odenhes@pamf.org

Abstract

Background: The ubiquitous use of electronic health records (EHRs) during medical office visits using a computer monitor and keyboard can be distracting and can disrupt patient-health care provider (HCP) nonverbal eye contact cues, which are integral to effective communication. Provider use of a remote medical scribe with face-mounted technology (FMT), such as Google Glass, may preserve patient-HCP communication dynamics in health care settings by allowing providers to maintain direct eye contact with their patients while still having access to the patient's relevant EHR information. The medical scribe is able to chart patient encounters in real-time working in an offsite location, document the visit directly into EHR, and free HCP to focus only on the patient.

Objective: The purpose of this study was to examine patient perceptions of their interactions with an HCP who used FMT with a remote medical scribe during office visits. This includes an examination of any association between patient privacy and trust in their HCP when FMT is used in the medical office setting.

Methods: For this descriptive, cross-sectional study, a convenience sample of patients was recruited from an outpatient dermatology clinic in Northern California. Participants provided demographic data and completed a 12-item questionnaire to assess their familiarity, comfort, privacy, and perceptions following routine office visits with an HCP where FMT was used to document the clinical encounter. Data were analyzed using appropriate descriptive and inferential statistics.

Results: Over half of the 170 study participants were female (102/170, 59.4%), 60.0% were Caucasian (102/170), 24.1% were Asian (41/170), and 88.8% were college-educated (151/170). Age ranged between 18 and 90 years (mean 50.5, SD 17.4). The majority of participants (118/170, 69.4%) were familiar with FMT, not concerned with privacy issues (132/170, 77.6%), and stated that the use of FMT did not affect their trust in their HCP (139/170, 81.8%). Moreover, participants comfortable with the use of FMT were less likely to be concerned about privacy ($P<.001$) and participants who trusted their HCP were less likely to be concerned about their HCP using Google Glass ($P<.009$). Almost one-third of them self-identified as early technology adopters (49/170, 28.8%) and 87% (148/170) preferred their HCP using FMT if it delivered better care.

Conclusions: Our study findings support the patient acceptance of Google Glass use for outpatient dermatology visits. Future research should explore the use of FMT in other areas of health care and strive to include a socioeconomically diverse patient population in study samples.

(*J Med Internet Res* 2018;20(6):e10762) doi:[10.2196/10762](https://doi.org/10.2196/10762)

KEYWORDS

acceptance, clinician burnout, communication, Google Glass, health care provider, patient, remote scribing, trust

Introduction

In an effort to improve health care quality, outcomes, and reduce health care costs, the Health Information Technology for Economic and Clinical Health (HITECH) Act [1] mandated the use of electronic health records (EHRs). Across the country, EHRs are now almost ubiquitously being used to support patient care by means of clinical decision support and reduction of medical errors; however, they have been associated with unintended consequences for health care providers (HCPs). HCPs are citing increased EHR administrative tasks as a key cause of clinician burnout [2]. In fact, time log studies have demonstrated that HCPs are spending more time in performing administrative tasks such as charting and less time face-to-face with their patients [3-5]. Increased administrative burden affects providers' work-life balance, leading to decreased physician job satisfaction and burnout [6]. Increased burnout may prompt providers to leave the field of medicine. Furthermore, the use of EHR changes the patient-provider communication dynamic by drawing HCP attention away from patients [7,8], despite evidence that direct eye contact promotes effective patient-HCP communication [9]. Poor communication between patients and HCP can also affect rapport, patient satisfaction, adherence to treatment, clinical outcomes, and patient trust [10,11], highlighting the importance and need for solutions to preserve the connection that patients need with their clinicians.

The integration of medical scribes into clinical care has been introduced as one method used to preserve patient-HCP eye contact and communication and reduce HCP workload and charting [12]. Having a scribe perform EHR documentation of the clinical visit, providers can focus their attention on the patient. In this capacity, medical scribes can be present in the examination room during the medical visit or could work remotely. Remote "virtual scribes" are also able to chart patient encounters in real-time working on or off site from a Health Insurance Portability and Accountability Act (HIPAA)-secure location, significantly reducing data entry workload that EHRs place on HCPs [13]. The addition of a virtual scribe can play a critical role in the provider-patient experience by decreasing charting and allowing the provider to fully engage with the patient and, thereby, positively impacting the patient experience [12].

Google Glass (GG), a face-mounted wearable technology, is emerging as a tool in health care settings. It has been used in several patient care areas including remote electrocardiogram interpretation [14], evaluation accuracy and triage times among paramedics in the field [15], and radiological intervention procedures [16]. Although such studies have demonstrated the value of using face-mounted technology (FMT) in health care, little is known about patient perceptions surrounding providers' use of this technology. Prochaska and colleagues [17] surveyed patient perceptions of GG in a sample of hospitalized patients. In that sample, most respondents were unfamiliar with GG. Nearly half of respondents were concerned about privacy, but

most patients were open to their providers using GG if it would improve their care.

At our organization, remote scribing via FMT was implemented to tackle the issue of burnout by connecting clinicians to remote medical scribes during patient visits and enabling real-time documentation in the EHR. Initial data has demonstrated a promising reduction in HCP EHR documentation burden. However, with the growing use of remote scribing via FMT in patient care, little empirical attention has been given to FMT and patient perceptions of FMT use in outpatient settings. The purpose of this study was to examine how the use of remote scribing with GG, by HCPs, is perceived by patients in an outpatient clinical setting.

Methods**Overview**

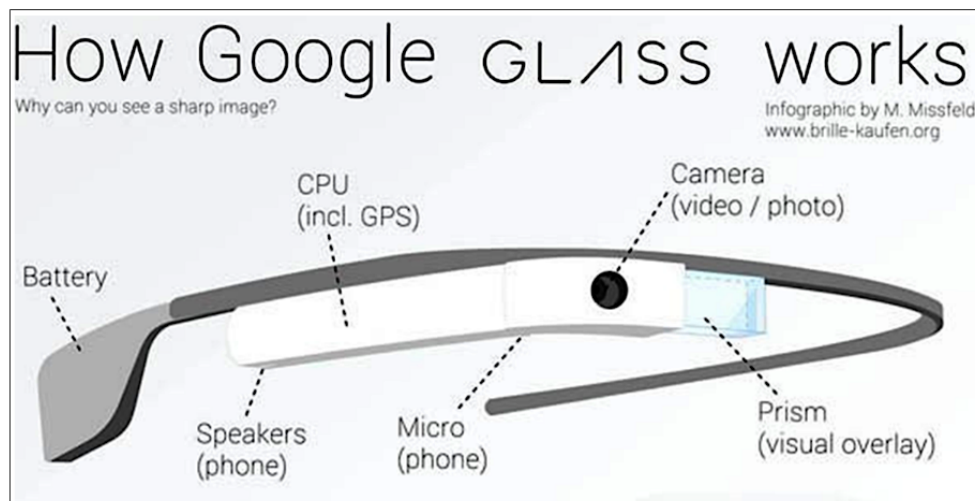
This descriptive cross-sectional study was approved by the Fresno State University and Sutter Health's Institutional Review Boards. A convenience sample of patients over 18 years of age, who could read and understand English, and who were being seen for an outpatient dermatology office visit by an HCP who used FMT were invited to participate. At the time of data collection, 4 out of 11 HCPs in the outpatient dermatology clinic used FMT.

Study Recruitment

Participants were approached by the principal investigator (PI) after their clinic office visit and were provided with a description of the study. Interested participants were taken to a private room in the clinic where the PI or research assistant verified that the study criteria were met and informed consent was obtained. Study questionnaires were completed on computer tablets, which took 5-10 minutes. Data collection occurred over a 4-day period in September 2017.

Measures

Participants were asked to provide general demographic data, including gender, age, race, and education level. Following this, they completed a 12-item questionnaire, which was adapted from the questionnaire developed by Prochaska and colleagues [17], with added questions regarding the level of technology and privacy. The items assessed patient familiarity, comfort, and privacy level with FMT as well as their trust, relationship, and communication with their HCP. Patient trust in HCP was rated on a 4-point Likert scale from 1 to 4 (more likely to trust provider, no change, less likely to trust provider, or I do not know). Relationship and communication with HCP was rated on a 10-point scale, from 1 (poor relationship/communication) to 10 (excellent relationship/communication). Participants were also asked to rate their level of technology adoption (innovator, early adopter, early majority, late majority, and laggard). Space was left at the end of the questionnaire for participants to add any additional comments.

Figure 1. Google glass specifications. CPU: central processing unit; GPS: Global Positioning System.

Technology and Equipment

GG is a face-mounted device that is similar to eyeglasses. The GG unit has the capability to connect to a phone via Wi-Fi enabling hands-free internet access. A small optical display is mounted above the right eye and a camera, a microphone, a speaker, and wireless connectivity is built into an eyeglass frame, which is operated through voice commands and a touchpad (see [Figure 1](#)). Each HCP in the study facility was assigned his or her own pair of GG. Custom lenses, which were compatible with GG, were available for HCPs requiring prescription glasses. Patients provided verbal permission for HCP to use GG at the beginning of each visit. In cases wherein permission was not granted, HCPs removed the GG device and remote documentation for the visit was not done.

The remote scribe observes the clinician–patient interaction and performs the documentation of the visit with HCP speaking “out loud” about the subjective history, objective findings, assessment, and treatment plan. Any comments or clarifications that the remote scribe has are communicated back via written messages that appear on the FMT display and are only visible to HCP. Remote scribes can also provide patient information within the field of vision via FMT, such as requests for data look up from the EHR, so that HCP can simultaneously perform other tasks or procedures. HCPs have the ability to receive data and input patient information through the small screen within the FMT lens, which is only discernible to the wearer, from the remote scribe. This reduces the need for HCP to look away to a desktop screen, allowing HCP to focus on the patient.

At the time of data collection for this study, HCPs had been using GG for approximately 12 months. Each HCP was assigned his or her own GG device and remote scribe. All communication between the remote scribe and HCP was encrypted and followed HIPAA operational, security, and privacy protocols to safeguard patient information. All remote scribe facilities completed a rigorous security and privacy review, which was conducted by an independent third-party auditor prior to HCP use with patients.

Data Analysis

Descriptive data, including participant characteristics, and some questionnaire data were analyzed using descriptive statistics (frequencies, means, and measures of central tendency). Appropriate inferential statistics, including chi-square and analysis of variance tests, were used to identify the associations between variables. Narrative comments were reviewed by the research team for common themes.

Results

Participant Characteristics

Of the 170 participants who completed study questionnaires, over half (102/170, 59.4%) were female. The largest ethnic demographic was Caucasian (102/170, 60.0%) and second was Asian (41/170, 24.1%). The majority were college-educated or more (151/170, 88.8%). Patient ages ranged from 18 to 90 years of age (mean 50.5, SD 17.4; see [Table 1](#)).

Patient Perception of Google Glass

Level of Technology

When asked to describe feelings about new technologies, 25 participants (14.7%) classified themselves as innovators (the first to adopt new technologies) and 24 (14.1%) as early majority (adopt new technology when it is still new but most people do not have it). Seventy-three (42.9%) classified themselves as early adopters (selective in adopting new technology), 41 (24.1%) rated themselves as late majority (adopt new technology after the majority of people are using it and it becomes commonplace), and 7 (4.1%) were laggards (one of the last to adopt new technology; you wait until all the bugs are out and it is inexpensive to purchase). Participants with a higher level of education were significantly more likely to be among the first to adopt a new technology ($\chi^2_{24}=64.8$, $P<.001$).

When asked about their concerns surrounding the use of FMT, the majority (122/170, 73.9%) stated that they had no concerns, few (15/170, 8.8%) stated that they have security-related concerns, and very few (2/170, 1.2%) stated that FMT might be distracting.

Table 1. Participant characteristics (N=170).

Characteristics	Value
Age (years), mean (SD), range	50.5 (17.4), 18-90
Gender, n (%)	
Female	101 (59.4)
Male	69 (40.6)
Ethnicity, n (%)	
Caucasian	102 (60.0)
Asian	41 (24.1)
Hispanic/Latino	12 (7.1)
African American	4 (2.4)
Other	11 (6.4)
Education level, n (%)	
Completed high school	11 (5.9)
Some college, no degree	16 (9.4)
College degree	60 (35.3)
Post graduate	75 (44.1)
Other/prefer not to answer	9 (5.3)

Familiarity and Comfort With Google Glass

A large number of participants (118/170, 69.4%) reported being very or somewhat familiar with FMT, 87.1% (148/170) were extremely or somewhat comfortable with their HCP using FMT during the office visit, and 87.1% (148/170) preferred their HCP use FMT if it helped them deliver better care. Additionally, participants who were comfortable with their HCP using FMT were less likely to be concerned about privacy ($X^2_{16}=89.40$, $P<.001$).

Privacy and Trust

Few (38/170, 22.4%) participants reported being very or somewhat concerned with privacy. The majority (139/170, 81.8%) reported no change in their level of trust related to the use of FMT with 12.9% (22/170) reporting that the use of FMT would increase trust in their HCP. A significant relationship was noted between the participants' privacy concerns relating to the use of FMT and trust in their HCP ($X^2_{12}=26.5$, $P<.009$).

Relationship and Communication

Participants' relationships with their HCP averaged to 9.4 (SD 0.93) and communication averaged to 9.5 (SD 1.10) on 10-point scales (Table 2). Chi-square tests of independence were performed to examine the relationships between variables (see Table 2 for complete survey results).

Participant Narrative Comments

Fifty-five participants (32.4%) provided narrative comments at the end of their surveys. Comments conveyed not noticing HCP was using FMT; for example, "I was so involved in our visit, I didn't even notice" and "I didn't really even notice FMT for most of the visit" as well as an overall feeling of FMT being associated with better care, "If it helps her keep track of my care, I am all for it," "I feel more details are being documented," and "If it helps with transcription then it is a great idea." Comments also conveyed patient satisfaction; for example, "If it provides more face time with the doctor, I think it is worth it," and "It is nice to have more interaction with the doctor versus them looking at the computer to take notes."

Table 2. Perception of Google Glass (N=170).

Question	n (%)
Are you familiar with Google Glass?	
Very familiar	16 (9.4)
Somewhat familiar	102 (60.0)
Neither familiar or unfamiliar	16 (9.4)
Somewhat unfamiliar	11 (6.5)
Very unfamiliar	25 (14.7)
How comfortable were you when your dermatology provider was wearing Google Glass?	
Extremely comfortable	110 (64.7)
Somewhat comfortable	38 (22.4)
Neither comfortable or uncomfortable	19 (11.2)
Somewhat uncomfortable	2 (1.2)
Extremely uncomfortable	1 (0.6)
Was privacy a concern when your dermatology provider was using Google Glass?	
Very concerned	10 (5.9)
Somewhat concerned	28 (16.5)
Neither concerned or unconcerned	35 (20.6)
Somewhat unconcerned	15 (8.8)
Very unconcerned	82 (48.2)
How did Google Glass affect your trust in your dermatology provider?	
More likely to trust my provider	22 (12.9)
No change	139 (81.8)
Less likely to trust my provider	2 (1.2)
I do not know	7 (4.1)
If your dermatology provider said that Google Glass >helped them to deliver better care, would you want them to wear Google Glass during your next visit?	
Yes	148 (87.1)
No	3 (1.8)
I do not know	11 (6.5)
I need to know more	8 (4.7)
Would you have concerns if your dermatology provider used Google Glass during a visit? Check all that apply.	
I would have no concerns	122 (73.9)
Security	15 (8.8)
Privacy	28 (16.5)
It may be distracting	1 (1.2)
Unfamiliar with Google Glass	7 (4.1)
Other: Security and Privacy	8 (4.7)
Multiple answers	11 (6.6)
Choose the phrase from the list below that best describes your feelings about new technologies. You are:	
An Innovator: First to adopt new technology	25 (14.7)
An Early Adopter: Selective in adopting new technology	73 (42.9)
An Early Majority: Adopt new technology when it is still new but most people do not have it	24 (14.1)
A Late Majority: Adopt new technology after the majority of people are using it and it becomes commonplace	41 (24.1)

Question	n (%)
A Laggard: One of the last to adopt new technology, you wait until all the bugs are out and it is inexpensive to purchase	7 (4.1)

Discussion

Principal Findings

To our knowledge, this is the first study to examine patient perceptions of FMT in an outpatient clinical setting; our findings build upon the work of Prochaska and colleagues [17] who examined patient perceptions of GG in a hospital setting. Although the primary reason FMT was implemented at our organization was to decrease charting documentation time and reduce clinician burnout, our results showed that FMT allowed providers to improve their interactions with patients. Remote medical scribes can alleviate HCP documentation burden and our findings support the use of virtual scribes and FMT, given that few participants expressed concern (38/170, 22.4%) with the use of FMT, 81.8% (139/170) reported no change in their level of trust with the use of FMT, and 87.1% (148/170) wanted their HCP to use FMT if it helped him or her deliver better care. Overall, our findings are in agreement with those of the study conducted by Prochaska and colleagues [17] who found that 65% (56/86) of respondents wanted their doctors to wear GG if it improved their care.

A key consideration with the deployment of remote scribes is how FMT might affect privacy and trust between the clinician and patient. Contextually, this is interesting because unlike in-person scribes that are physically present in the exam room at time of the clinical encounter, the remote scribe is not known or ever seen by the patient. In contrast to the study by Prochaska and colleagues [17], wherein nearly half of respondents were concerned about privacy, fewer participants (38/170, 22.3%) in our study were concerned with the privacy of their personal health information. Participant comments reflected little concern with trust and privacy, and the use of FMT did not change their level of trust in providers. Patients preferred their HCP use FMT ((148/170; 87.1%) if it helped deliver better care. Our institution's deployment of remote scribing over 4 years and patient experience of the benefits versus a survey of the theoretical use of FMT administered in a hospital setting may explain the difference in concerns about privacy in our sample.

Patients in our study were more familiar with GG (118/170, 69.4%) than those (23/86, 27%) in the study conducted by Prochaska [17]. The majority of our study participants (122/170, 71.7%) considered themselves as early adopters of technology which could be explained by the fact that the study was conducted in Silicon Valley, the nation's technology center. For many Silicon Valley residents, cutting edge technology is a part of their normal lives [18]. This familiarity and comfort may again be a consequence of an overall higher technology adoption in our geographical region. Although otherwise ethnically diverse, this study's population was largely college-educated (151/170, 88.8%), which could also have contributed to their comfort and acceptance of the technology.

Study participants gave HCPs, who used FMT, high ratings (9.5 out of 10) regarding the perceptions of effective communication. Patients perceived that HCPs using FMT were more attentive, had greater focus, and communicated greater empathy. Also, based on patient comments, participants were more satisfied with their visits and felt that HCPs delivered more personal experiences by spending less time on the computer. This is supported by literature, which demonstrates that HCPs who spend more time communicating face-to-face, focusing their attention more on the patient and less on EHR, can positively influence their communication with patients [6]. Furthermore, patients noted that the improved interaction with their HCP made them less aware of the use of the GG technology.

The use of remote scribing via GG in outpatient dermatology settings may reduce HCP documentation time, increase efficiency, reduce charting errors/omissions, and reduce workflow stress. Health care providers can simply interact with their patient, reviewing their clinical histories and examination findings, empowered by a remote scribe who enters data directly into the patient's EHR in real-time. This novel approach may facilitate a more meaningful use of EHR and realization of its benefits in clinical care without adding an administrative burden to HCPs. Owing to the fact that EHRs are now an integral part of clinical practice, health care organizations should continue to seek new methods of using EHR in ways that improve provider satisfaction, organizational efficiencies, and patient-provider interactions.

Limitations

Our findings should be interpreted with caution, given the largely insured and well-educated sample, limiting generalizability to lower socioeconomic populations. Additionally, the study sample was drawn from a Northern California clinic located in Silicon Valley, where technology may be more readily accepted than in other parts of the United States. This technology acumen could account for the participants' lack of concerns when using FMT, and it is possible that the acceptance of FMT would decrease in more rural and conservative areas. Lastly, data regarding participants' prior encounters with HCPs at the study site were not collected, further limiting study findings.

Conclusions

Our study findings identified a high level of patient acceptance of FMT in a dermatology clinic setting where FMT was implemented in an effort to reduce provider burnout through the use of virtual remote scribes. Future research examining wearable technology such as GG should strive to include patients from other clinical settings and from diverse geographic areas and socioeconomic backgrounds. Other outcomes, for example, HCP satisfaction with FMT, whether the use of FMT increases HCP efficiency/productivity, HCP relationships with medical scribes, and experiences of medical scribes should also be examined in future studies.

Acknowledgments

The corresponding author would like to thank the study participants for their time and the Sutter Health Office of Patient Experience and Augmedix for supporting the implementation of remote scribing via GG. The corresponding author thanks the contributing authors for their assistance in preparing this manuscript, and Martin Missfeldt for designing the infographic seen in [Figure 1](#). Sutter Health is a strategic investor in Augmedix Inc (San Francisco, CA) and the provider of the remote scribing solution.

Conflicts of Interest

None declared.

References

1. US Department of Health & Human Services. Title XIII Health information technology: HITECH Act 2009 URL: <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/hitechact.pdf> [accessed 2018-03-03] [WebCite Cache ID 6znTdkQnw]
2. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014;12(6):573-576 [FREE Full text] [doi: [10.1370/afm.1713](https://doi.org/10.1370/afm.1713)] [Medline: [25384822](https://pubmed.ncbi.nlm.nih.gov/25384822/)]
3. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. *Ann Fam Med* 2017 Sep;15(5):419-426 [FREE Full text] [doi: [10.1370/afm.2121](https://doi.org/10.1370/afm.2121)] [Medline: [28893811](https://pubmed.ncbi.nlm.nih.gov/28893811/)]
4. Payne TH, Corley S, Cullen TA, Gandhi TK, Harrington L, Kuperman GJ, et al. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *J Am Med Inform Assoc* 2015 Sep;22(5):1102-1110 [FREE Full text] [doi: [10.1093/jamia/ocv066](https://doi.org/10.1093/jamia/ocv066)] [Medline: [26024883](https://pubmed.ncbi.nlm.nih.gov/26024883/)]
5. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients And Desktop Medicine. *Health Aff (Millwood)* 2017 Apr 01;36(4):655-662 [FREE Full text] [doi: [10.1377/hlthaff.2016.0811](https://doi.org/10.1377/hlthaff.2016.0811)] [Medline: [28373331](https://pubmed.ncbi.nlm.nih.gov/28373331/)]
6. DiAngi Y, Longhurst CA, Payne TH. Taming the EHR (Electronic Health Record) - There is Hope. *J Fam Med* 2016;3(6):1072 [FREE Full text] [Medline: [27830215](https://pubmed.ncbi.nlm.nih.gov/27830215/)]
7. Asan O, Young HN, Chewning B, Montague E. How physician electronic health record screen sharing affects patient and doctor non-verbal communication in primary care. *Patient Educ Couns* 2015 Mar;98(3):310-316 [FREE Full text] [doi: [10.1016/j.pec.2014.11.024](https://doi.org/10.1016/j.pec.2014.11.024)] [Medline: [25534022](https://pubmed.ncbi.nlm.nih.gov/25534022/)]
8. Margalit RS, Roter D, Dunevant MA, Larson S, Reis S. Electronic medical record use and physician-patient communication: an observational study of Israeli primary care encounters. *Patient Educ Couns* 2006 Apr;61(1):134-141. [doi: [10.1016/j.pec.2005.03.004](https://doi.org/10.1016/j.pec.2005.03.004)] [Medline: [16533682](https://pubmed.ncbi.nlm.nih.gov/16533682/)]
9. Kraft-Todd GT, Reinero DA, Kelley JM, Heberlein AS, Baer L, Riess H. Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PLoS One* 2017 May;12(5):e0177758 [FREE Full text] [doi: [10.1371/journal.pone.0177758](https://doi.org/10.1371/journal.pone.0177758)] [Medline: [28505180](https://pubmed.ncbi.nlm.nih.gov/28505180/)]
10. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. *J Gen Intern Med* 2009 Mar;24(3):341-348 [FREE Full text] [doi: [10.1007/s11606-008-0892-6](https://doi.org/10.1007/s11606-008-0892-6)] [Medline: [19130148](https://pubmed.ncbi.nlm.nih.gov/19130148/)]
11. Tabler J, Scammon D, Kim J, Farrell T, Tomoiaia-Cotisel A, Magill M. Patient care experiences and perceptions of the patient-provider relationship: A mixed method study. *Patient care experiences and perceptions of the patient-provider relationship: A mixed method study* 2014;1(1):75-87 [FREE Full text]
12. Earls S, Savageau JA, Begley S, Saver BG, Sullivan K, Chuman A. Can scribes boost FPs' efficiency and job satisfaction? *J Fam Pract* 2017 Apr;66(4):206-214. [Medline: [28375393](https://pubmed.ncbi.nlm.nih.gov/28375393/)]
13. Brady K, Shariff A. Virtual medical scribes: making electronic medical records work for you. *J Med Pract Manage* 2013;29(2):133-136. [Medline: [24228379](https://pubmed.ncbi.nlm.nih.gov/24228379/)]
14. Jeroudi OM, Christakopoulos G, Christopoulos G, Kotsia A, Kypreos MA, Rangan BV, et al. Accuracy of remote electrocardiogram interpretation with the use of Google Glass technology. *Am J Cardiol* 2015 Feb 01;115(3):374-377. [doi: [10.1016/j.amjcard.2014.11.008](https://doi.org/10.1016/j.amjcard.2014.11.008)] [Medline: [25482681](https://pubmed.ncbi.nlm.nih.gov/25482681/)]
15. Cicero MX, Walsh B, Solad Y, Whitfill T, Paesano G, Kim K, et al. Do you see what I see? Insights from using google glass for disaster telemedicine triage. *Prehosp Disaster Med* 2015 Feb;30(1):4-8. [doi: [10.1017/S1049023X1400140X](https://doi.org/10.1017/S1049023X1400140X)] [Medline: [25571779](https://pubmed.ncbi.nlm.nih.gov/25571779/)]
16. Vorraber W, Voessner S, Stark G, Neubacher D, DeMello S, Bair A. Medical applications of near-eye display devices: an exploratory study. *Int J Surg* 2014 Dec;12(12):1266-1272 [FREE Full text] [doi: [10.1016/j.ijsu.2014.09.014](https://doi.org/10.1016/j.ijsu.2014.09.014)] [Medline: [25462704](https://pubmed.ncbi.nlm.nih.gov/25462704/)]

17. Prochaska M, Press V, Meltzer D, Arora V. Patient Perceptions of Wearable Face-Mounted Computing Technology and the Effect on the Doctor-Patient Relationship. *Appl Clin Inform* 2016 Dec 12;7(4):946-953 [[FREE Full text](#)] [doi: [10.4338/ACI-2016-06-LE-0094](#)] [Medline: [27730249](#)]
18. Ryan C. <https://census.gov>. 2017. Computer and Internet Use in the United States: 2015 URL: <https://census.gov/content/dam/Census/library/publications/2017/acs/acs-37.pdf> [accessed 2018-05-30] [[WebCite Cache ID 6znUioC78](#)]

Abbreviations

EHR: electronic health record

FMT: face-mounted technology

GG: Google Glass

HCP: health care provider

HIPAA: Health Insurance Portability and Accountability Act of 1996

HITECH: Health Information Technology for Economic and Clinical Health

PI: primary investigator

Edited by G Eysenbach; submitted 12.04.18; peer-reviewed by Y Huang, C Longhurst, L Downing; comments to author 02.05.18; revised version received 11.05.18; accepted 11.05.18; published 21.06.18.

Please cite as:

Odenheimer S, Goyal D, Jones VG, Rosenblum R, Ho L, Chan AS

Patient Acceptance of Remote Scribing Powered by Google Glass in Outpatient Dermatology: Cross-Sectional Study

J Med Internet Res 2018;20(6):e10762

URL: <http://www.jmir.org/2018/6/e10762/>

doi: [10.2196/10762](https://doi.org/10.2196/10762)

PMID: [29929947](https://pubmed.ncbi.nlm.nih.gov/29929947/)

©Sandra Odenheimer, Deepika Goyal, Veena Goel Jones, Ruth Rosenblum, Lam Ho, Albert S Chan. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 21.06.2018. 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.

Original Paper

Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study

Timothy C Guetterman¹, PhD; Tammy Chang^{1,2}, MD, MPH, MS; Melissa DeJonckheere¹, PhD; Tanmay Basu³, PhD; Elizabeth Scruggs⁴, MD; VG Vinod Vydiswaran^{5,6}, PhD

¹Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

²Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, United States

³Ramakrishna Mission Vivekananda Educational and Research Institute, Belur Math, West Bengal, India

⁴Department of Internal Medicine-Pediatrics, University of Michigan, Ann Arbor, MI, United States

⁵Department of Learning Health Sciences, Medical School, University of Michigan, Ann Arbor, MI, United States

⁶School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:

Timothy C Guetterman, PhD
Department of Family Medicine
University of Michigan
1018 Fuller Street
Ann Arbor, MI, 48154
United States
Phone: 1 734 936 7344
Fax: 1 734 998 7335
Email: tguetter@med.umich.edu

Abstract

Background: Qualitative research methods are increasingly being used across disciplines because of their ability to help investigators understand the perspectives of participants in their own words. However, qualitative analysis is a laborious and resource-intensive process. To achieve depth, researchers are limited to smaller sample sizes when analyzing text data. One potential method to address this concern is natural language processing (NLP). Qualitative text analysis involves researchers reading data, assigning code labels, and iteratively developing findings; NLP has the potential to automate part of this process. Unfortunately, little methodological research has been done to compare automatic coding using NLP techniques and qualitative coding, which is critical to establish the viability of NLP as a useful, rigorous analysis procedure.

Objective: The purpose of this study was to compare the utility of a traditional qualitative text analysis, an NLP analysis, and an augmented approach that combines qualitative and NLP methods.

Methods: We conducted a 2-arm cross-over experiment to compare qualitative and NLP approaches to analyze data generated through 2 text (short message service) message survey questions, one about prescription drugs and the other about police interactions, sent to youth aged 14-24 years. We randomly assigned a question to each of the 2 experienced qualitative analysis teams for independent coding and analysis before receiving NLP results. A third team separately conducted NLP analysis of the same 2 questions. We examined the results of our analyses to compare (1) the similarity of findings derived, (2) the quality of inferences generated, and (3) the time spent in analysis.

Results: The qualitative-only analysis for the drug question (n=58) yielded 4 major findings, whereas the NLP analysis yielded 3 findings that missed contextual elements. The qualitative and NLP-augmented analysis was the most comprehensive. For the police question (n=68), the qualitative-only analysis yielded 4 primary findings and the NLP-only analysis yielded 4 slightly different findings. Again, the augmented qualitative and NLP analysis was the most comprehensive and produced the highest quality inferences, increasing our depth of understanding (ie, details and frequencies). In terms of time, the NLP-only approach was quicker than the qualitative-only approach for the drug (120 vs 270 minutes) and police (40 vs 270 minutes) questions. An approach beginning with qualitative analysis followed by qualitative- or NLP-augmented analysis took longer time than that beginning with NLP for both drug (450 vs 240 minutes) and police (390 vs 220 minutes) questions.

Conclusions: NLP provides both a foundation to code qualitatively more quickly and a method to validate qualitative findings. NLP methods were able to identify major themes found with traditional qualitative analysis but were not useful in identifying nuances. Traditional qualitative text analysis added important details and context.

(*J Med Internet Res* 2018;20(6):e231) doi:[10.2196/jmir.9702](https://doi.org/10.2196/jmir.9702)

KEYWORDS

qualitative research; natural language processing; text data; methodology; coding

Introduction

Background

Qualitative research methods are increasingly being used in social and health-related research because of their ability to help investigators understand nuances, contexts, and the perspectives of participants in their own words. Qualitative data can include images and videos, but text-based data is the most prevalent. The usual sources of text-based data are open-ended survey items, interview or focus group transcripts, and health record documents. However, text-based data needs different approaches for analysis compared with quantitative data to be able to answer complex research questions.

Qualitative text analysis is a process of analyzing qualitative text data, such as open-ended survey responses and interview transcripts. The process generally involves reading the data, assigning qualitative codes as succinct descriptors of meaning to text segments [1-3], and identifying themes that capture the major inferences to address study aims or research questions. Although the demand for qualitative research is high, it is a relatively labor-intensive process as researchers seek an in-depth understanding. Specifically, text data are dense, and researchers often underestimate the amount of data gathered through qualitative methods [3]. For instance, a 30-min interview yields about 10 pages of single-spaced transcribed text.

The overarching goal of qualitative research is often to provide an in-depth and nuanced report, typically by writing themes and a rich, thick description [3,4] that conveys the findings vividly and contextualizes them. To achieve depth, qualitative sample sizes tend to be small to allow researchers to complete the analysis and gain a detailed understanding. Despite the small sample size, the coding process alone takes considerable time due to the need to read all data, consider meaningful codes, assign relevant codes to segments, and discuss and reach agreement with other analysts. When sample sizes are large, a similar analysis of the entire database may become prohibitive in terms of time and effort, leading researchers to focus on a smaller, purposive subsample. However, what if it were possible to analyze larger qualitative databases with sufficient depth while reducing these barriers? Such analysis could leverage the depth of qualitative data with the generalizability of a larger, probabilistic sample.

One potential solution to mitigate the resource constraints of qualitative analysis is natural language processing (NLP). NLP is an area of research and application that explores how computers and automated algorithms can be used to understand and manipulate natural language text to accomplish useful, meaningful tasks [5]. It allows for the analysis of substantially

larger text databases compared with the typical qualitative analysis methods and has been applied to data from electronic health records [6-8], PubMed [9], social media data [10], and text messages (short message service, SMS) [11]. However, research examining the methodological merits of NLP techniques is necessary to further consider NLP as a feasible and high-quality approach for qualitative analysis. Crowston et al [12] reported a case study about the use of NLP for qualitative analysis of messages to understand interactions between software teams. They found that NLP methods performed well in terms of an accurate number of codes identified and a reduction in the amount of text that humans would also have to code; they also increased the speed of coding. Aside from the groundbreaking study of Crowston et al [12], few methodological researchers have examined NLP through the viewpoint of qualitative analysis. Furthermore, we are aware of no other research that has directly compared qualitative analysis with those augmented with automated NLP approaches for text analysis. We conducted a study to compare NLP and qualitative text analysis on the basis of resources used, similarity of findings derived, and the quality of inferences generated. In this study, we applied both NLP and qualitative text analysis methods to a database of short open-ended survey responses from youth gathered via SMS text messages.

Qualitative and Natural Language Processing Analysis of Text Data

Our intention is to in part bridge the gap between qualitative text analysis and NLP analysis by discussing the merits and limitations of the two approaches and how they complement each other. To understand our methods of text analysis more completely, we have provided an overview of qualitative text analysis and NLP approaches. Here, we have discussed our detailed methods of analysis of text data to provide background explanation because in the method section, we focus on how we compared these two approaches. The following section illuminates the two approaches for those who may be unfamiliar with either. As NLP relies on a computer algorithm to categorize free text, we offer substantially more details about how it is performed to make the use of a computer to analyze text more accessible.

Qualitative Text Analysis

Although numerous methods for qualitative text analysis exist, the general approach involves reading the data and assigning codes to text segments [2]. A code is a concise qualitative label (eg, “exercise,” “staying healthy,” and “harmful”) to identify the meaning in a segment of text. For example, “I like to exercise” might be coded by investigators as “exercise.” Applying the qualitative thematic text analysis approach of

Kuckartz [2], investigators identify codes through the analysis based on reading the data. The process of assigning a code label is subjective; however, procedures are available to establish intercoder agreement through a consensus process. For example, 2 or more researchers may code the same text initially and discuss discrepancies, such as using different codes or synonyms that have virtually the same meaning for a code label. In addition, researchers might use sensitizing concepts [13] to identify some codes a priori based on a theoretical or a conceptual model guiding the research. Considerable debate prevails around that practice, which qualitative purists might consider as too postpositivistic or deductive [1].

Analysis continues as researchers refine the codes used, define each code more precisely, and iteratively develop the code book. Researchers then identify themes or categories that represent major findings of the analysis [2]. Identifying themes is a process of examining patterns and similarities among codes and then interrelating the themes. The themes then comprise the major findings of the analysis. In this study, we applied this thematic text analysis approach to analyze the text data gathered from youth.

Natural Language Processing Framework

NLP is a subfield of computer science and linguistics that deals with algorithms, methodologies, and tools to analyze natural language text and studies grammatical, syntactic, and semantic structure of text. In our study, for example, the SMS text messages comprised open-ended short responses by participants of the survey. Depending on the type of questions, the text responses could be a simple “next,” a preferred choice (eg, “yes” or “no”), or a more verbose answer, for example, in response to a “Why?” question. The preferred choice responses were cleaned by removing punctuations and capitalization and normalized by correcting spelling errors. The “Why?” question responses were also processed with these data cleaning steps, after which they were clustered based on similarity of words.

Word similarity can be computed based on the relative distance between words in a hierarchical word ontology known as WordNet. WordNet is a database for the English language that contains words and multiword phrases and organizes nouns, verbs, adjectives, and adverbs into more than 117,000 synonym sets (known as synsets) [14]. A synset is a set of words that have the same meaning (cognitive synonyms). The synsets are further organized into a hierarchy of “is-a” relations (hypernyms and hyponyms), which could be used to compute similarity of a pair of words. For example, the words “exercise” and “workout” are synonyms and “stretching,” “yoga,” and “calisthenics” are hyponyms of “exercise,” whereas “exertion” is the hypernym of “exercise.” Nouns indicating other exertion activities such as “straining” and “pull” are siblings of “exercise” in the WordNet hierarchy (ie, “straining,” “pull,” and “exercise” are all hyponyms of “exertion”). Based on edge distance between appropriate synsets in this tree-like structure, one could consider that exercise and workout are very similar (an edge distance of 0), exercise and yoga are quite similar (an edge distance of 1), whereas exercise and straining are even less similar (an edge distance of 2). In contrast, the words exercise and weight, although they appear to be related, have a large tree

edge distance and are considered not similar—while one is an activity, the other is an artifact. The performance of WordNet to measure the semantic orientations of adjectives has been measured against human judgment and was found to be very effective [15].

Several similarity measures have been proposed that use the edge distance and other factors in computing word-to-word similarity [16,17]. For example, Wei et al [18] used WordNet to find semantic similarity between words instead of content similarity to improve the effectiveness of automatic text clustering. They explored several semantic similarity measures over WordNet, for example, Leacock and Chodorow similarity [19] and Wu-Palmer similarity [20] for finding semantic similarity in the text clustering technique. They found that Wu-Palmer similarity was a better measure to capture the semantic similarity between words in a text clustering application. Wu-Palmer similarity between two words is the path length to the root node from the least common subsumer (LCS) of the two words in WordNet [20]. LCS is the most specific concept a pair of words share as an ancestor. For example, in the sample WordNet hierarchy given above, the LCS for “yoga” and “pull” is “exercise,” since it is the most specific ancestor of “yoga” and “pull.” In the Wu-Palmer similarity measure, the path length is normalized by dividing the sum of the path lengths from the individual words to the root word [20]. The Wu-Palmer similarity has been used to find semantic relatedness between concepts in many genres of text, including biomedical texts [17,21,22].

In this study, we used the Wu-Palmer similarity measure [23] to find semantic similarity between words. Researchers use the word similarity measures to identify clusters or concepts based on synonyms and very similar word pairs (eg, Wu-Palmer similarity >0.9). Assigning similar words and phrases to one concept group is similar to using a code label in qualitative text analysis. However, through the use of NLP, this process is automated. The detailed description of the proposed NLP framework to find different clusters of similar words from the short messages is as follows:

1. All nouns and pronouns were extracted from the short message texts and converted to lower case. We did not consider the other parts of speech because they were found to have little contribution toward identifying concepts from the SMS text messages. The same was observed empirically.
2. A vocabulary was created with unique nouns and adjectives (ie, multiple occurrences of a word are discarded).
3. Synsets of each word in the vocabulary are generated using WordNet. Note that a word may have more than one synset as described earlier.
4. A pair of words were grouped together if the Wu-Palmer similarity between any pair of synsets, one generated from the first word and the other from the second word, is >0.9. It may be noted that the Wu-Palmer similarity score ranges from 0 to 1, both inclusive, and 1 indicates the highest similarity. This step was repeated for all pairs of words in the vocabulary.
5. Derivationally related forms of each word in the word pairs were generated, for example, “honest” is derivationally related to “honesty” as generated by WordNet.

6. The most similar pair of words and their derivational forms were combined to create a cluster. This led to several word clusters to be created from the vocabulary.
7. A pair of clusters was merged, if they had at least 50% common members (ie, words). The process continued until no more merges could take place. The method terminated automatically upon satisfying the given condition and generated the final clusters of words. These clusters indicated different senses and semantic meanings present in the given database of SMS text messages.

Study Background and Context

To conduct this research, we used SMS text message survey data gathered from the MyVoice study [11]. MyVoice is a national SMS text message poll of youth, in which 3-5 SMS text message questions were sent to a national sample of participants aged 14-24 years on a weekly basis. Questions were typically open-ended, and topics range from current events to specific health concerns. The purpose of this study was to describe the merits of integrating NLP and manual qualitative data analysis based on our direct comparison of the 2 approaches applied to a dataset. Our goal was not to prefer one approach over the other but to critically reflect on the value and limitations of each approach alone in addition to the integration of the two. This article provides guidance to researchers in deciding to use NLP techniques and manual qualitative text analysis.

Methods

Overview

To compare NLP methods and qualitative analysis, we conducted a modified 2-arm cross-over experiment. Our primary outcomes were comparing the similarity of thematic findings and time taken to analyze based on person hours. We were also particularly interested in documenting the process for each arm. A different, experienced coding team conducted the analysis for each approach. Two qualitative analysis teams (each with 2 of the authors, MD, TC, ES, and TG) independently coded and generated findings from two different datasets, one focused on opinions about prescription drug use and the other focused on interactions with police. Our analysis followed the qualitative text analysis process as noted previously. The 2 teams used MAXQDA 12 qualitative software (VERBI GmbH; Berlin, Germany) to facilitate the analysis process. We randomly assigned the datasets to each team to begin the analysis. Simultaneously, a third team (VV and TB) independently conducted NLP analysis with each dataset. We computed word similarity over all pairs of nouns in responses to the 2 survey questions. Using the criterion of Wu-Palmer similarity >0.9 , we then grouped survey responses that contained synonyms and very similar word pairs into a concept group.

The qualitative analysis teams received the NLP results for the dataset they had already coded to augment their analysis and reconsider their conclusions. The cross-over then occurred as the 2 qualitative teams received the NLP data *first* from the other dataset they had no prior exposure to. The reason for having a cross-over design was to account for team effects and different question set effects. The teams proceeded with qualitative analysis after reviewing the NLP results and then

developed conclusions. In summary, we had the following 2 cross-team comparison conditions: (1) NLP-only followed by augmented qualitative or NLP analysis and (2) qualitative analysis-only followed by an augmented qualitative or NLP analysis.

Data Sources

We used a subset of open-ended text data identified from a larger project and applied each of the 2 coding approaches. Open-ended questions enabled the participants to construct a response without being constrained by response options and allowed them to explain their response. The data consisted of responses to an SMS text message-based survey with young adults aged 14-24 years (mean 18 years). We focused on 2 questions, one about drugs and the other about experiences with police, respectively. The drug question was “What do you think is more dangerous: taking a prescription drug that is not yours or an illegal drug? Why?” The police question was “What have your experiences with police been like?”

Analysis

To address our methodological purpose, we carefully examined the results of our data analysis. Our goals for this methodological analysis were to understand (1) how the thematic findings differ when beginning with qualitative-only coding followed by the qualitative/NLP augmented coding compared with beginning with NLP followed by NLP/qualitative augmented coding; (2) how the process led to additions, deletions, changes in codes or changes in the definitions of codes; and (3) how the conclusions derived differ between the 2 teams. We also calculated the percent agreement between the 2 researchers in each team in coding of the same dataset to check for consistency. We found acceptable agreement [1] with 62 of 88 (70%) coded segment agreements and 72 of 84 (87%) agreements, respectively, for the 2 teams. More importantly, each team engaged in the process of discussing codes line-by-line and reconciling differences through consensus. In addition, the teams recorded process notes, including the time spent in each analysis. Finally, we compared the inferences generated by each standalone approach and a qualitative/NLP augmented approach.

Results

Overview

We have briefly summarized the findings of our analysis of text data using the following different approaches: qualitative coding, qualitative followed by NLP-augmented coding, NLP-only coding, and NLP followed by qualitative-augmented coding. The findings reported in tables are summaries from the teams after their independent analysis and before any discussion between teams. After reviewing the coding and conclusions, we have discussed the results of our methodological analysis of this process.

Comparing Qualitative, Natural Language Processing, and Augmented Coding Approaches for Text Analysis

In total, 84 individuals answered at least one of the 2 sets of questions; 58 answered the drug question and 68 answered the police question, which were the focus of our analysis. The

demographic questions were not required, and demographics were available for 66 of the 84 individuals (Table 1).

Drug Question (What do You Think is More Dangerous: Taking a Prescription Drug That is Not Yours or an Illegal Drug? Why?)

The thematic findings of the 2 teams were relatively consistent (Tables 2 and 3). The qualitative-only analysis for this question yielded 4 major findings. The perspectives of youth respondents included that prescription and illegal drugs are equally dangerous and that the degree of danger depends on the situation. Other youth explained that either drug could be more dangerous, depending on the intention of use and whether the drug was prescribed. The NLP-only findings were somewhat similar but also driven by word frequencies. The major NLP-only findings were that (1) youth were divided as to which was more dangerous, (2) more noted that illegal drugs were more dangerous, but (3) 11 youth responded that the danger depends on the situation. However, the context was missing from the NLP-only analysis. For example, some wrote about side effects and harm to the body, but we could not determine from the NLP results whether these comments referred to illegal drugs or prescription drugs that are not theirs. Finally, the qualitative followed by NLP-augmented results were more comprehensive. For example, a qualitative-only thematic finding was that either prescription or illegal drugs could be more dangerous and stigma was an issue. However, when we examined the NLP-generated data, we recognized that “stigma” was incomplete and actually more complex. We added that stigma was related to the discreetness of obtaining prescription drugs compared with buying street drugs.

The 2 teams—one began with qualitative-only followed by NLP-augmented coding and the other began with NLP-only followed by qualitative-augmented coding—reached very similar conclusions. From the augmented analysis, both teams added to findings about the legality of the drug, whether the ingredients are known, and Food and Drug Administration approval. For one team, these new points rose to the level of major conclusions, while the other tended to add them as details, perhaps reflecting more of a stylistic difference between teams. However, for both teams, the legality issue was clearly a major thematic finding that the augmented qualitative and NLP approach added.

Police Question (What Have Your Experiences With Police Been Like?)

Tables 4 and 5 compare thematic findings for the 2 teams and different approaches to coding the police data. Again, findings were relatively consistent. The qualitative-only results for this question yielded the following 4 primary findings: (1) about one-third of youth had no real interaction with police to comment on, (2) the majority who had interaction had a positive experience, (3) some noted concerns about racism, and (4) a small group described the importance of police to maintain public safety. The NLP-only analysis yielded the following 4 slightly different findings: (1) some had few experiences, (2) several youth had positive experiences, though some were “bad,” (3) individual relationships and characteristics affected

experiences, and (4) youth could point to specific “situations” with police.

As with the drug question, the qualitative- and NLP-augmented results for the police question were the most comprehensive. Based on the additional review of the NLP police data, “good/positive” and “bad” commentary were most frequent, which reinforces the conclusion from the qualitative-only phase. The NLP data did reflect several occasional references to race (white) or gender (women). Furthermore, multiple mentions of police force as respectful and assurance of security lend credence to the qualitative conclusion that police are needed to maintain public safety. In general, the NLP data were unable to pick up on nuance (eg, that some individuals described both positive and negative experiences at the same time) but ultimately reinforced the basic conclusions noted by the qualitative team. Furthermore, NLP data revealed that gender, in addition to race, was related to participant’s experiences and feelings surrounding the police. In brief, the augmented NLP and qualitative analysis increased the depth of understanding (ie, details and frequencies), but the overall findings were quite similar.

Methodological Comparison of Methods

Our analysis revealed 3 methodological insights related to the process of using NLP data, how NLP and qualitative-augmented analysis tended to yield more information, and observations about how the ordering of analysis affected the process. First, we developed a process to use and review NLP data, which was somewhat different from our qualitative coding process. In using NLP, we first reviewed the entire file sent by the NLP team. It consisted of a spreadsheet of words, synonyms in the dataset, and relative frequencies (see Table 6 for an example of NLP output).

We looked for response phrases that simply repeated words in the question, typically listed with high frequency in the list, and discounted those. We reviewed all words and examples, highlighting those that stood out. For analyses that began with qualitative coding, followed by NLP, we were particularly keen to find NLP concepts that we missed qualitatively. At this point, it was helpful to create a conceptual model by drawing a map relating codes. If we had findings from an initial qualitative analysis, we then integrated NLP data with what we knew from the qualitative findings. Finally, we noted a potential difference in achieving data saturation between NLP and qualitative analysis. Data saturation is the point at which themes are sufficiently complex and collecting additional data is not adding to findings [24]. In NLP analysis, the point of saturation tended to occur while we reviewed more frequent words and before we reached less frequent words. In NLP, saturation seemed to be dependent on the frequency of ideas rather than the complexity of themes. In contrast, in qualitative coding, we reached saturation at some point while reading through the data. We noted that with qualitative coding, saturation was more dependent on the order in which text responses appeared throughout this study, and we continued to analyze all responses.

The different approaches yielded similar thematic findings, but the NLP- or qualitative-augmented coding produced more information than the qualitative-only or NLP-only approaches.

Table 1. Participant demographic information.

Variable	Drug response (n=48)	Police response (n=59)	Drug or police response (n=66)
Age, mean (SD)	18.5 (2.2)	18.3 (2.5)	18.3 (2.4)
Gender, n (%)			
Female	28 (58.3)	33 (55.9)	37 (56.1)
Male	18 (37.5)	25 (42.4)	27 (40.9)
Other	2 (4.2)	1 (1.7)	2 (3.0)
Race, n (%)			
White	26 (54.2)	36 (61.0)	38 (57.6)
Black	8 (16.7)	9 (15.3)	11 (16.7)
Asian	7 (14.6)	7 (11.9)	8 (12.1)
Other (including multiracial)	7 (14.6)	7 (11.9)	9 (13.6)
Hispanic ^a	0 (0.0)	3 (6.7)	3 (6.4)
Education, n (%)			
<High school	19 (39.6)	28 (47.5)	31 (47.0)
High school grade	7 (14.6)	5 (8.5)	7 (10.6)
Some college	17 (35.4)	19 (32.2)	20 (30.3)
College grade (BA+)	5 (10.4)	7 (11.9)	8 (12.1)
Parent education^a, n (%)			
High school or less	1 (3.5)	2 (4.4)	2 (4.3)
Some college or 2-year degree	4 (13.8)	5 (11.1)	5 (10.6)
BA but less than Masters	7 (24.1)	8 (17.8)	9 (19.2)
Masters but less than PhD	9 (31.0)	18 (40.0)	19 (40.4)
PhD	8 (27.6)	12 (26.7)	12 (25.5)
Primarily living with^a, n (%)			
Parents	22 (75.9)	33 (73.3)	35 (74.5)
Dorm	0 (0.0)	1 (2.2)	1 (2.1)
Sharing an apartment with other people	7 (24.1)	8 (17.8)	8 (17.0)
Other	0 (0.0)	3 (6.7)	3 (6.4)
Family size^a, n (%)			
1-3	6 (20.7)	9 (20.0)	10 (21.3)
4-6	17 (58.6)	29 (64.4)	29 (61.7)
7-10	5 (17.2)	7 (15.6)	7 (14.9)
11+	1 (3.5)	0 (0.0)	1 (2.1)
Parent's marital status^a, n (%)			
Married or together	21 (72.4)	36 (80.0)	36 (76.6)
Divorced or separated	7 (24.1)	7 (15.6)	9 (19.2)
Other (widowed, unsure)	1 (3.5)	2 (4.4)	2 (4.3)

^aSample sizes are as follows: Drug response (n=29); police response (n=45); and drug or police response (n=47). Participants were not required to provide demographic information, so the n for respective demographic questions in this table is lower than the total number of participants. Because some responded to both questions, we have 3 columns of demographic information. There are fewer responses for ethnicity, parent's education, primary living situation, family size, and parent's marital status due to those questions not being asked to the subset of individuals who had demographics requested twice. The third column displays data for those who responded to at least one question.

Table 2. Comparison of findings derived from qualitative-only and qualitative followed by natural language processing-augmented approaches to coding for the drug question (n=58). Key aspects of each finding are italicized.

Theme	Qualitative only ^a	Qualitative (natural language processing augmented) ^b
Prescription drugs and illegal drugs	Prescription drugs and illegal drugs are <i>equally dangerous</i> because both are serious, could harm you, and are illegal.	Prescription drugs and illegal drugs are <i>equally dangerous</i> because both are serious, could harm you, and are illegal.
Danger	<i>Danger depends</i> on the situation, the amount of drug taken, type of drug, whether or not it was prescribed to you. Distinction between a medical danger versus a legal danger.	<i>Danger depends</i> on the situation, the amount of drug taken, type of drug, whether or not it was prescribed to you. Distinction between a medical danger versus a legal danger.
Respondent chose either/or	Either Rx ^c drugs or illegal drugs could be more dangerous based on addictiveness, accessibility, prevalence, overdose, or danger. Side effects: known or unknown. Stigma.	Either Rx drugs or illegal drugs could be more dangerous based on addictiveness, accessibility, prevalence, overdose, or danger. Side effects: known or unknown. Stigma of getting drugs off the street versus discreetness of “popping” Rx pills.
Intention or appropriateness	Is the drug safe for everyone or unsafe for some people depending on whether prescription was prescribed to you.	Is the drug safe for everyone or unsafe for some people depending on whether prescription was prescribed to you.
Ingredients	—	What the drugs consisted of. Mixing Rx versus unknown contents of street drugs versus taking Rx you do not know what they are.
Legality	—	Legal more prominent; the <i>legality of using other prescriptions and the legality of street drugs</i> and what that meant about street drug safety or regulations.
Government involvement	—	<i>Food and Drug Administration approval</i> and regulations versus “illegal” for a reason.
Harm to body	—	Mortality was often mentioned (overdose, “something that could kill you”).
Specific drugs mentioned	—	“Weed, heroin, cocaine, meth, alcohol, smoking” and using them for comparisons for safety and side effects or addictiveness.

^aTime required (person min): 270.^bTime required (person min): 180.^cRx: prescription medication.

Table 3. Comparison of findings derived from natural language processing-only and natural language processing followed by qualitative-augmented approaches to coding for the drug question (n=58). Key aspects of each finding are italicized.

Theme	Natural language processing only ^a	Natural language processing (qualitative augmented) ^a
Prescription drugs and illegal drugs	Respondents seemed <i>divided</i> between whether illegal drugs or medicines were more dangerous.	Of the 58 respondents, 24 noted that <i>illegal drugs were more dangerous</i> , 15 thought both were <i>equally dangerous</i> , and 11 answered <i>prescription drugs</i> . 10 argued that it depended on the context.
Danger	11 respondents noted that it <i>depends</i> or similar as to what is more dangerous. Some noted it depends on the reason for using either.	For some, the question of danger <i>depends</i> on who owns the drug, the situation, the type of drug, and the ease of access. Several felt that the answer depended on what type of drug (either illegal or prescription), how much, and for what. A few felt that certain illegal drugs (eg, marijuana) were less dangerous than legal drugs (eg, alcohol).
Government involvement	It seemed that more wrote that <i>illegal drugs were more dangerous</i> .	<i>Illegal drugs may be more dangerous</i> because they are not medically cleared, federally regulated, nor approved by the Food and Drug Administration. They may contain traces of other substances. Respondents felt uncertain about where the illegal drug might come from or what it really contained.
Prescription drugs	—	<i>Prescription drugs may be more dangerous</i> because they are easy to access and not illegal, so more people may feel comfortable taking them.
Intention or appropriateness	—	Several respondents noted that taking a prescription drug that is not yours is <i>illegal</i> too.
Side effects and harm to the body	Respondents wrote about <i>side effects</i> (n=5) and <i>harm to the body</i> (n=6), but we cannot determine whether it referred to prescription medicine or illegal drugs.	—

^aTime required (person min): 120.**Table 4.** Comparison of findings derived from qualitative-only and qualitative followed by natural language processing-augmented approaches to coding for the police question (n=68). Key aspects of each finding are italicized.

Theme	Qualitative only ^a	Qualitative (natural language processing augmented) ^b
Interactions	For those who had interaction, the majority of participants described <i>interactions as positive</i> (eg, “pleasant,” “positive,” “decent”). Others reported negative interactions (eg, “bad,” “not so good”). Some of the individuals who described positive experiences also gave negative experiences, such as describing a sense of fear (“Positive but I am always scared interacting with them”).	For those who had interaction, the majority of participants described <i>interactions as positive</i> (eg, “pleasant,” “positive,” “decent”). Others reported negative interactions (eg, “bad,” “not so good”). Few described the interactions as negative or bad. Some of the individuals who described positive experiences also gave negative experiences, such as describing a sense of fear (“Positive but I am always scared interacting with them”).
Racism/gender differences	Some individuals wrote about major concerns with <i>racism among police</i> .	Some individuals wrote about major concerns with <i>racism among police</i> . Several mentioned <i>gender differences</i> in how they were treated.
Public safety	A small group described police as <i>needed in order to maintain public safety</i> .	A small group described police as <i>needed in order to maintain public safety</i> .

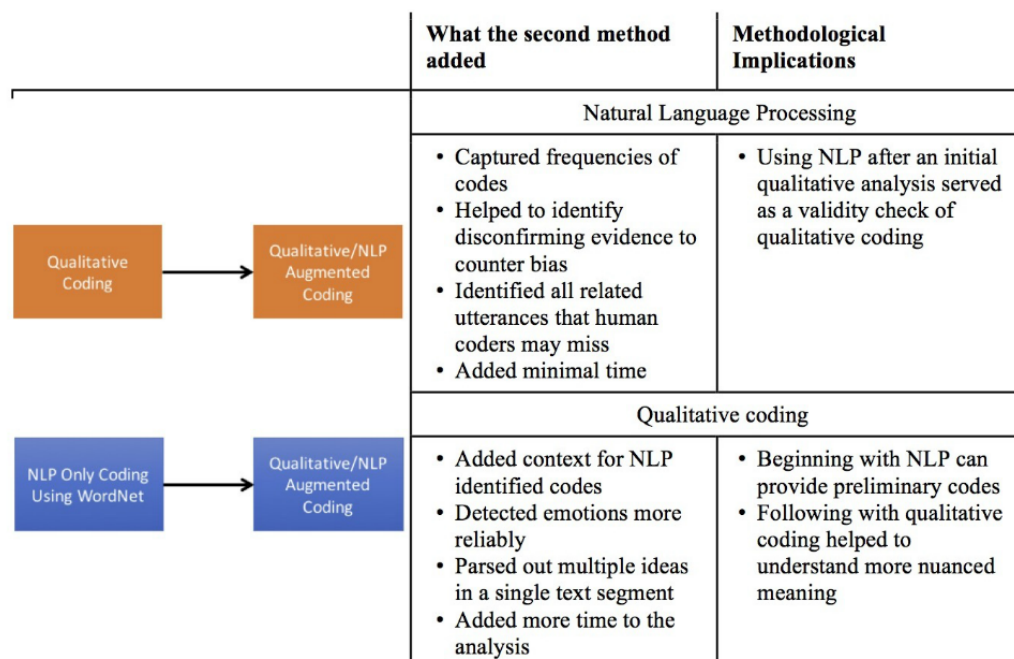
^aTime required (person min): 270.^bTime required (person min): 120.

Table 5. Comparison of findings derived from different approaches to coding for the police question (n=68). Key aspects of each finding are italicized.

Theme	Natural language processing only ^a	Natural language processing (qualitative augmented) ^b
Number of experiences	Some youth reported <i>few experiences</i> with police.	Some youth reported <i>few or no experiences</i> with police (n=17). Others reported “inconsequential” interaction (n=2).
Interaction	Many youths reported good or <i>positive experiences</i> with police. Other words in this theme include friendly, respectful, pleasant, and helpful. Others reported “ <i>bad</i> ” <i>experiences</i> , unpleasant, aggressive, and mean (less frequency words).	Many youths reported good or <i>positive experiences</i> with police (n=32). Other words in this theme include friendly, respectful, pleasant, and helpful. Others reported “ <i>bad</i> ” <i>experiences</i> , unpleasant, aggressive, mean (n=7). Some reported good and bad (n=2).
Situations	Youth can point to “ <i>situations</i> ,” “ <i>moments</i> ,” and “ <i>times</i> ” <i>when they interacted with police</i> . Some of these situations were at concerts, sporting events, or at their schools.	Youth can point to “ <i>situations</i> ,” “ <i>moments</i> ,” and “ <i>times</i> ” <i>that they interacted with police</i> . Some of these situations were at concerts, sporting events, or at their schools.
Avoidance	—	<i>Avoid situations with police</i> , makes people nervous, excessive force, seem mean, cause fear.
Individual characteristics	<i>Individual characteristics and relationships</i> come into play: white, woman, young, friends, or parents.	<i>Individual characteristics and relationships</i> come into play: white, woman, young, friends, or parents. Specifically, race (n=5) and gender (n=2).

^aTime required (person min): 40.^bTime required (person min): 180.**Table 6.** Example natural language processing output from the drug dataset.

Code word	Frequency	Similar words	Data segments
Medicines	98	Medication, medicine, medicate, prescription, drug	“an illegal drug” “true for specific drugs” “addictive than prescription medication” “whereas prescription medications are legal” “most dangerous drugs out there” “because prescription drugs are specifically”
Illegal	48	—	“is the illegal drug yours” “doing an illegal drug since” “compared to illegal drugs” “both are illegal in my”
Prescription	26	Prescriptions	“addictive than prescription medication” “some prescription medicines are” “least the prescription medication is” “very dangerous prescription medicines but”
Dangerous	14	—	“are equally dangerous” “is more dangerous because it” “is physically/mentally dangerous and illegal”
Depends	11	—	“it depends on which”
More	9	—	“can be more addictive than” “medicines are more dangerous than” “a lot more overdoses than” “drug is more dangerous than” “can be more powerful and”

Figure 1. Summary of methodological findings. NLP: natural language processing.

Qualitative coding was beneficial in our analysis because it added critical contextual understanding that helped to interpret responses. For example, in the police data, the qualitative coding helped the analysts to identify emotions, such as fear, that were not evident in the NLP analysis. Qualitative coding also helped us to identify full quotes that we might include in a report of findings. On the other hand, NLP was beneficial in terms of concept and code frequencies. It can add a validity check, which is concerned with the accuracy of qualitative findings [3]. Although we were able to determine the relative frequency (eg, quantifying comments by “all,” “most,” “some,” “few,” etc) of codes through qualitative coding, the process of finding NLP word frequencies was more efficient. Therefore, the augmented approach of combining both NLP and qualitative approaches leveraged the strengths of each. The NLP- or qualitative-augmented coding led to more conclusions and more depth within conclusions.

Our process notes yielded further insight into the ordering of augmenting one method with another. As anticipated, the NLP-only approach was quickest at 120 minutes for the drug question and 40 minutes for the police question, whereas the qualitative-only approach was longest at 270 minutes for both questions. However, we found a difference, an approach beginning with qualitative analysis followed by qualitative- or NLP-augmented analysis took longer time than that beginning with NLP for both drug (450 vs 240 minutes) and police questions (390 vs 220 minutes). As noted in Figure 1, including the second method added different benefits depending on the order. Our subjective preference—from both teams—was to begin with qualitative coding and then augment it with NLP afterward, which was unexpected. Prior to the study, we anticipated that it would be more helpful to begin with NLP, but all 4 analysts felt that it was more difficult to start with NLP word frequencies because we were missing the context of the codes. On the other hand, beginning with qualitative coding,

we gained a stronger sense of the data and found it helpful to add NLP as a validation strategy. Despite our subjective preference, we did not find substantive differences in the findings when beginning with one method versus the other.

Discussion

Principal Findings

In brief, our primary conclusion is that NLP is a viable method for coding text data, and it is particularly useful when augmenting manual qualitative coding. Each method alone yielded relatively consistent results, but the combination (ie, augmented) analysis clearly added more detail across the 2 questions, a finding consistent with the study of Crowston et al [12]. The final, augmented findings from the 2 teams were relatively similar, which suggests that we are not merely detecting differences among teams but also obtained insights into the ordering of the augmenting of approaches. We observed a similar pattern across questions, in which the augmented analysis yielded the most complete details. Finally, though we noted that we achieved data saturation with both NLP and qualitative approaches, saturation may be a nonissue for NLP of data gathered through probabilistic sampling to achieve generalization.

Limitations

Several limitations of this study and the approach must be noted. First, this research was based on responses to 2 short open-ended survey questions. Our findings may not transfer to longer responses such as interview transcripts because NLP’s challenge with nuance and contextual responses might increase. Regarding time efficiency, the times we recorded were based on a relatively small text database of short snippets. Although the augmented approach took more time in our comparison, we anticipate that with larger datasets, the efficiency of an NLP-augmented approach will prevail and take less time. Further research is

needed to compare the approaches with interviews and other text data. Second, it is possible that some of our findings actually reflect differences among the 2 qualitative teams. However, we attempted to counterbalance those concerns through the cross-over design. Third, each method has its own limitations. Qualitative coding generally eschews frequencies, whereas NLP privileges that concept. Thus, a natural question is whether the two are comparable. Furthermore, qualitative coding is limited to a purposeful, rather than a probabilistic sample that is generalizable. In addition, WordNet is carefully curated and does not reflect current events, slang, or nonsemantic use of terms. For this reason, it does have some limitations for analysis of adolescent language and SMS text message data. Other limitations are related to our particular simple NLP approach. Other more complex approaches, such as sentiment analysis, may capture the context better. A future study can directly compare NLP output with qualitative analysis. Nevertheless, we believe that the limitations of each approach add further credence to the idea of an augmented qualitative and NLP approach.

NLP Coding and Context

It is important to note that NLP helped quantify, organize, and categorize responses quickly—providing an accurate overview of the major themes. NLP by itself can miss the context of what was being said, especially when emotion was involved (eg, “fear” of police), similar to the findings of Crowston et al [12]. We recommend a qualitative-augmented approach to understand the context. Practically, it might involve taking a subsample of a larger NLP analysis to augment with qualitative coding to understand the context.

Moreover, NLP has the potential to add value to manual qualitative analysis procedures. NLP does capture the relative frequency of particular codes and words, which can add important information about the weight of what was communicated to complement the context provided by qualitative methods. In general, NLP can provide a validity check of qualitative findings by providing a second method to triangulate findings or by helping the researcher to systematically search for disconfirming evidence. After establishing preliminary themes, qualitative researchers might use NLP to engage in this qualitative validation procedure to look for contrary data, exceptions, and alternate perspectives in examining whether evidence supports or disconfirms findings as reported by Creswell and Miller [25]. NLP offers the ability to search for all text and utterances that human readers may overlook. In addition to these potential benefits, using NLP itself adds minimal time.

Nature of Data Analyzed

We noticed another contextual issue based on the nature of the question analyzed. The drug question required the participants

to compare (ie, an illegal drug or a prescription drug that is not yours). NLP alone is not well equipped to understand the context of comments and which of the 2 types of drugs participants were commenting on. Alternatively, the question could have been separated into 2 items to facilitate an NLP-only analysis. The police dataset had similar nuances with added qualifiers provided by some. Therefore, we recommend any comparative questions be analyzed through an NLP and qualitative augmented approach.

Although our data were from a larger study of adolescent perspectives on policy and health topics, the methods we used are agnostic of the topic domain. The augmented approach can be applied to a wide variety of health, medical, and topics from other domains. Furthermore, it is applicable to other data sources, including traditional survey responses, social media snippets, or documents. Both the questions we analyzed consisted of relatively short text segments and similar sample sizes of 58 and 68, respectively. Although we found no time savings with the NLP and qualitative augmented approach, as the sample size increases, the efficiencies of NLP will likely be clearer. One way to achieve time efficiency is by applying NLP to the entire dataset and qualitatively coding a smaller subsample. The implication for future research and development is to advance how the method can be applied to longer text, such as qualitative semistructured interview transcripts and more complex responses rather than the domain of inquiry, as this technology should be applicable to any text data. Therefore, we urge further research applying an NLP and qualitative augmented approach to interview transcripts and other forms of text data.

Conclusions

NLP methods were able to identify major themes found with traditional qualitative analysis, but the approach was not useful at identifying nuances. Subsequent traditional qualitative text analysis added important details and context. Researchers using NLP techniques might want to consider analyzing even a portion of data with qualitative text analysis to ensure that important context is not missed.

NLP provides both a foundation to code qualitatively more quickly and a method of validation. NLP can help researchers conduct qualitative analysis more quickly because a coding rubric may become apparent in the NLP output. Findings from NLP by itself may be appropriate for analysis that must be done rapidly for a focused question (ie, policy questions, formative program evaluation could improve processes in real-time, community needs assessments). Finally, NLP can add a validity check of qualitative findings by adding frequency counts and larger sample sizes to the conclusions drawn from qualitative analysis when needed to address the research questions.

Acknowledgments

We wish to acknowledge Melissa Plegue for her assistance in compiling demographic data of respondents. In addition, we are grateful to Lilly Pritula for editorial assistance in preparing this manuscript.

Conflicts of Interest

None declared.

References

1. Miles M, Huberman A, Saldaña J. Qualitative data analysis: A methods sourcebook. Thousand Oaks, CA: SAGE; 2014.
2. Kuckartz U. Qualitative text analysis: A guide to methods, practice and using software. London: Sage; 2014.
3. Creswell JW, Poth CN. Qualitative inquiry and research design: Choosing among five approaches. Thousand Oaks, CA: Sage; 2018.
4. Merriam SB, Tisdell EJ. Qualitative research: A guide to design and implementation. 4th edition. San Francisco, CA: John Wiley & Sons; 2016.
5. Chowdhury GG. Natural language processing. *Ann Rev Info Sci Tech* 2005 Jan 31;37(1):51-89. [doi: [10.1002/aris.1440370103](https://doi.org/10.1002/aris.1440370103)]
6. Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, et al. Natural Language Processing-Enabled and Conventional Data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study. *JMIR Med Inform* 2016 Oct 28;4(4):e35 [FREE Full text] [doi: [10.2196/medinform.5544](https://doi.org/10.2196/medinform.5544)] [Medline: [27793791](https://pubmed.ncbi.nlm.nih.gov/27793791/)]
7. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. *J Biomed Inform* 2018 Jan;77:34-49 [FREE Full text] [doi: [10.1016/j.jbi.2017.11.011](https://doi.org/10.1016/j.jbi.2017.11.011)] [Medline: [29162496](https://pubmed.ncbi.nlm.nih.gov/29162496/)]
8. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based Real-Time Case Finding for the Population Health Management of Patients With Diabetes Mellitus: A Prospective Validation of the Natural Language Processing-Based Algorithm With Statewide Electronic Medical Records. *JMIR Med Inform* 2016 Nov 11;4(4):e37 [FREE Full text] [doi: [10.2196/medinform.6328](https://doi.org/10.2196/medinform.6328)] [Medline: [27836816](https://pubmed.ncbi.nlm.nih.gov/27836816/)]
9. Alvaro N, Miyao Y, Collier N. TwiMed: Twitter and PubMed Comparable Corpus of Drugs, Diseases, Symptoms, and Their Relations. *JMIR Public Health Surveill* 2017 May 03;3(2):e24 [FREE Full text] [doi: [10.2196/publichealth.6396](https://doi.org/10.2196/publichealth.6396)] [Medline: [28468748](https://pubmed.ncbi.nlm.nih.gov/28468748/)]
10. Park A, Hartzler AL, Huh J, McDonald DW, Pratt W. Automatically Detecting Failures in Natural Language Processing Tools for Online Community Text. *J Med Internet Res* 2015 Aug 31;17(8):e212 [FREE Full text] [doi: [10.2196/jmir.4612](https://doi.org/10.2196/jmir.4612)] [Medline: [26323337](https://pubmed.ncbi.nlm.nih.gov/26323337/)]
11. DeJonckheere M, Nichols LP, Moniz MH, Sonnevile KR, Vydiswaran VGV, Zhao X, et al. MyVoice National Text Message Survey of Youth Aged 14 to 24 Years: Study Protocol. *JMIR Res Protoc* 2017 Dec 11;6(12):e247 [FREE Full text] [doi: [10.2196/resprot.8502](https://doi.org/10.2196/resprot.8502)] [Medline: [29229587](https://pubmed.ncbi.nlm.nih.gov/29229587/)]
12. Crowston K, Allen EE, Heckman R. Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology* 2012 Nov;15(6):523-543. [doi: [10.1080/13645579.2011.625764](https://doi.org/10.1080/13645579.2011.625764)]
13. Padgett DK. Qualitative and mixed methods in public health. Thousand Oaks, CA: Sage; 2012.
14. Miller GA. WordNet: a lexical database for English. *Commun. ACM* 1995 Nov;38(11):39-41. [doi: [10.1145/219717.219748](https://doi.org/10.1145/219717.219748)]
15. Kamps J, Marx M, Mokken R, De RM. Using WordNet to Measure Semantic Orientations of Adjectives. 2004 Presented at: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC); 2004; Lisbon, Portugal p. 1115-1118.
16. Pedersen T, Patwardhan S, Michelizzi J. WordNet Similarity: measuring the relatedness of concepts. In: *Demonstration Papers at HLT-NAACL*. Boston, Massachusetts: Association for Computational Linguistics; 2004.
17. Do Q, Roth D, Sammons M, Tu Y, Vydiswaran V. Robust, light-weight approaches to compute lexical similarity. *Computer Science Research and Technical Reports*, University of Illinois 2009;9 [FREE Full text]
18. Wei T, Lu Y, Chang H, Zhou Q, Bao X. A semantic approach for text clustering using WordNet and lexical chains. *Expert Systems with Applications* 2015 Mar;42(4):2264-2275. [doi: [10.1016/j.eswa.2014.10.023](https://doi.org/10.1016/j.eswa.2014.10.023)]
19. Leacock C, Chodorow M. Combining local contextWordNet similarity for word sense identification. In: *WordNet: An electronic lexical database*. Edited by Fellbaum C. Cambridge, MA: MIT Press; 1998:A.
20. Wu Z, Palmer M. Verbs semanticslexical selection. 1994 Presented at: Proceedings of the 32nd annual meeting on Association for Computational Linguistics; 1994; Las Cruces, NM p. 133-138.
21. McInnes BT, Pedersen T. Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. *J Biomed Inform* 2013 Dec;46(6):1116-1124 [FREE Full text] [doi: [10.1016/j.jbi.2013.08.008](https://doi.org/10.1016/j.jbi.2013.08.008)] [Medline: [24012881](https://pubmed.ncbi.nlm.nih.gov/24012881/)]
22. McInnes BT, Pedersen T, Liu Y, Melton GB, Pakhomov SV. Knowledge-based method for determining the meaning of ambiguous biomedical terms using information content measures of similarity. *AMIA Annu Symp Proc* 2011;2011:895-904 [FREE Full text] [Medline: [22195148](https://pubmed.ncbi.nlm.nih.gov/22195148/)]
23. MAXQDA (computer program). Berlin, Germany: VERBI GmbH; 2017. URL: <https://www.maxqda.com/>
24. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? *Field Methods* 2016 Jul 21;18(1):59-82. [doi: [10.1177/1525822X05279903](https://doi.org/10.1177/1525822X05279903)]
25. Creswell JW, Miller DL. Determining Validity in Qualitative Inquiry. *Theory Into Practice* 2000 Aug;39(3):124-130. [doi: [10.1207/s15430421tip3903_2](https://doi.org/10.1207/s15430421tip3903_2)]

Abbreviations

NLP: natural language processing

LCS: least common subsumer

SMS: short message service

Edited by G Eysenbach; submitted 20.12.17; peer-reviewed by B Jin, S Kim, Y Wang; comments to author 20.03.18; revised version received 14.05.18; accepted 15.05.18; published 29.06.18.

Please cite as:

Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VGV

Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study

J Med Internet Res 2018;20(6):e231

URL: <http://www.jmir.org/2018/6/e231/>

doi: [10.2196/jmir.9702](https://doi.org/10.2196/jmir.9702)

PMID: [29959110](https://pubmed.ncbi.nlm.nih.gov/29959110/)

©Timothy C Guetterman, Tammy Chang, Melissa DeJonckheere, Tanmay Basu, Elizabeth Scruggs, VG Vinod Vydiswaran. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.06.2018. 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.

Review

Web-Based Versus Usual Care and Other Formats of Decision Aids to Support Prostate Cancer Screening Decisions: Systematic Review and Meta-Analysis

Sofia Baptista^{1,2}, MD, MSc; Elvira Teles Sampaio², MD, MSc; Bruno Heleno³, MD, PhD; Luís Filipe Azevedo^{1,4}, MD, PhD; Carlos Martins^{1,4}, MD, PhD

¹Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal

²Serpa Pinto Family Health Unit, Agrupamento de Centros de Saúde Porto Ocidental, Porto, Portugal

³Chronic Diseases Research Centre, NOVA Medical School, NOVA University of Lisbon, Lisboa, Portugal

⁴Centre for Health Technology and Services Research, University of Porto, Porto, Portugal

Corresponding Author:

Sofia Baptista, MD, MSc

Department of Community Medicine, Information and Health Decision Sciences

Faculty of Medicine

University of Porto

Alameda Professor Hernâni Monteiro

Porto, 4200 - 319

Portugal

Phone: 351 910357362

Email: sofiabaptista@gmail.com

Abstract

Background: Prostate cancer is a leading cause of cancer among men. Because screening for prostate cancer is a controversial issue, many experts in the field have defended the use of shared decision making using validated decision aids, which can be presented in different formats (eg, written, multimedia, Web). Recent studies have concluded that decision aids improve knowledge and reduce decisional conflict.

Objective: This meta-analysis aimed to investigate the impact of using Web-based decision aids to support men's prostate cancer screening decisions in comparison with usual care and other formats of decision aids.

Methods: We searched PubMed, CINAHL, PsycINFO, and Cochrane CENTRAL databases up to November 2016. This search identified randomized controlled trials, which assessed Web-based decision aids for men making a prostate cancer screening decision and reported quality of decision-making outcomes. Two reviewers independently screened citations for inclusion criteria, extracted data, and assessed risk of bias. Using a random-effects model, meta-analyses were conducted pooling results using mean differences (MD), standardized mean differences (SMD), and relative risks (RR).

Results: Of 2406 unique citations, 7 randomized controlled trials met the inclusion criteria. For risk of bias, selective outcome reporting and participant/personnel blinding were mostly rated as unclear due to inadequate reporting. Based on seven items, two studies had high risk of bias for one item. Compared to usual care, Web-based decision aids increased knowledge (SMD 0.46; 95% CI 0.18-0.75), reduced decisional conflict (MD -7.07%; 95% CI -9.44 to -4.71), and reduced the practitioner control role in the decision-making process (RR 0.50; 95% CI 0.31-0.81). Web-based decision aids compared to printed decision aids yielded no differences in knowledge, decisional conflict, and participation in decision or screening behaviors. Compared to video decision aids, Web-based decision aids showed lower average knowledge scores (SMD -0.50; 95% CI -0.88 to -0.12) and a slight decrease in prostate-specific antigen screening (RR 1.12; 95% CI 1.01-1.25).

Conclusions: According to this analysis, Web-based decision aids performed similarly to alternative formats (ie, printed, video) for the assessed decision-quality outcomes. The low cost, readiness, availability, and anonymity of the Web can be an advantage for increasing access to decision aids that support prostate cancer screening decisions among men.

(*J Med Internet Res* 2018;20(6):e228) doi:[10.2196/jmir.9070](https://doi.org/10.2196/jmir.9070)

KEYWORDS

decision making; decision aid; internet; patient participation; prostate; screening

Introduction

Prostate Cancer and Screening

According to the GLOBOCAN worldwide estimates of cancer incidence and mortality produced by the International Agency for Research on Cancer, there were 1,111,700 new cases of prostate cancer and 307,500 prostate cancer deaths in 2012, making prostate cancer the second most commonly diagnosed cancer in men and the fifth leading cause of cancer deaths among men [1].

Screening for prostate cancer remains a controversial issue, particularly after data from two major trials were released. The United States Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial found no benefits from using prostate-specific antigen (PSA) screening for prostate cancer diagnoses [2]. The results from the European Randomised Study of Screening for Prostate Cancer with data truncated at 13 years concluded that one prostate cancer death would be avoided and 27 excess cases detected per 781 men invited for screening with PSA [3]. Overdiagnosis was estimated to be as high as 41%. The proportion of prostate cancer that would never have led to clinical symptoms resulted in unnecessary biopsy procedures and treatment with potential side effects, which may include urinary, sexual (eg, erectile dysfunction), and gastrointestinal complications [4]. The CAP Randomized Clinical Trial was recently published and reported no significant difference in prostate cancer mortality with PSA screening after a median follow-up of 10 years but an increase in the detection of low-risk prostate cancer cases [5].

Although often encouraged by media and health care providers, prostate cancer screening is currently recommended by only a few organizations. After reviewing the available evidence, the United States Preventive Services Task Force released a draft recommendation in 2017, assigning a “C” grade recommendation to prostate cancer screening in men 55-69 years old, stating that the potential benefits and adverse effects of PSA-based screening are closely balanced in that age group. The decision about whether to be screened should be an individual one based on conversations with the physician about the benefits and adverse effects of screening in order to help men make a decision based on personal values and preferences [6]. Many experts defend a shared decision-making process involving doctor and patient, using validated decision aids. In fact, many guidelines issued by medical organizations such as the European Association of Urology [7], the American Cancer Society [8], and the American College of Physicians [9] support a shared decision-making process for prostate cancer screening.

Shared Decision Making and Decision Aids

According to the International Patient Decision Aids Standards Collaboration (IPDAS), decision aids are evidence-based tools designed to help people participate in decision making about health care options with the aim of improving the quality of the

decision. Many study groups have focused on the development of decision aids to support shared decision making [10,11].

As established in the original IPDAS background document, two constructs are critical for establishing the effectiveness of a decision aid: (1) the quality of both the decision-making process and (2) the actual decision. For the quality of the decision-making process, five attributes are defined, all of which are measured by different scales: (1) recognizing that a decision needs to be made, (2) feeling informed about the options, (3) understanding what values matter most for the decision, (4) discussing preferences with their practitioner, and (5) being involved in decision making. Concerning the decision quality, two core attributes should be measured: (1) a patient's knowledge of the options and outcomes and (2) agreement between the chosen option and the features that matter most for the patient [12].

The most recent systematic review and meta-analysis assessing the impact of decision aids for screening decisions concluded that decision aids can increase patient knowledge, make people feel clearer about their values, reduce decisional conflict, and promote an active patient role in decision making [13]. The authors state that more studies are needed to deepen understanding of format issues such as Web-based delivery of decision aids. In addition, if new studies can be included in the systematic review, it may be possible to sort out the reasons for heterogeneity of results (eg, the format of the decision aid). Another systematic review, focusing on decision aids for prostate cancer screening, reported similar results [14].

Decision aids may be implemented in different formats, including written (eg, pamphlet/booklet), multimedia (eg, video, DVD), or Web-based. Syrowatka et al, in a systematic review and meta-analysis that assessed computer-based decision aids for any preference-sensitive medical decision, concluded that decision aids are associated with a significant improvement in knowledge and decrease in decisional conflict. However, results were limited by high levels of heterogeneity [15]. Nevertheless, the scope of the latter review was broader, including any preference-sensitive medical decision. Thus, it did not specifically address prostate cancer screening. In addition, the authors included all decision aid formats that could be accessed with a computer (eg, Web-based, videobooklet, CD-ROM). With the increasing use and ease of access to the internet, the Web has been proposed as a promising way of delivering decision aids. Therefore, it is important to assess the impact of Web-based decision aids in the prostate cancer screening decision-making process, but the number of studies addressing this subject to date have been scarce and showed mixed results.

The IPDAS Collaboration identified 12 dimensions to assess quality of patient decision aids, one of which focused on the delivery of decision aids on the internet [16]. In fact, several theories point out the potential benefits of the internet to provide broad long-term dissemination of information that can be targeted and tailored to patient needs and preferences. Hence, IPDAS emphasized that a comprehensive systematic review

focusing on the internet delivery of decision aids was needed [17].

To our knowledge, ours is the first systematic review and meta-analysis to compare Web-based decision aids with usual care and other formats of decision aids. We sought to investigate their impact on decision quality for men making a screening decision regarding prostate cancer.

Methods

Criteria for Considering Studies for This Review

We included randomized controlled trials (RCTs) involving men who had not been previously diagnosed with prostate cancer and who were making screening decisions concerning prostate cancer. We included studies comparing Web-based decision aids to several parameters: (1) no intervention/usual care or (2) alternative decision aids formats. For interventions to be considered Web-based, they had to correspond to any program accessed over a network connection using HTTP or through a Web-based app. According to this definition, materials such as CD-ROMs or DVDs, although computer-based, were not considered Web-based. Thus, studies with such interventions were excluded. We included studies in which at least one quality of decision-making outcome (eg, knowledge, decisional conflict, and involvement in decision making) was reported. Screening behavior, either the intention to undergo PSA screening or undergoing the actual PSA screening, were defined as secondary outcomes.

Search Methods for Identification of Studies

Electronic Searches

Our search strategy for this review included searching electronic medical and social science databases: (1) PubMed, (2) Cumulative Index to Nursing and Allied Health Literature (CINAHL), (3) PsycINFO, and (4) Cochrane CENTRAL (Cochrane Central Register of Controlled Trials). Whenever possible, the search strategies ([Multimedia Appendix 1](#)) used a combination of free text and database-specific subject headings. The search was conducted in November 2016.

Searching Other Resources

We also searched trial registries (World Health Organization, National Institutes of Health, ClinicalTrials.gov), reference lists of included trials, and the Decision Aid Library Inventory.

Data Collection and Analysis

Selection of Studies

Two reviewers screened the titles and abstracts of all retrieved articles after employing the search strategy. Those included after screening were accessed in full text. Authors were contacted to clarify study eligibility. Disagreements were resolved by consensus among 3 reviewers.

Data Extraction and Management

Data extraction was performed independently by 2 reviewers. Extracted data included study design and setting, numbers, and

other characteristics of study participants and interventions in addition to outcomes and other information thought to be relevant. Whenever different publications reported on the same trial, the data corresponding to the latest follow-up were included. For cluster RCTs, we collected effect estimates and standard errors from analyses that took the clustering into account. Study authors were contacted when more detailed information was needed. Disagreements were resolved by consensus. When available, data resulting from imputation were used in the analysis in accordance with an intention-to-treat approach.

Assessment of Risk of Bias in Included Studies

Assessment of risk of bias was performed using the Cochrane tool for judging risk of bias [18].

Measures of Treatment Effect and Data Synthesis

We used mean differences (MD) for continuous variables that were measured with the same instrument, standardized mean differences (SMD) when a similar outcome was assessed with different instruments, and relative risks (RR) for dichotomous variables. Continuous variables were standardized to a scale of 0-100. In cases where outcome data (eg, standard deviations) were missing, we tried to reach one of the study authors by email to request the complete measures. If we were unsuccessful in obtaining the data from authors, we derived standard deviations from standard errors or confidence intervals [18].

We analyzed studies comparing Web-based decision aids to usual care separately from studies comparing Web-based decision aids to decision aids presented in a diverse format. Review Manager 5.3.5 software was used to estimate meta-analytical-weighted treatment effects across studies [19]. Data analysis was conducted with a random-effects model given the heterogeneity among studies being pooled.

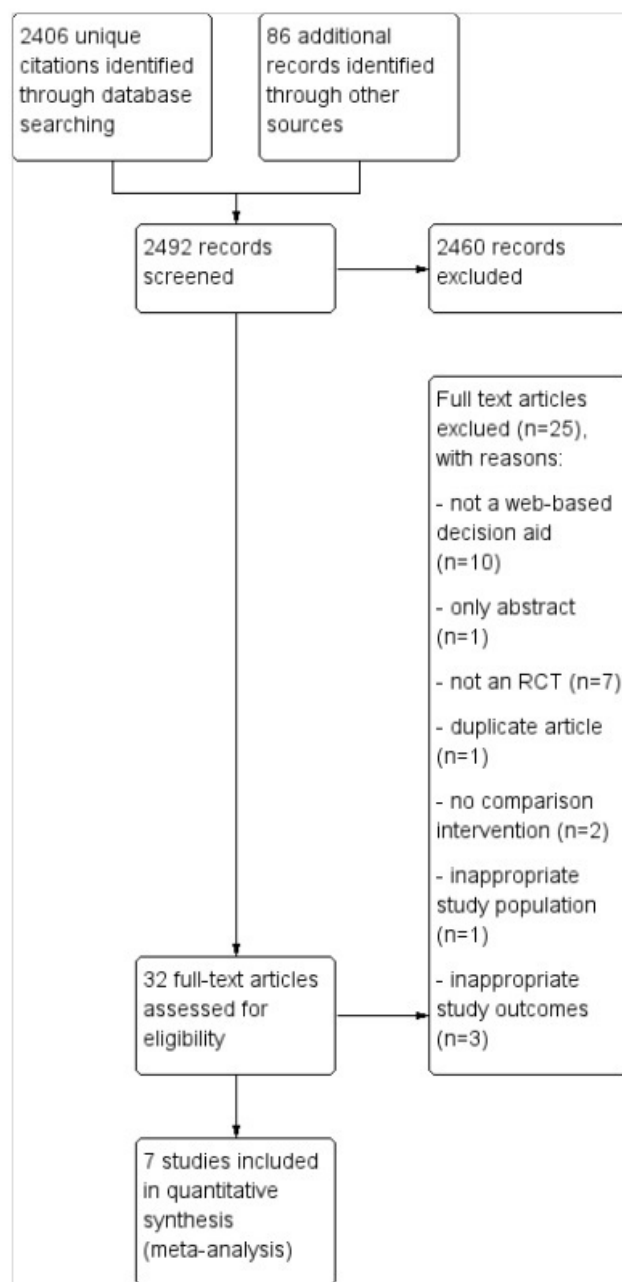
Sensitivity Analysis

A sensitivity analysis was conducted in order to reassess the effect measures after excluding trials classified as having high risk of bias for any of the chosen parameters and after applying the fixed-effects model. For the knowledge outcome, a sensitivity analysis was done using MDs as an alternative to SMDs. Although we opted for SMD to pool knowledge across studies since different constructs were used to measure this outcome, the use of MD could also be defensible, as the scale itself is the same.

Results

Results of the Search and Description of Studies

The electronic database search retrieved 2406 unique citations (2536 records), and 86 additional citations were identified through other sources ([Figure 1](#)). Of the 32 full-text articles assessed, 25 were excluded ([Figure 1](#) describes the reasons, and further details are provided in [Multimedia Appendix 2](#)). We contacted 4 investigators to clarify methodological issues and to complete the extracted data when necessary.

Figure 1. Study flow diagram. RCT: randomized controlled trial.

Included Studies

Seven studies met our inclusion criteria and were included in the meta-analysis. The characteristics of included studies are presented in [Table 1](#).

The studies were published between 2003 and 2013; five were based in the United States [20-26], one in Australia [25], and one in the United Kingdom [26]. A total of 4714 men with ages ranging from 45-75 years participated in the seven included studies. Five of the studies recruited men from a primary care setting [20-22,24,26], one through a radio and newspaper advertisement [25], and another from an industrial manufacturing worksite [23]. Six of the studies randomized individual patients [21,22,24-26], while one used the worksite as the unit of randomization [23]. For this review, we referred to control interventions as usual care unless they fulfill the

definition of a patient decision aid. Among the included studies, five studies compared Web-based decision aid performance to the usual care [22-24,26], four studies compared Web-based decision aids to printed decision aids [20,24-26], and two studies compared Web-based decision aids to video decision aids [20,25]. In addition to containing information about the options and outcomes regarding prostate cancer screening, all decision aids used in the studies provided a values clarification tool, except for one [25]. All included studies assessed knowledge, and six measured decisional conflict [21-26]. Others reported outcomes included anxiety [25,26], satisfaction with decision [20,24], decision-making role [21,23,25], and intention to undergo and actually undergoing PSA screening [21-26]. All studies used a parallel design, except Allen et al, who used a cluster RCT. Allen et al used a generalized estimating equations analysis and thus properly accounting for the cluster design and the possible associated unit-of-analysis errors.

Table 1. Characteristics of included studies.

Study	Methods	Participants	Comparison	Outcomes
Frosch 2003 [20]	RCT ^a , 2 groups: video decision versus Web-based decision aids	Men aged ≥50 years considering PSA ^b screening in a preventive medicine clinic (USA): 112 (Web)/ 114 (video)	Decision aid: Web-based, with information, pros and cons of PSA testing, experiences of other patients, values clarification exercise Active comparator: Video, same content of the Web decision aid	Primary outcome measures: (1) participant ratings of convenience, effort required, and satisfaction with the intervention, (2) knowledge about prostate cancer screening and treatment, and (3) choice of undergoing PSA test
Krist 2007 [21]	RCT, comparing Web-based decision and paper-based decision aids versus no intervention (usual care)	Men aged 50-70 years considering PSA screening in a primary care setting (USA): 226 (Web)/ 196 (paper)/ 75 (usual care-control)	Decision aid: Web-based information about prostate cancer, screening, screening benefits, and known risks, current uncertainties. The website was reviewed by a general decision aid expert and several content experts. Active comparator: print brochure, which duplicated the content of the website Comparator: usual care	Primary outcome: patient-reported control preferences scale score. Other: Prostate cancer screening knowledge, time spent discussing screening, topics covered in the discussion, decisional conflict scale score and whether a PSA test was ordered
Frosch 2008 [22]	RCT, 4 groups: Web-based decision aid (1) vs Web decision aid + chronic disease trajectory (2) vs chronic disease trajectory (3) vs usual care (internet info) (4)	Men aged ≥50 years considering PSA screening in a preventive medicine clinic (USA): 155 (1) + 152 (2) + 153 (3) + 151 (4)	Decision aid: information about prostate cancer screening and treatment, with physician and patient testimonials contrasting different preferences and decisions Active comparator: chronic disease trajectory model that prompted patients to express utilities for outcomes associated with a prostate cancer life course by contrasting screening with no screening in its impact on quality of life and longevity Comparator: links to public websites on prostate cancer screening maintained by the American Cancer Society and the Centers for Disease Control and Prevention	Primary outcome measures: (1) knowledge; (2) actual option; (3) decisional conflict. Other outcomes: (1) treatment preference if cancer diagnosed and (2) concern about prostate cancer
Allen 2010 [23]	RCT, 2 groups (Web and control)	Men aged ≥45 years considering PSA screening (USA): 398 (Web)/ 414 (no intervention)	Decision aid: Web-based (content based on expert opinion and guidelines from IPDAS ^c). Comparator: no intervention	Primary outcomes: (1) decisional status, (2) prostate cancer knowledge, (3) decision self-efficacy, (4) consistency between values, and (5) screening decision. Secondary outcomes: (1) preference for control in decision making and (2) decisional conflict
Taylor 2013 [24]	RCT, 3 groups (Web, paper, and control)	Men aged 45-70 years considering PSA screening in a primary care setting (USA): 631 (Web)/ 630 (paper)/ 632 (usual care)	Decision aid: Web-based Active comparator: printed decision aid Comparator: usual care Both decision aids share same content: (1) description of screening tests and possible results, (2) information about treatment options, risks and adverse effects, (3) a review of prostate cancer risk factors and encouragement to discuss screening with a physician, and (4) a 10-item values clarification tool; and resources for more information. Web-based decision aids also included interactive features (eg, testimonials, interactive values clarification tools)	Knowledge, decisional conflict scale, satisfaction with decision scale, prostate cancer screening uptake. Measured at baseline, and then after 1 and 13 months

Study	Methods	Participants	Comparison	Outcomes
Ilic 2008 [25]	RCT, 3 groups (Web, paper, and video)	Men aged ≥ 45 years considering PSA screening in Australia, recruited by radio and newspaper advertisements: 56 (Web)/ 50 (pamphlet)/ 55 (video)	Decision aid: Web-based Active comparator: pamphlet Active comparator: video Decision aid contents: (1) epidemiology on prostate cancer, (2) diagnostic process, (3) treatment options, and (4) the associated benefits/risks	Primary outcome: decisional conflict. Secondary outcomes: (1) knowledge and (2) anxiety, consumer decision-making role and screening interest
Evans 2010 [26]	RCT, 4 groups: 2 intervention groups (Web- and paper-based) and 2 control groups (questionnaire and usual care)	Men aged 50-75 years considering PSA screening in a primary care setting (UK): 129 (Web)/ 126 (paper)/127 (questionnaire)/ 132 (usual care)	Decision aid: Web-based - Prosdex: information, pros and cons of PSA testing, other patient experiences, values clarification exercise Active comparator: paper version with the text of the website Active comparator: questionnaire Comparator: usual care	Primary outcome: knowledge of prostate cancer and PSA. Other: attitudes towards PSA testing; behavior (intention to undergo PSA testing), anxiety, decisional conflict, and actually undergoing of PSA test (at 6 months)

^aRCT: randomized controlled trial.

^bPSA: prostate-specific antigen.

^cIPDAS: International Patient Decision Aids Standards.

Risk of Bias in Included Studies

Assessments of the risk of bias for each study are summarized in [Multimedia Appendix 3](#) and the authors' support for each judgment are presented in [Multimedia Appendix 4](#). Random sequence generation was rated as being at low risk of bias in most of the studies (6/7, 86%) and unclear risk of bias in one study. Allocation concealment was considered low risk of bias in five studies (5/7, 71%) and unclear risk in the remaining two studies.

Blinding of participants and personnel was assessed as being at low risk of bias in one study (1/7, 14%), unclear risk of bias in four studies (4/7, 57%), and high risk in two studies (2/7, 29%) [20,21]. All studies were evaluated as being a low risk of bias regarding blinding of outcome assessment. All studies were rated as low risk of attrition bias that relates to incomplete outcome data. Five studies (5/7, 71%) were classified as unclear risk of bias regarding selective reporting due to the lack of information about public registration of the trial protocol. The other two studies had a registered protocol and were rated as low risk of bias for the selective reporting parameter. When assessing other sources of bias, six studies were rated as low risk of bias (6/7, 86%). The remaining study was considered unclear risk of bias as study groups were not similar in size [20].

Effects of Interventions

The summary of the findings is found in [Multimedia Appendix 5](#).

Knowledge

All seven studies assessed patient knowledge in the meta-analysis. Studies tested knowledge through questionnaires based on the content of the decision aids. The number of correct

answers was transformed into a scale ranging from 0% (no correct answers) to 100% (all correct answers).

Web-Based Decision Aids Versus Usual Care

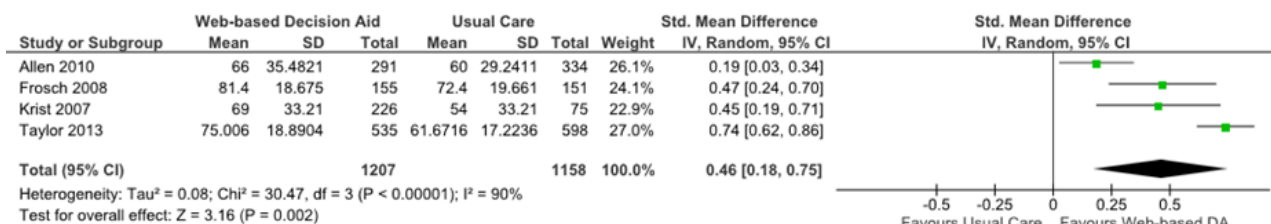
Four studies included knowledge comparisons for this outcome in the meta-analysis. One study used a different way for grading the questionnaire (1 point for a correct answer, 0 for any unanswered item, and -1 for an incorrect answer), so data could not be transformed for the scale described above. In addition, no standard deviations could be obtained; thus data could not be pooled [26]. Compared to the usual care, patients allocated to Web-based decision aids had higher average knowledge scores (SMD 0.46; 95% CI 0.18-0.75; [Figure 2](#)). The study that was not included in the meta-analysis showed a higher statistically significant average score for the Web-based decision aid group in comparison with the usual care. Four of the five RCTs assessing knowledge for Web-based decision aids compared to usual care demonstrated a statistically significant improvement in knowledge in the Web-based decision aid group [19-24]. Taylor et al reassessed knowledge at 13 months, and the Web-based decision aid group continued to register a statistically significant increase in median scores compared to the usual care group [25].

Web-Based Versus Printed Decision Aids

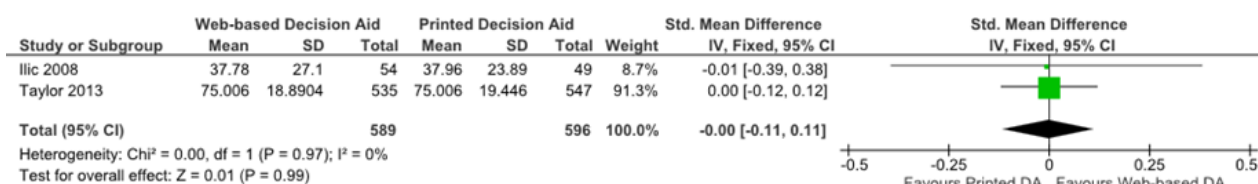
Four studies assessed knowledge for the comparison of Web-based to printed decision aids, but only data from two studies could be pooled. The scale used by Evans et al was not convertible to 0%-100% scale [24]. Additionally, standard deviations for study results could not be obtained for the Evans et al and Krist et al studies; these studies did not find any differences between groups regarding this outcome. No differences in the average knowledge scores were found for this comparison (SMD 0.00; 95% CI -0.11 to 0.11; [Figure 2](#)) [21,26].

Figure 2. Forest plots of standardized mean differences for knowledge. A) Web-based decision aids (DA) versus usual care, B) Web-based decision aids versus printed decision aids, C) Web-based decision versus video decision aids.

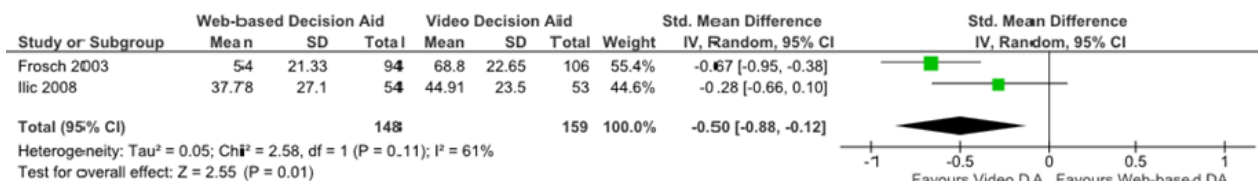
A



B



C



Web-Based Versus Video Aids

With regard to the comparison of Web-based decision to video decision aids, the Web-based group registered lower average knowledge scores (SMD -0.50 ; 95% CI -0.88 to -0.12 ; Figure 2) for the pooled data for two studies. Frosch et al found a smaller nonstatistically significant difference between the two groups when only the participants who reviewed the complete set of materials were considered for analysis [22].

Decisional Conflict

Six of the seven studies measured patient decisional conflict using the decisional conflict scale [20-25]. The decisional conflict scale consists of five subscales, and total scores range from 0 (no decisional conflict) to 100 (extremely high decisional conflict). When comparing Web-based aids to usual care or alternative formats of decision aids, a negative score corresponds to a reduction in decisional conflicts, which favors Web-based decision aids.

Web-Based Decision Aids Versus Usual Care

Five studies compared Web-based decision aids to usual care in terms of decisional conflict [22,23,26]. It was not possible to pool data from two studies due to lack of standard deviation of the results [21,26]. Krist et al did not find a significant difference between the two groups in contrast with the findings of Evans et al who reported a significant higher decisional

conflict for the usual care group [21,26]. Frosch et al reported the results using subscales without providing standard deviation data; this study finding showed significantly higher decisional conflict for the usual care group in the subscales of “feeling informed” and “support in decision making”, and no difference was found for the subscales “uncertainty” and “having made an effective decision” [22]. The overall MD for decisional conflict comparing Web-based decision aids versus usual care was -7.07% (95% CI -9.44 to -4.71 ; Figure 3).

Web-Based Decision Versus Printed Decision Aids

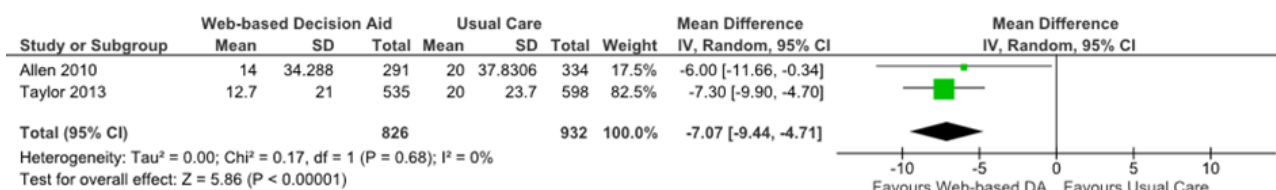
Four studies assessed decisional conflict by comparing Web-based decision to printed decision aids [21,24-26]. The MD for pooled data from two studies was 0.68 (95% CI -1.46 to 2.83 ; Figure 3). Data from the Evans et al and Krist et al studies could not be included for meta-analysis because standard deviations could not be obtained [21,26]. Reported mean scores for decisional conflicts were similar for the two groups in the Krist et al study [21]. Evans et al did not find any statistically significant differences for decisional conflict when a Web-based decision aid was compared to a printed one [26].

Web-Based Decision Versus Video Decision Aids

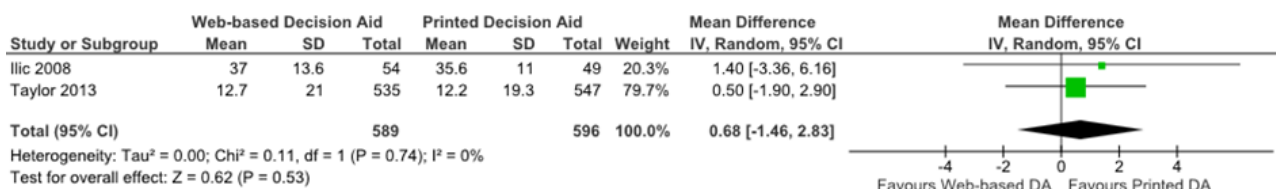
Ilic et al did not find any statistical differences regarding mean decisional conflict scores for patients when exposed to Web-based decision compared to video decision aids [25].

Figure 3. Forest plots of mean differences for decisional conflict. A) Web-based decision aids (DA) versus usual care, B) Web-based decision aids versus printed decision aids.

A



B



Participation in Decision Making

Three of the seven studies evaluated participation in decision making using the Control Preferences Scale (CPS) [21,23,25], which consists of five statements (A to E), two of which reflect patient controlled decision making, another one refers to shared decision making, and the last two reflect practitioner-controlled decision making. Decision aids are intended to enhance a patient's active role in decision making. Therefore, a pooled $RR > 1$ for group differences in CPS statements A to C favors Web-based decision aids, and a pooled $RR < 1$ for statements D and E also favors Web-based decision aids.

Web-Based Decision Aids Versus Usual Care

When comparing Web-based decision aids to usual care in terms of a patient-controlled or active role in the decision-making process, the pooled RR was 1.06 (95% CI 0.97-1.16; Figure 4). For the practitioner-controlled role, a pooled RR of 0.50 was obtained, which compared Web-based decision aids to usual care (95% CI 0.31-0.81; Figure 4).

Web-Based Decision Versus Printed Decision Aids

Regarding patients who assumed an active role according to the CPS, the pooled RR was 0.96 (95% CI 0.77-1.19; Figure 5). The pooled RR for the same comparison for a collaborative role in decision making was 1.12 (95% CI 0.78-1.60; Figure 5). Finally, when pooling data that compared Web-based decision to printed decision aids in terms of a passive role according to the CPS, the RR obtained was 0.83 (95% CI 0.47-1.48; Figure 5).

Web-Based Decision Versus Video Decision Aids

Ilic et al was the only study assessing participation in decisions comparing Web-based decision to video decision aids. No statistically significant differences between groups were found for active (RR 0.89; 95% CI 0.66-1.21), collaborative (RR 1.15; 95% CI 0.68-1.95), or passive patient role in decision making according to the CPS (RR 1.47; 95% CI 0.26-8.46) [25].

Screening Behavior: Preferred Option

Three studies investigated the preferred patient options concerning prostate screening using the PSA test [20,22,23]; Evans et al and Ilic et al evaluated answers using a 5-point Likert-like response scale [25,26]. Allen et al reported agreement with the statement "want to be screened" [23].

Web-Based Decision Aids Versus Usual Care

The pooled RR for two studies comparing the preference for having a PSA test for patients using a Web-based decision aid in comparison to usual care was 0.84 (95% CI 0.59-1.21; Figure 6).

Web-Based Decision Versus Printed Decision Aids

When comparing Web-based decision to printed decision aids, the overall pooled RR indicating a preference for PSA screening was 0.93 (95% CI 0.61-1.41; Figure 6).

Web-Based Decision Versus Video Decision Aids

Ilic et al assessed the preference for the PSA screening test with a 5-point response scale. When comparing those who responded either "definitely want" and "probably want" in the Web-based decision aid and video groups, there was no significant difference (RR 1.29; 95% CI 0.99-1.67) [25].

Screening Behavior: PSA Test

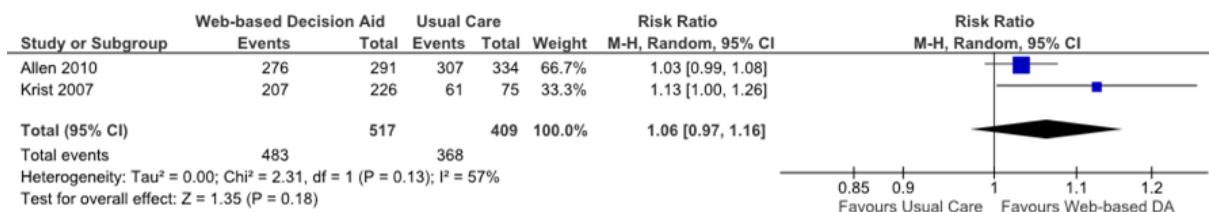
Using different methods, five studies investigated the actual choice of PSA screening: (1) Evans et al asked general practitioners to review participant's medical records [26], (2) Frosch et al also relied on medical records [22], (3) Taylor et al assessed patients' self-reported PSA screening at 13 months [24], (4) Krist et al used patients' reports of PSA tests ordered [19], and (5) Frosch et al searched for PSA test requests in medical records [20].

Web-Based Decision Aids Versus Usual Care

When comparing Web-based decision aids to usual care groups after pooling data for screening uptake, the obtained RR was 1.0 (95% CI 0.89-1.11; Multimedia Appendix 6).

Figure 4. Forest plots of relative risks for participation in decision making. A) Patient controlled or shared decision making: Web-based decision aid (DA) versus usual care, B) Practitioner controlled decision making: Web-based decision aid versus usual care.

A



B

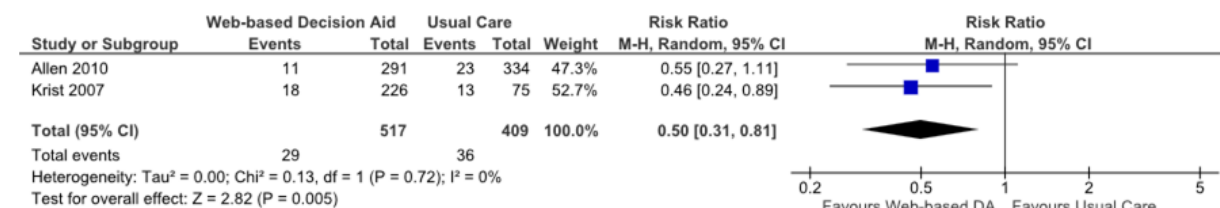
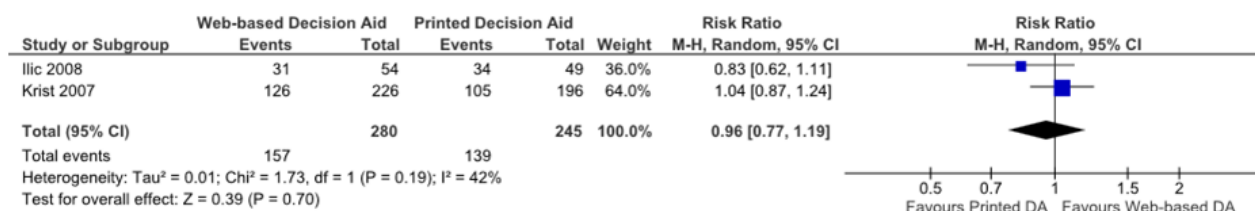
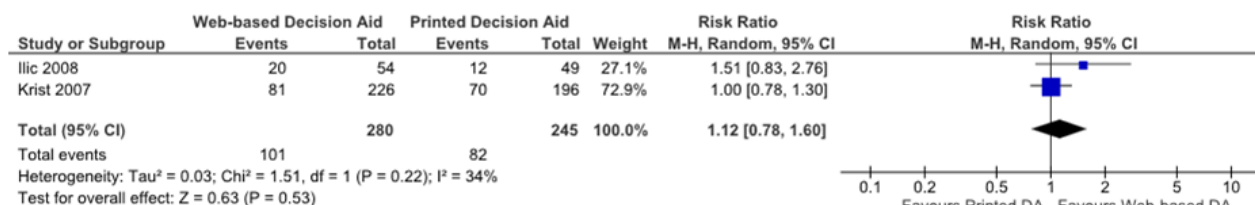


Figure 5. Forest plot of relative risks for participation in decision making: Web-based decision aids versus printed decision aids (DA). A) Patient controlled, B) Shared decision making, C) Practitioner controlled.

A



B



C

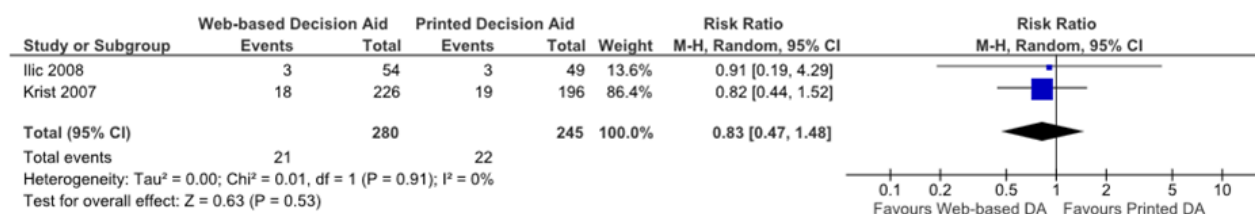
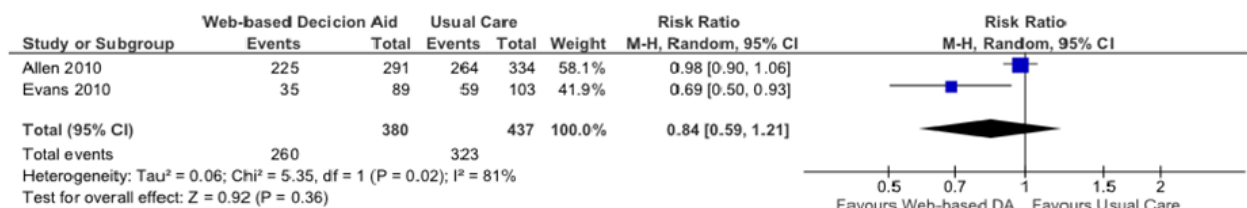
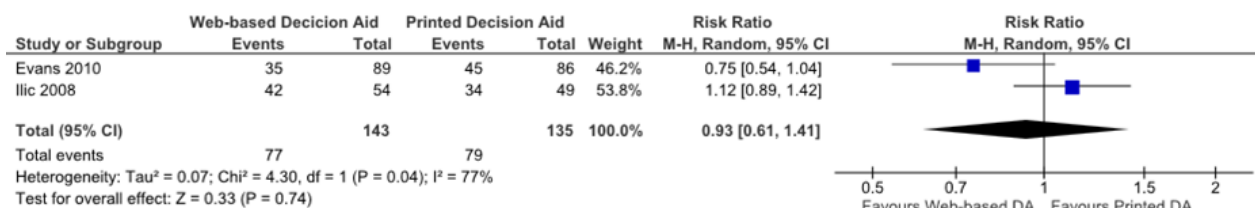


Figure 6. Forest plot of relative risks for screening behavior - preference for prostate-specific antigen test. A) Web-based decision aids (DA) versus usual care, B) Web-based decision aids versus printed decision aids.

A



B



Web-Based Decision Aids Versus Printed Decision Aids

The overall pooled RR was 1.04 (95% CI 0.97-1.12; [Multimedia Appendix 6](#)) when comparing PSA tests in patients exposed to Web-based decision aids to patients receiving printed decision aids.

Web-Based Decision Aids Versus Video Decision Aids

The only study assessing PSA test uptake in patients receiving a Web-based decision aid versus patients receiving a video decision aid revealed a slight difference for the comparison (RR 1.12; 95% CI 1.0-1.25) [20].

Sensitivity Analysis

We investigated the potential bias resulting from including studies that were assessed as high risk of bias for any of the seven criteria considered, which resulted in exclusion of two studies for this analysis [20,21]. Most results remained similar, with the following exception: the differences in knowledge between Web-based decision and video groups became nonsignificant. The difference for a practitioner-controlled role in decision making between Web-based decision aids and usual care also became nonsignificant. After applying a fixed-effect model, the results were also compared to the results retrieved in the first analysis with the random-effect model. Results were similar for all outcomes and comparisons, with one exception. When comparing Web-based decision aids to usual care for a patient active or collaborative role in decision, with a slight decimal change, the difference became significant. These similar results of fixed-effect model to random-effect model diminish risk of bias due to “small study effects” (ie, the potential of the included small studies to overestimate effect sizes).

Taking into account IPDAS criteria [11], decision aids should offer a values clarification tool. After a verification process, which included contact with authors, Ilic et al was the only one of the included studies that did not contain such tool. We conducted the analysis after removal of this study. The results were similar to those obtained when the study was included.

When using mean differences instead of standardized mean differences for knowledge, results were similar for the comparisons of Web-based decision aids versus usual care (MD 10.66%; CI 95% 6.78-14.53), Web-based decision versus printed decision aids (MD -0.01%; CI 95% -2.23 to 2.22), and Web-based decision versus video decision aids (MD -11.9%; CI 95% -19.19 to -4.61).

Heterogeneity

Statistically significant heterogeneity was found for knowledge when comparing Web-based decision aids to usual care.

Discussion

Principal Findings and Evidence

In comparison with usual care, Web-based decision aids significantly increased knowledge, reduced decisional conflict, and reduced the practitioner-controlled role in the decision-making process. No differences were found regarding patients assuming an active or collaborative role in decision making or in terms of screening behavior. When comparing Web-based decision with printed decision aids, no differences were found for knowledge, decisional conflict, participation in decision, or screening behavior. Compared to video decision aids, Web-based decision aids showed lower average knowledge scores and a slight decrease in PSA screening uptake, while no differences were found regarding participation in decision making. None of the studies assessed decisional conflict for these comparisons.

There is high-quality evidence that suggests that Web-based decision aids when compared to printed decision aids perform similarly in improving men's knowledge regarding prostate cancer screening and reducing decisional conflict. There is moderate-quality evidence that Web-based decision compared to printed decision aids show no differences in screening behavior. There is also low-quality evidence that Web-based

decision aids resulted in lower knowledge scores when compared with video decision aids.

Interpretation the Context of Existing Literature

Our results are similar to those from other systematic reviews and meta-analyses indicating the superiority of decision aids (in any format) in comparison with usual care, but to our knowledge, our study is the first to compare Web-based decision aids to alternative formats in the context of prostate cancer screening decisions. Stacey et al, in a Cochrane meta-analysis of decision aids for people facing screening or treatment issues, found decision aids to improve people's knowledge, reduce decisional conflict, promote an active patient role in decision making, and reduce the number of patients choosing to undergo PSA screening [13]. In a systematic review and meta-analysis of features of computer-based decision aids for any preference-sensitive medical decision, Syrowatka et al indicated that decision aids are associated with significant knowledge improvement and decrease in decisional conflict. However, results were limited by high levels of heterogeneity [15]. Ilic et al, in a systematic review assessing the effectiveness of decision aids for decision making in prostate cancer testing, also reported a reduction in decisional conflict and a statistically significant improvement in knowledge [14].

Syrowatka et al reported that computer-based decision aids were associated with significant improvements in knowledge and decisional conflict compared to usual care or alternative aids [15]. On the other hand, our results show Web-based decision aids perform similarly to printed decision aids in terms of decision quality outcomes. Comparisons should be made with caution, since the systematic review by Syrowatka et al addressed all computer-based decision aids (not only Web-based) and any preference-sensitive medical decision. It may not be surprising that Web-based and printed decision aids perform similarly in the context of a trial, since the contents of the decision aid in both arms is the same. Therefore, the current evidence supports the use of decision aids. However, either decision aid format may be used depending on individual patient's preference. In addition, we hypothesize that in a busy day-to-day clinical practice, with limited time to talk to patients, Web-based decision aids may have greater potential, allowing patients to easily access and review the material prior to the encounter with the physician, which may impact the shared decision-making process.

Only two studies compared Web-based to video decision aids, which limits our conclusions, particularly concerning the fact that Web-based decision aids showed lower average knowledge scores [20,25]. However, we can hypothesize that older men making a prostate cancer screening decision may not be very familiar with internet use. Of note, in Frosch et al, the video decision aid population arm had a specifically allocated time to watch the video, which was different from the Web-based decision aid group. This difference may have increased adherence to video visualization. This study showed that the video group had significantly more probability of reviewing the materials, while in the Web-based decision aid group, only 53.5% watched the entire presentation, and 39.5% had not reviewed any part of it. In fact, the authors reported that for

those in the Web-based decision aid group who reviewed the entire presentation, knowledge scores were similar to those from the video group [20]. More studies are needed in order to address the comparison of decision aids reviewed at an assigned versus self-allocated time.

Strengths and Limitations

Among the included studies, the risk of bias was higher for the blinding of participants and personnel criteria. Post-hoc analysis removing studies at high risk of bias yielded the same results, except that no differences were found for knowledge comparing Web-based to video decision aids and no differences for a practitioner-controlled role in decision making when comparing Web-based decision aids to usual care.

Several limitations must be considered while analyzing our conclusions. For most outcomes, the number of studies was low, making it difficult to assess for publication bias. The different contents of the decision aids of each study also limited their comparison. We tried to pool data only when the same scale or procedure was used to evaluate each outcome. However, different ways of formulating the questions may also pose a limitation to our conclusions. The way studies measured screening behavior could also have introduced bias. The methods used, such as self-reported screening, review of medical records, and evaluation of intention to undergo screening, may not be reliable. PSA screening decisions may have to be made by men several times in their life, so lack of long-term follow-up in the included studies also limits the understanding of the impact of Web-based decision aids over time. Taylor et al was the study with the longest follow-up (13 months) [24].

Some studies mention visualization rates of the decision aids. Allen et al refers rates from 23%-59% in the intervention Web-based decision aid group [23]. In Frosch et al, 86.7% of the Web-based decision aid arm participants at least clicked on the link provided [22]. It is possible that if all men effectively viewed the materials, the impact of decision aids could be amplified. Results could also have been influenced due to the fact that people in the printed decision aid arms could have the decision aid with them while answering the questionnaires.

Finally, participants in the majority of included studies were white men with high educational levels, which limits the generalization of results for other populations, such as low literacy men and cultural minorities. Allen et al addressed a low literacy population in nonclinical settings (worksites) that differed from the other included studies [23]. Taylor et al also included many participants from a low socioeconomic background [24]. More trials should focus on these populations and investigate the delivery of decision aids in different settings.

We hypothesize that the statistically significant heterogeneity found for knowledge when comparing Web-based decision aids and usual care in conjunction with Web-based to video decision aids may be due to the fact that the tests used for this assessment in the included studies were not standardized.

Conclusions

According to this analysis, the Web format seems to have a similar effect to printed or video decision aids in terms of

increase in knowledge and decrease in decisional conflict. This provides evidence to use decision aids to support a patient's prostate cancer screening decision, whether it is a Web-based decision aid or alternative format like a printed or video version. Of note, most included studies were published 7 or more years ago. In the last several years, the internet has become even more ubiquitous and easy to use with many public places providing it at no cost, which may be an important feature to increase access to decision aids. Another potential advantage, especially for health preference sensitive issues, is the anonymity that Web allows. Increasingly more decision aids will likely become

available through this media, and more men will be skilled enough to search for them online. More RCTs are needed to further compare the impact of these alternative decision aid formats in decision making and to analyze their influence not only in the short term, but also over time. In addition, more studies are needed to deepen our understanding of the unique features of Web-based decision aids, such as virtual connectivity, interactivity, tailoring, as well as to compare Web-based decision aids with video and printed decision aids in terms of implementation and dissemination strategies and cost-effectiveness analyses [17].

Acknowledgments

This study was supported by the European Regional Development Fund through the operation (POCI-01-0145-FEDER-007746) funded by the Programa Operacional Competitividade e Internacionalização–Compete2020 and by National Funds through Fundação para a Ciência e a Tecnologia within the Center for Health Technology and Services Research, R&D Unit (reference UID/IC/4255/2013).

Authors' Contributions

SB, CM, and BH were responsible for study conception and design; SB and ETS acquired data; SB and LFA analyzed and interpreted data; SB and ETS drafted the paper; and CM, BH, and LFA revised it critically. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies for electronic databases.

[PDF File (Adobe PDF File), 134KB - [jmir_v20i6e228_app1.pdf](#)]

Multimedia Appendix 2

Characteristics of excluded studies.

[PDF File (Adobe PDF File), 34KB - [jmir_v20i6e228_app2.pdf](#)]

Multimedia Appendix 3

Risk of bias graph and risk of bias summary.

[PDF File (Adobe PDF File), 115KB - [jmir_v20i6e228_app3.pdf](#)]

Multimedia Appendix 4

Characteristics of included studies.

[PDF File (Adobe PDF File), 86KB - [jmir_v20i6e228_app4.pdf](#)]

Multimedia Appendix 5

Summary of findings.

[PDF File (Adobe PDF File), 44KB - [jmir_v20i6e228_app5.pdf](#)]

Multimedia Appendix 6

Forest plots of RRs for screening behavior - PSA test uptake.

[PDF File (Adobe PDF File), 35KB - [jmir_v20i6e228_app6.pdf](#)]

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015 Mar;65(2):87-108 [FREE Full text] [doi: [10.3322/caac.21262](https://doi.org/10.3322/caac.21262)] [Medline: [25651787](https://pubmed.ncbi.nlm.nih.gov/25651787/)]
2. Andriole GL, Crawford ED, Grubb RL, Buys SS, Chia D, Church TR, PLCO Project Team. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst* 2012 Jan 18;104(2):125-132 [FREE Full text] [doi: [10.1093/jnci/djr500](https://doi.org/10.1093/jnci/djr500)] [Medline: [22228146](https://pubmed.ncbi.nlm.nih.gov/22228146/)]
3. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 2014 Dec 6;384(9959):2027-2035 [FREE Full text] [doi: [10.1016/S0140-6736\(14\)60525-0](https://doi.org/10.1016/S0140-6736(14)60525-0)] [Medline: [25108889](https://pubmed.ncbi.nlm.nih.gov/25108889/)]
4. Heijnsdijk EAM, Wever EM, Auvinen A, Hugosson J, Ciatto S, Nelen V, et al. Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med* 2012 Aug 16;367(7):595-605 [FREE Full text] [doi: [10.1056/NEJMoa1201637](https://doi.org/10.1056/NEJMoa1201637)] [Medline: [22894572](https://pubmed.ncbi.nlm.nih.gov/22894572/)]
5. Martin RM, Donovan JL, Turner EL, Metcalfe C, Young GJ, Walsh EI, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial. *JAMA* 2018 Dec 06;319(9):883-895. [doi: [10.1001/jama.2018.0154](https://doi.org/10.1001/jama.2018.0154)] [Medline: [29509864](https://pubmed.ncbi.nlm.nih.gov/29509864/)]
6. US Preventive Services Task Force. Prostate Cancer Screening Draft Recommendations URL: <https://screeningforprostatecancer.org/frequently-asked-questions/> [accessed 2017-09-27] [WebCite Cache ID 6tnLHMri6]
7. EAU Guidelines on Prostate Cancer 2017. URL: <http://uroweb.org/individual-guidelines/oncology-guidelines/> [accessed 2017-09-27] [WebCite Cache ID 6tnLe8uog]
8. Wolf AMD, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, Volk RJ, American Cancer Society Prostate Cancer Advisory Committee. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin* 2010;60(2):70-98 [FREE Full text] [doi: [10.3322/caac.20066](https://doi.org/10.3322/caac.20066)] [Medline: [20200110](https://pubmed.ncbi.nlm.nih.gov/20200110/)]
9. Qaseem A, Barry MJ, Denberg TD, Owens DK, Shekelle P, Clinical Guidelines Committee of the American College of Physicians. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med* 2013 May 21;158(10):761-769. [doi: [10.7326/0003-4819-158-10-201305210-00633](https://doi.org/10.7326/0003-4819-158-10-201305210-00633)] [Medline: [23567643](https://pubmed.ncbi.nlm.nih.gov/23567643/)]
10. Sheridan SL, Harris RP, Woolf SH, Shared Decision-Making Workgroup of the U.S. Preventive Services Task Force. Shared decision making about screening and chemoprevention. A suggested approach from the U.S. Preventive Services Task Force. *Am J Prev Med* 2004 Jan;26(1):56-66. [Medline: [14700714](https://pubmed.ncbi.nlm.nih.gov/14700714/)]
11. Volk R, Lewellyn-Thomas H. The IPDAS Background Document: An Introduction. 2012. URL: <http://ipdas.ohri.ca/IPDAS-Introduction.pdf> [accessed 2017-09-27] [WebCite Cache ID 6tnMW557p]
12. Sepucha KR, Borkhoff CM, Lally J, Levin CA, Matlock DD, Ng CJ, et al. Establishing the effectiveness of patient decision aids: key constructs and measurement instruments. *BMC Med Inform Decis Mak* 2013;13 Suppl 2:S12 [FREE Full text] [doi: [10.1186/1472-6947-13-S2-S12](https://doi.org/10.1186/1472-6947-13-S2-S12)] [Medline: [24625035](https://pubmed.ncbi.nlm.nih.gov/24625035/)]
13. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2017 Dec 12;4:CD001431. [doi: [10.1002/14651858.CD001431.pub5](https://doi.org/10.1002/14651858.CD001431.pub5)] [Medline: [28402085](https://pubmed.ncbi.nlm.nih.gov/28402085/)]
14. Ilic D, Jammal W, Chiarelli P, Gardiner RA, Hughes S, Stefanovic D, et al. Assessing the effectiveness of decision aids for decision making in prostate cancer testing: a systematic review. *Psychooncology* 2015 Apr 15:-. [doi: [10.1002/pon.3815](https://doi.org/10.1002/pon.3815)] [Medline: [25873433](https://pubmed.ncbi.nlm.nih.gov/25873433/)]
15. Syrowatka A, Krömker D, Meguerditchian AN, Tamblyn R. Features of Computer-Based Decision Aids: Systematic Review, Thematic Synthesis, and Meta-Analyses. *J Med Internet Res* 2016 Jan 26;18(1):e20 [FREE Full text] [doi: [10.2196/jmir.4982](https://doi.org/10.2196/jmir.4982)] [Medline: [26813512](https://pubmed.ncbi.nlm.nih.gov/26813512/)]
16. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, International PDAS Collaboration. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* 2006 Aug 26;333(7565):417 [FREE Full text] [doi: [10.1136/bmj.38926.629329.AE](https://doi.org/10.1136/bmj.38926.629329.AE)] [Medline: [16908462](https://pubmed.ncbi.nlm.nih.gov/16908462/)]
17. Hoffman AS, Volk RJ, Saarimaki A, Stirling C, Li LC, Härter M, et al. Delivering patient decision aids on the Internet: definitions, theories, current evidence, and emerging research areas. *BMC Med Inform Decis Mak* 2013;13 Suppl 2:S13 [FREE Full text] [doi: [10.1186/1472-6947-13-S2-S13](https://doi.org/10.1186/1472-6947-13-S2-S13)] [Medline: [24625064](https://pubmed.ncbi.nlm.nih.gov/24625064/)]
18. Higgins JPT. The Cochrane Collaboration. 2011. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 URL: <http://www.cochrane-handbook.org> [accessed 2018-06-13] [WebCite Cache ID 708osXApq]
19. Review Manager (RevMan) 5.2.6. Copenhagen: The Nordic Cochrane Centre, the Cochrane Collaboration; 2013. URL: <http://community.cochrane.org/help/tools-and-software/revman-5> [accessed 2018-06-18] [WebCite Cache ID 70GujeiD6]
20. Frosch DL, Kaplan RM, Felitti VJ. A randomized controlled trial comparing internet and video to facilitate patient education for men considering the prostate specific antigen test. *J Gen Intern Med* 2003 Oct;18(10):781-787 [FREE Full text] [Medline: [14521639](https://pubmed.ncbi.nlm.nih.gov/14521639/)]
21. Krist AH, Woolf SH, Johnson RE, Kerns JW. Patient education on prostate cancer screening and involvement in decision making. *Ann Fam Med* 2007;5(2):112-119 [FREE Full text] [doi: [10.1370/afm.623](https://doi.org/10.1370/afm.623)] [Medline: [17389534](https://pubmed.ncbi.nlm.nih.gov/17389534/)]

22. Frosch DL, Bhatnagar V, Tally S, Hamori CJ, Kaplan RM. Internet patient decision support: a randomized controlled trial comparing alternative approaches for men considering prostate cancer screening. *Arch Intern Med* 2008 Feb 25;168(4):363-369. [doi: [10.1001/archinternmed.2007.111](https://doi.org/10.1001/archinternmed.2007.111)] [Medline: [18299490](https://pubmed.ncbi.nlm.nih.gov/18299490/)]
23. Allen JD, Othus MKD, Hart A, Tom L, Li Y, Berry D, et al. A randomized trial of a computer-tailored decision aid to improve prostate cancer screening decisions: results from the Take the Wheel trial. *Cancer Epidemiol Biomarkers Prev* 2010 Sep;19(9):2172-2186 [FREE Full text] [doi: [10.1158/1055-9965.EPI-09-0410](https://doi.org/10.1158/1055-9965.EPI-09-0410)] [Medline: [20716619](https://pubmed.ncbi.nlm.nih.gov/20716619/)]
24. Taylor KL, Williams RM, Davis K, Luta G, Penek S, Barry S, et al. Decision making in prostate cancer screening using decision aids vs usual care: a randomized clinical trial. *JAMA Intern Med* 2013 Oct 14;173(18):1704-1712 [FREE Full text] [doi: [10.1001/jamainternmed.2013.9253](https://doi.org/10.1001/jamainternmed.2013.9253)] [Medline: [23896732](https://pubmed.ncbi.nlm.nih.gov/23896732/)]
25. Ilic D, Egberts K, McKenzie JE, Risbridger G, Green S. Informing men about prostate cancer screening: a randomized controlled trial of patient education materials. *J Gen Intern Med* 2008 Apr;23(4):466-471 [FREE Full text] [doi: [10.1007/s11606-007-0466-z](https://doi.org/10.1007/s11606-007-0466-z)] [Medline: [18373146](https://pubmed.ncbi.nlm.nih.gov/18373146/)]
26. Evans R, Joseph-Williams N, Edwards A, Newcombe RG, Wright P, Kinnersley P, et al. Supporting informed decision making for prostate specific antigen (PSA) testing on the web: an online randomized controlled trial. *J Med Internet Res* 2010;12(3):e27 [FREE Full text] [doi: [10.2196/jmir.1305](https://doi.org/10.2196/jmir.1305)] [Medline: [20693148](https://pubmed.ncbi.nlm.nih.gov/20693148/)]

Abbreviations

CENTRAL: Cochrane Central Register of Controlled Trials
CINAHL: Cumulative Index to Nursing and Allied Health Literature
CPS: Controlled Preferences Scale
IPDAS: International Patient Decision Aids Standards Collaboration
MD: mean difference
PSA: prostate specific antigen
RCT: randomized controlled trial
RR: relative risk
SMD: standardized mean difference

Edited by G Eysenbach; submitted 27.09.17; peer-reviewed by A Syrowatka, S Chambers, S Roy; comments to author 18.02.18; revised version received 14.04.18; accepted 08.05.18; published 26.06.18.

Please cite as:

Baptista S, Teles Sampaio E, Heleno B, Azevedo LF, Martins C

Web-Based Versus Usual Care and Other Formats of Decision Aids to Support Prostate Cancer Screening Decisions: Systematic Review and Meta-Analysis

J Med Internet Res 2018;20(6):e228

URL: <http://www.jmir.org/2018/6/e228/>

doi: [10.2196/jmir.9070](https://doi.org/10.2196/jmir.9070)

PMID: [29945857](https://pubmed.ncbi.nlm.nih.gov/29945857/)

©Sofia Baptista, Elvira Teles Sampaio, Bruno Heleno, Luís Filipe Azevedo, Carlos Martins. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.06.2018. 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.

Original Paper

Impact of a Web-Based Electronic Health Record on Behavioral Health Service Delivery for Children and Adolescents: Randomized Controlled Trial

Eric J Bruns¹, PhD; Alyssa N Hook¹, BS; Elizabeth M Parker¹, PhD, MHS; Isabella Esposito¹, BS; April Sather¹, MPH; Ryan M Parigoris¹, BA; Aaron R Lyon¹, PhD; Kelly L Hyde², PhD

¹Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States

²FidelityEHR, Santa Fe, NM, United States

Corresponding Author:

Eric J Bruns, PhD

Department of Psychiatry and Behavioral Sciences

University of Washington School of Medicine

2815 Eastlake Ave E

Suite 200

Seattle, WA, 98102

United States

Phone: 1 206 685 2477

Fax: 1 206 685 3430

Email: ebruns@uw.edu

Abstract

Background: Electronic health records (EHRs) have been widely proposed as a mechanism for improving health care quality. However, rigorous research on the impact of EHR systems on behavioral health service delivery is scant, especially for children and adolescents.

Objective: The current study evaluated the usability of an EHR developed to support the implementation of the Wraparound care coordination model for children and youth with complex behavioral health needs, and impact of the EHR on service processes, fidelity, and proximal outcomes.

Methods: Thirty-four Wraparound facilitators working in two programs in two states were randomized to either use the new EHR (19/34, 56%) or to continue to implement Wraparound services as usual (SAU) using paper-based documentation (15/34, 44%). Key functions of the EHR included standard fields such as youth and family information, diagnoses, assessment data, and progress notes. In addition, there was the maintenance of a coordinated plan of care, progress measurement on strategies and services, communication among team members, and reporting on services, expenditures, and outcomes. All children and youth referred to services for eight months (N=211) were eligible for the study. After excluding those who were ineligible (69/211, 33%) and who declined to participate (59/211, 28%), a total of 83/211 (39%) children and youth were enrolled in the study with 49/211 (23%) in the EHR condition and 34/211 (16%) in the SAU condition. Facilitators serving these youth and families and their supervisors completed measures of EHR usability and appropriateness, supervision processes and activities, work satisfaction, and use of and attitudes toward standardized assessments. Data from facilitators were collected by web survey and, where necessary, by phone interviews. Parents and caregivers completed measures via phone interviews. Related to fidelity and quality of behavioral health care, including Wraparound team climate, working alliance with providers, fidelity to the Wraparound model, and satisfaction with services.

Results: EHR-assigned facilitators from both sites demonstrated the robust use of the system. Facilitators in the EHR group reported spending significantly more time reviewing client progress ($P=.03$) in supervision, and less time overall sending reminders to youth/families ($P=.04$). A trend toward less time on administrative tasks ($P=.098$) in supervision was also found. Facilitators in both groups reported significantly increased use of measurement-based care strategies overall, which may reflect cross-group contamination (given that randomization of staff to the EHR occurred within agencies and supervisors supervised both types of staff). Although not significant at $P<.05$, there was a trend ($P=.10$) toward caregivers in the EHR group reporting poorer shared agreement on tasks on the measure of working alliance with providers. No other significant between-group differences were found.

Conclusions: Results support the proposal that use of EHR systems can promote the use of client progress data and promote efficiency; however, there was little evidence of any impact (positive or negative) on overall service quality, fidelity, or client satisfaction. The field of children's behavioral health services would benefit from additional research on EHR systems using designs that include larger sample sizes and longer follow-up periods.

Trial Registration: ClinicalTrials.gov NCT02421874; <https://clinicaltrials.gov/ct2/show/NCT02421874> (Archived by WebCite at <http://www.webcitation.org/6yyGPJ3NA>)

(*J Med Internet Res* 2018;20(6):e10197) doi:[10.2196/10197](https://doi.org/10.2196/10197)

KEYWORDS

mental health; medical informatics; electronic health records; child; adolescent; integrated care; care coordination

Introduction

Background

Electronic health record (EHR) systems are a type of health information technology (HIT) that has been widely proposed as a mechanism for improving the quality and positive impact of health care services [1-4]. Research suggests that a well-implemented and fully-integrated EHR systems can promote complete record-keeping and more efficient access to documentation, facilitating information sharing and better coordination of care [1,5,6]. Other proposed, but less well-validated, benefits of EHR systems include: (1) facilitating the use of standardized assessments that can promote progress monitoring, (2) better linkage to evidence-based interventions, (3) more effective communication between providers and supervisors, and (4) use of data to promote quality improvement and research [7].

Given the potential benefits, the use of EHR technology in healthcare has been a high policy priority for well over a decade, as evidenced by enabling legislation such as the Health Information Technology for Economic and Clinical Health (HITECH) Act [8], which authorized incentive payments through Medicare and Medicaid to eligible providers. Accordingly, research on EHRs in general healthcare has proliferated over the past decade. Research has examined rates of uptake of EHR systems and related HIT across healthcare settings [9-11], illuminated factors related to adoption and perceived usability [6,9], and enumerated barriers, challenges, and strategies to promote implementation [4-7,12,13]. Importantly, research has also examined impacts of EHR adoption, with comprehensive reviews showing a mix of positive, negative, and null outcomes. In general, studies have found structural and process benefits, such as productivity and work practices, but less impact on clinical outcomes [14-16].

Electronic Health Records in Behavioral Healthcare

In contrast to general healthcare, EHRs in behavioral health (ie, substance abuse and mental health services) has lagged substantially in both policy and research. Behavioral health providers were excluded from incentive programs such as those promoted by HITECH, rendering most behavioral healthcare providers unable to qualify for incentive payments [17]. Thus, it is not surprising that utilization of EHR systems in specialized behavioral health settings and addiction treatment centers is still quite limited [14], with a 2012 study finding full EHR

adoption in only about 20 percent of 505 behavioral health organizations [11].

Research on behavioral health information technology in general—and EHR implementation and impacts specifically—is also sparse by comparison to general healthcare [18,19]. Research that does exist has tended to find parallels to general healthcare. For example, the most commonly implemented EHR components for behavioral healthcare include maintaining documentation on clients and services provided, billing, scheduling, and clinic-wide reporting [5,6,17]. Functions such as information exchange, progress monitoring, and quality assurance—components that are arguably most likely to directly impact the content and quality of services delivered—were endorsed less frequently [14,20].

Barriers to behavioral health providers' EHR adoption and implementation also have been found to parallel those for general healthcare providers, with financial barriers related to procuring and maintaining EHRs most prominent, but also including issues related to technical support, lack of enthusiasm among providers, and the time and effort required for training and implementation [14,19,21,22]. Unique concerns have also been raised, such as poor alignment with existing behavioral health workflows, lack of fit between the types of information maintained by behavioral health providers (which may be more narrative) and typical EHR structures and functions (which are often more quantitative and categorical), and negative impacts on provider–client communication that may impede therapeutic alliance [14,23-25]. Although some studies have found that behavioral health providers prefer using EHR systems over paper records [25], others have found relatively low rates of satisfaction with the usability and helpfulness of EHR systems and the need for “work arounds” [22,26-28]. For example, a survey of 46 children's behavioral health providers conducted by our research team (personal communication with Coldiron, Hensley, and Hadfield, 2018) found a mean (SD) System Usability Scale (SUS) [29] score of only 48.4 (22.7) for the EHRs being used in organizations, well below the cut-off for acceptable (mean score of 65) or even “marginal” (mean score of 50) [30].

In sum, EHR systems continue to be promoted in behavioral healthcare as a potential means of improving practice efficiency and effectiveness. However, behavioral health-focused EHR systems lag substantially behind those for general healthcare in enabling policy and subsequent adoption, and behavioral health providers tend to be more skeptical of benefits than healthcare

providers. Research is also scant by comparison, with most research to date focused on rates of adoption and barriers to EHR use. One recent review concluded that “comparative studies exploring EHR implementation within behavioral health settings are currently absent in the literature” [25]. While a few studies have suggested EHRs may promote better coordination among primary care and behavioral health providers [9,18], little research is available to shed light on EHR systems’ impact on practice, process, and client outcomes.

Electronic Health Record Systems and Care Coordination

One area of behavioral healthcare that may especially benefit from an expansion of the EHR research base is care coordination for individuals with multiple and complex behavioral health needs. Effective care coordination requires a range of practitioner communication, service provision, and administrative activities with the potential to be facilitated by technology. Examples include: sharing of information among providers, accessibility of records by clients and their families, such as in personalized health records [10], access to a diverse provider registry, billing for multiple services and strategies, and cost and outcomes monitoring at the client, program, and system levels [31,32].

Research is now emerging that demonstrates EHRs’ potential for positive impact within coordinated care models. For example, Matiz and colleagues [33] found that enhancements to the EHR that added a care plan template were associated with a fourfold increase in care plan use. King and colleagues found that EHR use was associated with physicians’ adherence to research-based care coordination processes [34]. And Hsiao et al found that physicians using EHR were more likely to receive patient information needed for care coordination than those who did not [35].

While the above studies underscore the potential for EHR systems and other types of HIT to facilitate implementation of effective care coordination, none focused on behavioral healthcare, and none used an experimental design. Overall, despite the potential implications for decision-making among providers, managed care entities, and state behavioral health authorities, research is limited regarding how EHR adoption may affect implementation quality, client satisfaction, and adherence to defined practice models.

The Current Study

In the current study, we examined usability, and short-term impacts of an EHR developed to support the implementation of care coordination for children and youth with complex behavioral health needs and their families using the Wraparound process [36,37]. This EHR software was found in development studies to have adequate usability under controlled conditions [32]. In this study, we conducted a randomized pilot test of the EHR, assigning Wraparound facilitators working in two provider organizations across two states to either use the new EHR (19/34, 56%) or continue to implement Wraparound services as usual (SAU) using paper-based documentation (15/34, 44%).

In our research we sought to determine the following: (1) to understand providers’ perceptions of the EHR’s feasibility, acceptability, and contextual appropriateness in the “real world” of implementing Wraparound care coordination, and (2) comparing Wraparound facilitators randomly assigned to use the EHR versus paper-based SAU, determine how EHR implementation affected relevant work practices and service processes, such as supervision, fidelity to the Wraparound practice model, collection and use of progress data, teamwork and alliance, and parent satisfaction with care.

Methods

Overview of Study Design

The study was conducted in two sites. Site 1 was a Wraparound agency located in a diverse, largely rural region of a Southeastern US state. Site 2 was a regional mental health center providing Wraparound and other services in a small, predominantly white city and surrounding region in a Midwestern US state.

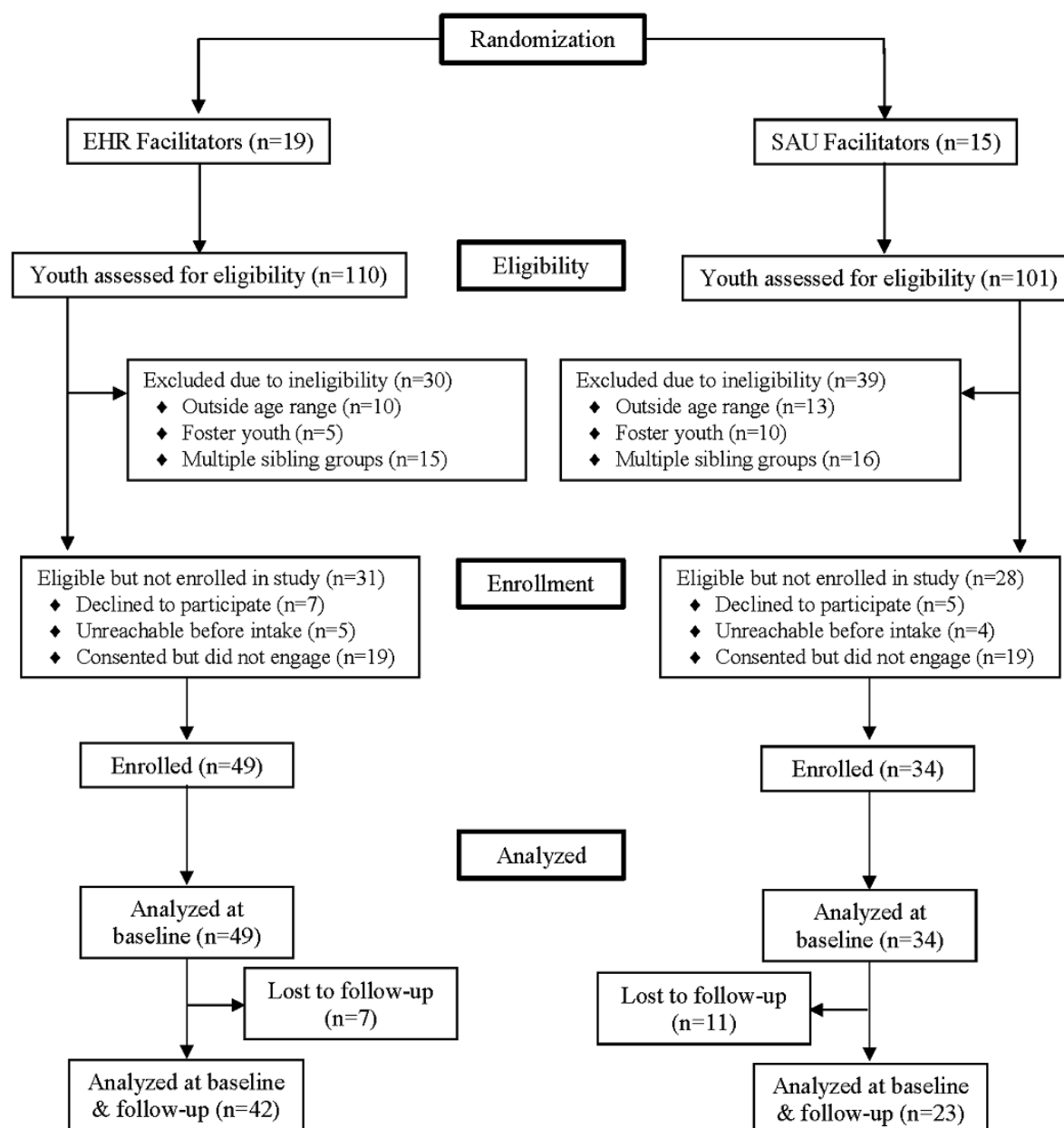
The study employed a blocked randomized control design with Wraparound facilitators (typically Bachelor’s or Master’s level mental health practitioners). Wraparound facilitators (also care coordinators) were randomly assigned to two conditions, EHR or SAU. A pool of 34 (29 in Site 1 and five in Site 2) randomized facilitators were stratified by the two sites and five supervisors (three in Site 1 and two in Site 2) to balance clustering effects. Randomization was conducted by the independent academic partner at the University of Washington.

All facilitators continued to provide Wraparound care coordination as they did before the study, with one exception: Facilitators assigned to the EHR condition were trained and supported to use an online EHR software package (see below for details). Facilitators not assigned to use the EHR continued to provide SAU. All supervisors were also trained to use the EHR and were encouraged to use the system when supervising facilitators in the EHR group. However, it is important to note that all supervisors were asked to supervise facilitators in both study groups.

Participants

Youth

To be eligible for the study, children and youth had to be between 5 and 18 years old and experiencing serious emotional and behavioral disturbance, defined as having a mental health diagnosis as designated in the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [38] and functional impairment that “substantially interferes with or limits the child from developing social, behavioral, cognitive, communicative or adaptive skills or his activities relating to family, school or community.” Youth in foster care were not eligible for the study due to issues of obtaining consent for youth in state custody. Youth in multiple sibling groups referred for services were also not eligible due to clustering effects and subsequent difficulty in interpreting results for such families.

Figure 1. Participant flow through the study.

The study was initiated in November 2015 in Site 1 and January 2016 in Site 2. All 34 facilitators and five supervisors in both sites consented to participate. For eight months after study inception, 211 children and youth enrolled in Wraparound in the two sites were referred to the study. Of those, 69/211 (33%) were found to be ineligible (31/69 [45%] due to being members of sibling groups, 15/69 [22%] due to being foster youth, and 23/69 [33%] due to being out of the age range). Of those remaining, 12/211 (6%) declined to participate, 38/211 (18%) consented to be contacted but did not respond to outreach from the research team, and 9/211 (4%) consented to be in the study but were not responsive to requests to conduct an intake interview. Thus, 83/211 (39%) children and youth and their caregivers were formally enrolled in the study, 49/83 (59%) served by facilitators in the EHR condition and 34/83 (41%) served by facilitators in the SAU condition. Of these, 18/83

(22%) were lost to data collection follow up leaving a final sample of 65/83 (78%) children and youth for whom longitudinal data were available, 42/65 (65%) in the EHR group and 23/65 (35%) in the SAU group. A detailed Consolidated Standards of Reporting Trials (CONSORT) diagram is provided in [Figure 1](#).

As shown in [Table 1](#), the group of children and youth on which analyses were conducted was majority male (54/83, 65%), with a mean age of 11.4 (SD 3.73) years. Approximately half (42/83, 51%) were from a racial or ethnic minority group (African American (37/83, 45%), mixed race (5/84, 6%), and 1/83 (1%) of Hispanic ethnicity). The most common Axis I diagnoses across children and youth were attention disorders (33/83, 40%), mood disorders (20/83, 24%), oppositional and conduct disorders (11/83, 13%), and anxiety disorders including PTSD (11/83, 13%).

Table 1. Baseline characteristics of children and youth, caregivers, and facilitators by study group.

Characteristics	EHR ^a	SAU ^b	Total
Youth, n	49	34	83
Age in years, mean (SD)	11.35 (3.69)	11.56 (3.85)	11.43 (3.73)
Female, n (%)	18 (36.73)	11 (32.35)	29 (34.94)
Race, n (%)			
African American	19 (38.78)	18 (52.94)	37 (44.58)
White	27 (55.10)	14 (41.18)	41 (49.40)
Mixed	3 (6.12)	2 (5.88)	5 (6.02)
Repeated a grade, n (%)	19 (40.43)	12 (35.29)	31 (38.27)
Ever been in foster care, n (%)	9 (18.37)	6 (17.65)	15 (18.07)
Brief Problem Checklist-Total problem score	14.02 (4.44)	13.56 (4.78)	13.83 (4.56)
Strengths and Difficulties Questionnaire-Total Score	21.26 (7.22)	21.47 (5.95)	21.94 (6.71)
Caregiver, n	49	34	83
Age in years, mean (SD)	38.49 (8.47)	39.32 (11.19)	38.83 (8.47)
Female, n (%)	45 (91.84)	33 (97.1)	78 (93.98)
Race, n (%)			
African American	18 (36.73)	17 (50)	35 (42.17)
White	30 (61.22)	17 (50)	47 (56.63)
Other	1 (2.04)	0	1 (2.04)
Adjusted gross income (US), n (%)			
< \$19,000	26 (54.17)	23 (39.70)	49 (60.49)
\$20,000-\$39,000	15 (31.25)	7 (21.21)	22 (27.16)
>\$40,000	7 (14.58)	3 (9.09)	10 (12.35)
Relationship to youth, n (%)			
Biological parent	37 (75.51)	26 (76.47)	63 (75.90)
Adoptive parent	4 (8.16)	1 (2.94)	5 (6.02)
Grandparent	3 (6.12)	5 (14.17)	8 (9.64)
Other	5 (10.20)	2 (5.88)	7 (8.43)
Facilitator, n	18	13	31
Female, n (%)	11 (61.1)	11 (84.6)	22 (71)
Race, n (%)			
African American	5 (27.8)	2 (15.4)	7 (22.6)
White	12 (66.7)	10 (76.9)	22 (71)
Hispanic	0	1 (7.7)	1 (3.2)
Other	1 (5.6)	0	1 (3.2)

^aEHR: electronic health records.^bSAU: services as usual.

A total of 29/83 (35%) children and youth had more than one Axis I disorder, and 63/79 (80%) scored in the clinical range for the Total Difficulties Score on the Strengths and Difficulties Questionnaire [39]. As shown in Table 1, there were no significant differences between groups at baseline on any of these measures.

Caregivers

Across both groups, 63/83 (76%) of children and youth were cared for by biological parents, 8/83 (10%) by a grandparent, 7/83 (8%) by other individuals (e.g., a family friend), and 5/83 (6%) by adoptive parents. A large majority of caregivers were female (78/8, 94%); 47/83 (57%) were white and 35/83 (42%) were African American. A majority of caregivers had a

household income of less than US \$19,000 (49/81, 60%). As shown in Table 1, there were no between-group differences on any variables at baseline.

Facilitators

Three facilitators were lost to attrition in Site 1 (all before study youth were assigned to them); thus, children and youth in the study were served by a total of 31 Wraparound facilitators, 26/31 (84%) in Site 1 and 5/31 (16%) in Site 2. A majority were female (22/31, 71%) and white 22/31 (71%), while 7/31 (23%) were African-American. As shown in Table 1, there were no differences in demographics between facilitators at baseline.

Intervention Conditions

Electronic Health Record Condition

Facilitators assigned to the EHR condition used an online software system that was developed through a partnership between a university research team and a small behavioral health-focused software developer. In addition to standard EHR fields (eg, youth and family information, diagnoses, assessment data, progress notes), the software maintained information on all elements of the Wraparound team and Wraparound plan in formats that align with the defined practice model for Wraparound care coordination [40,41]. For example, the software is organized via tabs that correspond to the sequence of activities that engage the family and build a plan that serves as the focus of coordinated Wraparound teamwork. Examples of information entered and maintained include the family's background and the reason for referral, youth and family strengths, a team mission statement, and priority need statements in the family's own words. Each need statement is connected to specific strategies and one or more outcomes statements on which data must be entered over time. If a strategy is a billable service, the facilitator can enter the service, service provider information, and the number of units authorized.

Other functions supported by the EHR system include individualized permission levels that allow for the sharing of information among youth and families, providers, and other team members. As such, caregivers and youth had access to certain records within the system, such as meeting schedules, plans of care, and progress monitoring dashboards. Reporting functions include individual youth-, supervisor-, and administrator-level data aggregation and reporting on services, expenditures, and outcomes. Facilitator workflow is supported by a hyperlinked Task List that tracks the completion of necessary care coordination steps and tasks as well as completion of required fields and elements of the Wraparound plan. The EHR also sends system-generated emails that obtain electronic signatures and automated reminders for upcoming meetings. Finally, the system promotes outcomes monitoring and feedback via collection of data on progress toward youth and family needs statements (eg, on a 0-10 scale). Brief process (eg, connection to professional helpers and social supports) and standardized outcomes measures are also incorporated into the system.

Although all efforts on the part of Wraparound care coordinators were completed within the EHR system, documentation by other involved health professionals (eg, primary or specialty care

physicians, child welfare case workers, mental health therapists) was not completed within the same EHR system. Evaluation reports, medical records, and other documentation can, however, be uploaded to the record via secure upload. See work published by Bruns and colleagues [32] for more details on the system.

Facilitators in the EHR group were trained on and supported to use the software via a sequence of activities that included: (1) an online training; (2) a two-day in-person training from the software developer's training team; and (3) monthly web-based check-in calls with five small cohorts of facilitators organized by supervisor. For two months after initial training (but before enrollment of study families), EHR-assigned facilitators were supported to continue learning the functions of the EHR with the two youth/families on their caseloads who were most recently enrolled in services. EHR users also had the availability of help desk support. The research team sent regular reports of EHR system use and data completeness for study enrolled families to facilitators and supervisors to help encourage full use of the system.

Services as Usual Condition

Facilitators in the control group completed research measures as described below, but did not participate in the training or use of the EHR. Rather, SAU-assigned facilitators maintained documentation, as usual, involving traditional paper case files. For facilitators in the SAU group, intake paperwork, progress notes, Wraparound plans, meeting minutes, and assessments all continued to be typed and hand-written and stored in a paper file and/or Excel files. Supervisors of SAU facilitators continued to review information on family needs, plans, and progress using paper and Excel files in their management and supervision.

Measures

Electronic Health Record System Activity

The research team monitored use of the software by EHR-assigned facilitators and reviewed activity logs by the facilitator in monthly consultation. The research team also compiled and fed these data back in initial months of the study to ensure the system was being used as intended by EHR group members. The activity monitor recorded each movement the user made within the system (ie, "Visited Custom Assessment Report Page" or, "Visited Add/Edit User Page") to capture how facilitators were utilizing the system. These data were then aggregated into categories (ie, "Maintaining Service Notes" or, "Updating & Developing the Plan of Care") to assess the percentage of time users were spending on each type of function in the EHR.

Demographic Information

The *Family Information Form* obtains data on youth and family demographics (eg, age; gender, and race of children/youth and caregivers; family income), diagnosis, and other information related to the family's history and home composition. The survey was administered to caregivers at baseline.

Provider Perceptions of Electronic Health Records

All supervisors and EHR-assigned facilitators completed two measures of EHR usability, acceptability, feasibility, and appropriateness six months after initiation of the study and

training on the system. The *System Usability Scale (SUS)* is a widely used, 10-item measure of perceptions of the usability of a technology system developed by Brooke [42,43]. Items such as, “I thought the system was easy to use,” and, “I felt very confident using the system” are rated on a five-point Likert scale. Resulting total scores range from 1-100. Scores below 50 indicate unacceptable usability, 50-70 indicate marginal usability, and greater than 70 indicate acceptable usability [29,42]. This scale is well-validated and has been found to have high inter-rater and test-retest reliability, excellent internal consistency ($\alpha=.91$) [44], and significant associations with alternative usability evaluation approaches [45].

The *System Acceptability and Appropriateness Scale (SAAS)* is an 11-item measure that evaluates HIT acceptability, utility and fit with service context. Items such as “How relevant is the technology to your client population?” are rated on a 1 (Not at All) to 5 (Extremely) point Likert scale and result in two subscale scores: Acceptability and Appropriateness. The SAAS was adapted from existing measures of intervention and HIT acceptability, including the Usage Rating Profile Intervention [46], Treatment Acceptability Rating Form-Revised (TARF-R) [47], and Intervention Rating Profile-15 [48]. Versions of the SAAS have been shown to possess acceptable technical adequacy ($\alpha>.70$) and criterion-related validity [49].

Provider Workflow and Behaviors

Four measures focused on provider workflow and behaviors. Two measures focused on supervision practices, one on facilitator attitudes toward standardized assessments, and one on facilitator behaviors related to measurement-based care.

The *Supervision Process Questionnaire (SPQ)* asked supervisors and facilitators to evaluate the percentage of time spent during supervision in nine different areas (eg, crisis assessment, client progress review, case conceptualization). Subscale or total scores were not calculated; instead, data were analyzed at the individual item level to evaluate between-group differences in supervision foci. Preliminary studies have found adequate interrater reliability [50].

The *Brief Supervision Practice Checklist-Adapted (BSPC)* is an eight-item survey administered to supervisors and facilitators that collects information on types of supervision practices and asks individuals in both roles to rate on a five-point scale (from Never to Almost Always) the degree to which different types of supervision practices are provided (eg, “supervisor discussed techniques to encourage family engagement;” and “supervisor reviewed youth and family progress”). Originally developed by Dorsey and colleagues [51] for supervision of clinicians practicing individual therapy, the measure was revised to better align with Wraparound care coordination. Formal reliability and validity have not been reported; however, internal consistency for the current sample was found to be good ($\alpha=.93$). Total scores were calculated for BSPC items, given that, unlike the SPQ, all items are proposed to evaluate a latent variable focused on effective Wraparound supervision. The SPQ and BSPC were administered to supervisors and facilitators in both groups at the initiation of the study (before training on the EHR), and six months later.

The *Attitudes Toward Standardized Assessment Scale (ASA)* is a 22-item measure of practitioner perceptions and attitudes about using standardized assessments in clinical practice. Items are scored on a 1 (Strongly Disagree) to 5 (Strongly Agree) scale and yield three subscales with adequate or better reliabilities: Benefit over Clinical Judgment, Psychometric Quality, and Practicality ($\alpha=.75$) [52]. Ratings have been associated with a greater likelihood of standardized assessment use. Facilitators in both groups completed the ASA at baseline and six-month time points.

The *Current Assessment Practice Evaluation-Revised (CAPER)* is a 10-item measure that assesses practitioners’ self-ratings of behaviors related to measurement-based care (MBC) across different phases of intervention (eg, at intake, ongoing during termination, discharge). As described in a recently submitted paper by Lyon and colleagues, CAPER subscales demonstrated good reliability as well as convergent and divergent validity with clinician attitudes about MBC in the expected directions (personal communication by Aaron Lyon, 2017). Facilitators responded to items such as “In the last two weeks, for how many youth/families did you administer a standardized assessment measure?” and “...for how many families did you systematically track an individualized outcome variable?” Facilitators in both groups completed the CAPER every other week for eight months, for a total of 16 biweekly surveys.

Wraparound Implementation and Service Process

Wraparound Implementation and Service Process was evaluated using three measures. The *Team Climate Inventory, short version (TCI)* is a 14-item survey that evaluates five relevant aspects of health care teamwork (Shared Vision, Participation Safety, Support for Innovation, Task Orientation, Interaction Frequency) using a five-point Likert scale. The scale has extensive support for reliability and factor structure; and validity is found in association with healthcare quality, patient satisfaction, and outcomes, including alpha coefficients of the subscales ranging from 0.73-0.80 [53].

The *Wraparound Fidelity Index, Brief Version (WFI-EZ)* is a widely-used, reliable and valid self-report measure of fidelity to the Wraparound process, based on the original Wraparound Fidelity Index, version 4 [54]. Items in the 25-item fidelity section of the measure are rated on the Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). The WFI-EZ yields scores for five theory- and research-based Wraparound practice domains (eg, Team-based, Outcomes-based, Family-driven) and a Total Score. Internal consistency for all items has been found to be good ($\alpha=.89$; personal communication with Ryan Parigoris, 2017). Evidence for validity includes differentiation among programs using a method of known groups approach as well as significant correlations between total fidelity scores and alternate measures of fidelity (personal communication with Ryan Parigoris, 2017) [55].

The *Working Alliance Inventory (WAI)* measures alliance between clinicians and clients on three domains: bond, goals, and tasks [56]. Based on the WAI short form, this measure was revised to reflect the alliance between Wraparound facilitators and families. Items are rated on a seven-point Likert scale, with response options ranging from Never to Always. The measure

results in an overall alliance score, as well as three subscale scores tied to the domains. Reliability has been found to be good for the client form [57], and adequate for provider versions [56].

Caregivers completed the TCI, WFI-EZ, and WAI four months after entry to Wraparound services.

Client Satisfaction

The *Client Satisfaction Questionnaire (CSQ)* is a widely used, well-validated measure of satisfaction with behavioral health services [58]. Items such as “How would you rate the quality of service your child received?” are rated on a four-point Likert scale ranging from Poor to Excellent. For this study, the eight-item short form (CSQ-8) was used, which has an internal consistency of .93 [59]. The CSQ-8 was administered to caregivers four months after entry to Wraparound services.

Facilitator Satisfaction

The *Therapist Satisfaction Index (TSI)* is a 14-item self-report measure to assess practitioners’ affinity for the intervention being used, perceived effectiveness, capacity for individualization and flexibility, and applicability to children and youth they work with. Items such as “The caregivers I work with seem to like the Wraparound approach” are rated on a five-point Likert scale from Strongly Disagree to Strongly Agree. Cronbach’s alpha for the total score has been found to be .83 [60]. A version with items revised to be appropriate for Wraparound was administered to facilitators at baseline and six months after study initiation.

Procedures

The study protocol was approved by the institutional review board at the University of Washington. Provider staff (supervisors and facilitators) were consented by the research coordinator after an on-site study introduction. Provider staff in both conditions completed EHR perception surveys and workflow and provider behavior instruments six months after training on the EHR. Enrollment of children, youth, and caregivers (and baseline interview completion) began two months after initial training. As described above, facilitators completed measures related to implementation specific to each enrolled youth and family four months after the child or youth was enrolled in services. Facilitators also completed a brief online survey about their use of measurement-based care (via the CAPER) bi-weekly for the duration of the study.

Intake coordinators at both study locations assigned all children and youth newly enrolled in the two Wraparound programs a study identification number and assessed them for eligibility. If determined to be eligible, Wraparound facilitators presented eligible youths’ parent or guardian with information about the study and sought to obtain consent to be contacted by the research team. After consent to contact was obtained, a member of the research team contacted the parent/guardian via phone and further explained the parameters of the research study, and formally enrolled those who agreed to participate. Interviews were conducted by a research assistant via phone at baseline and four months.

Data Analysis

Equivalence of groups at baseline was assessed using t-tests and chi-square tests. Differences between EHR and SAU in provider workflow, implementation and service processes, and client/facilitator satisfaction were examined using t-tests, and hierarchical linear models were also conducted to account for the nested nature of the data. Facilitators (level 1) were nested within supervisors (level 2), who were nested within site (level 3), or caregivers (level 1) were nested within facilitator (level 2), who were nested within site (level 3). To explore the impact of nesting, intraclass correlation coefficients (ICCs) at the supervisor and site levels were examined, both of which were very low (<0.05) for all major outcomes. Therefore, two-level hierarchical linear models with a random intercept for a site were run. To account for missing data, which ranged between 17/83 (20%) to 20/83 (24%), multivariate normal multiple imputations were used with 100 imputations. Auxiliary variables were included to aid the imputation. These analyses were conducted using Stata Version 13.1.

Longitudinal outcomes were tested through two-level growth curve models using HLM 7.0 [61] with observations/time (level 1) nested within facilitators (level 2). The data were also nested by site (level 3), but due to the low ICCs, a dummy variable was created and included in the model. Estimated scores and rates of change over time for the outcome variables were modeled. Random intercepts for facilitator were included and random slopes for observations/time were examined and retained when statistically significant. Data were modeled using full maximum likelihood estimation.

Although a large number of comparisons were made, we chose not to use a correction primarily because it would result in extremely small P values (or alphas). All tests were planned *a-priori* so we looked for consistency and examined patterns among the results. Also, because of small sample sizes and the exploratory nature of this study, we flagged results that trended toward significance (between-group differences at $P<.10$ level) for inclusion in the interpretation of results.

Results

Group Comparability

The EHR and SAU groups were compared at baseline on several demographic variables (see Table 1). As shown, there were no significant differences between groups on any variables, including total scores on two commonly used measures of child emotional and behavioral functioning, the Brief Problem Checklist [62] and the Strengths and Difficulties Questionnaire [39].

EHR System Activity

Table 2 presents a summary of EHR activity by facilitators in each site for months 1 and 2 when activity was recorded and fed back during EHR consultation with the research team. As shown, EHR facilitators from both sites demonstrated robust use of the system. Facilitators in Site 1 demonstrated a greater mean number of clicks during the first month at 1,473 (SD 61.45) and second month at 1,060 (SD 58.90) than facilitators from Site 2 with 866 (SD 26.43) in month 1 and 612 (SD 16.91)

in month 2. Use by facilitators in both sites was greater in month 1, during which it was necessary to transfer data from paper records for enrolled families. During the second month of use, nearly two-thirds (642/1060, 61%) of the clicks for facilitators in Site 1 were used in communicating with the team or updating and developing the plan of care. In contrast, a larger proportion of clicks for facilitators from Site 2 (395/612, 65%) were used in the system managing information and updating/maintaining service notes. Such differences were attributed to different organizational priorities and approaches to implementing Wraparound.

Provider Perceptions

Results from measures focused on practitioner perspectives on the EHR including the SUS and SAAS are summarized in [Table 3](#). Scores on the individual items of the SUS ranged from

1.6-2.7, with a total average score of 54.72 (range 30-70.3). Scores on the acceptability subscale ranged from 2.6-3.6 and scores on the appropriateness subscale ranged from 2.9-3.4.

Provider Workflow and Behaviors

Workflow outcomes were assessed using measures of supervision activity (SPQ and BSPC), use of measurement-based care (CAPER), and attitudes toward standardized assessment (ASA). Results are presented in [Table 4](#). There were no differences between groups on the ASA scale at baseline. Results from the HLM suggested facilitators in the EHR group reported lower scores on the psychometric quality subscale, on average, compared to those in the SAU group. Facilitator reports did not differ by treatment group for the “benefit over clinical judgment” and “practicality” subscales.

Table 2. Summary of system activity (number of clicks) by site and time for the first two months of the study.

EHR ^a function	Site 1 (n=18), mean (%)		Site 2 (n=13), mean (%)	
	Month 1	Month 2	Month 1	Month 2
Communicating with the team	305 (20.7)	431 (40.7)	138 (15.9)	15 (2.4)
Core Assessments	4 (0.3)	25 (2.4)	25 (2.8)	34 (5.6)
Maintaining service notes	177 (12.0)	162 (15.3)	129 (14.8)	142 (23.2)
Managing information	324 (22.0)	13 (1.2)	273 (31.5)	253 (41.4)
Updating and developing the Wraparound plan	396 (26.9)	211 (20.0)	136 (15.7)	76 (12.4)
User settings	266 (18.1)	217 (20.5)	167 (19.2)	92 (15.1)
Total	1473 (100)	1060 (100)	866 (100)	612 (100)

^aEHR: electronic health record.

Table 3. Electronic health record software acceptability, appropriateness, and usability at the six-month follow-up (n=18 facilitators).

Variable	Mean (SD)
System Usability Scale - Total usability score (0-100 scale)	54.72 (12.54)
System Acceptability and Appropriateness (0=lowest to 5=highest)	
Acceptability	
Satisfied with current version of the technology	2.83 (0.92)
Believe technology to be organized/well-constructed	3.11 (1.08)
Satisfied with content of technology system	2.83 (0.79)
Satisfied with the technology's overall ease of use	2.61 (0.92)
Comfortable interacting with the technology	3.56 (0.78)
The technology is intuitively appealing	3.17 (1.04)
Appropriateness	
The technology is compatible with agency's mission or service provision mandate	3.39 (0.70)
The technology is relevant to client population	3.11 (0.90)
The technology fits with current treatment modality, theoretical orientation, or skill set	3.33 (0.84)
The technology is compatible with workflow timing	2.94 (0.87)
The technology fits with overall approach to service delivery and the setting in which care is provided	3.17 (0.87)

Table 4. Summary of workflow outcomes by study group.

Variable	EHR ^a (n=18) facilitators, mean (SD)	SAU ^b (n=13) facilitators, mean (SD)	<i>P</i> value	Intercept coefficient (SE)	Intervention coefficient (SE)	Baseline coefficient (SE)
Attitudes towards Standardized Assessments						
Benefit Over Clinical Judgement	2.77 (0.64)	2.89 (0.33)	.52	2.63 (0.44) ^c	0.18 (0.16)	0.07 (0.15)
Psychometric Quality	3.35 (0.43)	3.37 (0.36)	.89	1.80 (0.47) ^c	−0.30 (0.10) ^d	0.50 (0.13) ^c
Practicality	3.21 (0.39)	3.06 (0.31)	.28	2.45 (0.56) ^c	−0.13 (0.12)	0.23 (0.17)
Brief Supervision Practice Checklist						
Supervision Score	3.56 (0.18)	3.32 (0.27)	.45	0.53 (0.47)	0.09 (0.22)	0.76 (0.13) ^c
Supervision Process Questionnaire						
Administrative tasks	8.78 (6.59)	11.53 (7.47)	.29	9.38 (2.26) ^c	−3.36 (2.03) ^c	0.19 (0.15)
Facilitator personal support	10 (6.89) ^f	3.46 (4.74) ^f	.01	5.22 (1.32) ^c	1.42 (1.84)	0.16 (0.14)
Reviewing progress toward needs	4.44 (4.08)	5.77 (4)	.38	6.29 (1.62) ^c	3.33 (1.56) ^f	0.11 (0.19)
Skills coaching and training	12.72 (8.90)	17.23 (9.82)	.19	11.51 (2.78) ^c	−2.32 (2.36)	0.25 (0.13) ^f
Reviewing plans of care	15.33 (7.14)	16.54 (5.55)	.62	5.16 (3.27)	2.42 (1.99)	0.47 (0.16) ^d
Crisis assessment management	8.67 (6.71)	6.62 (4.14)	.34	7.58 (1.38) ^c	−1.01 (1.46)	0.08 (0.13)
Case conceptualization	5.83 (2.96)	5.54 (4.10)	.82	4.30 (1.25) ^d	0.26 (1.12)	0.26 (0.17)
Youth family engagement	13.06 (2.71) ^f	9.46 (3.82) ^f	.01	6.58 (2.60) ^f	−0.22 (1.63)	0.42 (0.24) ^c
Natural support engagement	10.28 (4.36)	10.31 (4.75)	.99	11.07 (2.30) ^c	0.49 (1.66)	−0.14 (0.19)
Support relationship	5.33 (3.24) ^f	9.23 (4.00) ^f	.01	4.74 (3.19)	0.49 (1.39)	0.52 (0.22) ^f
Facilitator professional role	5.56 (3.38)	4.31 (2.81)	.29	3.14 (0.94) ^d	−0.54 (0.93)	0.43 (0.15) ^d

^aEHR: electronic health record.^bSAU: services as usual.^c*P*<.001^d*P*<.01^e*P*<.10^f*P*<.05

As shown, there were no differences between groups on the BSPC score at baseline. However, there were differences on the SPQ. At baseline, facilitators in the EHR group reported greater average scores on the facilitator personal support and youth family engagement subscales compared to facilitators in the SAU group. In comparison, facilitators in the EHR group reported lower average scores on the support relationship subscale compared to the SAU group. In the HLM, reports on the BSPC did not differ when comparing facilitators in the EHR group to the SAU group. On the SPQ, facilitators in the EHR group reported significantly higher scores on the reviewing progress toward needs subscale, compared to those in the SAU group. Additionally, facilitators in the EHR group reported lower scores on the administrative tasks subscale, on average,

compared to those in the SAU group (result approached significance, *P*<.10).

Wraparound Implementation and Service Process

Implementation and fidelity outcomes are presented in [Table 5](#). There were no significant differences between treatment groups across the WAI, TCI, WFI-EZ fidelity total score, or the CSQ (total score) variables at the four-month follow-up. Based on results of HLM, caregiver reports on these measures did not differ significantly for those in the EHR group compared to the SAU group over time. However, a trend (*P*=.10) was found whereby facilitators in the EHR group reported lower scores compared to those in the SAU group on the task subscale of the WAI.

Table 5. Summary of caregiver and facilitator-reported implementation and fidelity outcomes by study group.

Variable	EHR ^a , mean (SD)	SAU ^b , mean (SD)	<i>P</i> value	Intercept coefficient (SE)	Intervention coefficient (SE)	Baseline coefficient (SE)
Caregiver						
Team Climate Inventory						
Vision	17.10 (3.46)	17.57 (2.31)	.56	17.75 (0.73) ^c	−0.65 (0.88)	N/A ^d
Participative safety	18 (0.39)	17.30 (3.52)	.38	17.91 (0.83) ^c	−0.75 (0.80)	N/A
Task orientation	12.98 (2.50)	13.13 (2.16)	.80	13.24 (0.57) ^c	−0.25 (0.65)	N/A
Support innovation	12.88 (2.67)	13 (2.22)	.86	12.89 (0.72) ^c	−0.17 (0.66)	N/A
Working Alliance Inventory						
Goal	23.81 (6.39)	24.35 (4.16)	.72	24.39 (1.32) ^c	−0.68 (1.54)	N/A
Task	23.55 (6.60) ^e	26.04 (3.90) ^e	.1	25.77 (1.59) ^c	−2.66 (1.53) ^e	N/A
Bond	25.36 (5.55)	26.57 (3.46)	.35	26.23 (1.39) ^c	−1.29 (1.30)	N/A
Total score	72.71 (17.56)	76.96 (10.23)	.29	76.19 (4.24) ^c	−4.33 (4.22)	N/A
WFI-EZ Fidelity						
Outcomes	0.74 (0.26)	0.75 (0.12)	.95	0.73 (0.07) ^c	−0.01 (0.06)	N/A
Teamwork	0.70 (0.21)	0.66 (0.19)	.50	0.65 (0.06) ^c	0.03 (0.05)	N/A
Natural supports	0.58 (0.20)	0.62 (0.16)	.36	0.63 (0.04) ^c	−0.05 (0.05)	N/A
Needs	0.72 (0.20)	0.74 (0.12)	.68	0.74 (0.04) ^c	−0.03 (0.05)	N/A
Strengths	0.80 (0.20) ^e	0.71 (0.14) ^e	.09	0.71 (0.05) ^c	0.07 (0.05)	N/A
Parent and Child Satisfaction Scale						
Total score	3.40 (0.69)	3.44 (0.59)	.84	3.37 (0.20) ^c	−0.04 (0.18)	N/A
Facilitator						
Therapist Satisfaction Index						
Total score	2.77 (0.64) ^e	2.89 (0.33) ^e	.08	−3.06 (1.66) ^e	3.10 (2.14)	0.99 (0.20) ^f

^aEHR: electronic health record. n=42 and n=18 for caregiver and facilitator groups, respectively.

^bSAU: services as usual; multiple imputation (mi) used to handle missing data (mi=100 imputed datasets). n=23 and n=13 for caregiver and facilitator groups, respectively.

^c*P*<.001

^dN/A: not applicable.

^e*P*<.10

^f*P*<.01

Table 6. Time and facilitator level indicators of current practices as reported on the Current Assessment Practice Evaluation Revised.

Variable	Fixed effects coefficient	Standard error	Random effects, SD	Variance
Administered Standardized Assessment				
Intercept	13.26	17.08	N/A ^a	N/A
Month (L1) ^b	1.30 ^c	0.39	N/A	N/A
Intervention (L2) ^{d,e}	8.87	7.34	N/A	N/A
Site (L2) ^{d,f}	11.77	15	N/A	N/A
Facilitator	N/A	N/A	19.48 ^g	379.6
Month	N/A	N/A	1.62 ^h	2.62
Given Feedback about Assessment				
Intercept	18.27	15.83	N/A	N/A
Month (L1) ^b	1.40 ^g	0.36	N/A	N/A
Intervention (L2) ^{d,e}	7.68	6.78	N/A	N/A
Site (L2) ^{d,f}	6.29	14.13	N/A	N/A
Facilitator	N/A	N/A	17.44 ^g	304.2
Month	N/A	N/A	1.40 ^c	1.95
Systematically Tracked Outcome				
Intercept	44.69 ^c	12.47	N/A	N/A
Month (L1) ^b	0.87 ^h	0.35	N/A	N/A
Intervention (L2) ^{d, e}	-2.43	6.34	N/A	N/A
Site (L2) ^{d,f}	4.16	12.05	N/A	N/A
Facilitator	N/A	N/A	15.79 ^g	249.45
Month	N/A	N/A	1.16 ^h	1.34
Given Feedback on Outcome				
Intercept	51.13 ^c	14.27	N/A	N/A
Month (L1) ^b	0.76 ^h	0.34	N/A	N/A
Intervention (L2) ^{d,e}	-0.98	6.28	N/A	N/A
Site (L2) ^{d,f}	-2.63	13.71	N/A	N/A
Facilitator	N/A	N/A	16.26 ^g	264.51
Month	N/A	N/A	1.10 ^h	1.21
Plan of Care Altered Based on Assessment				
Intercept	42.95 ^g	8.57	N/A	N/A
Month (L1) ^b	1.07 ^g	0.29	N/A	N/A
Intervention (L2) ^{d,e}	2.61	4.99	N/A	N/A
Site (L2) ^{d,f}	-18.03 ^c	5.83	N/A	N/A
Facilitator	N/A	N/A	10.14 ^c	102.92
Month	N/A	N/A	0.95 ^h	0.91
Assessment Used to Choose Service				
Intercept	4.59	15.5	N/A	N/A

Variable	Fixed effects coefficient	Standard error	Random effects, SD	Variance
Month (L1) ^b	0.86 ^g	0.24	N/A	N/A
Intervention (L2) ^{d,e}	3.83	5.9	N/A	N/A
Site (L2) ^{d,f}	12.25	13.1	N/A	N/A
Facilitator	N/A	N/A	14.51 ^g	210.67
Month	N/A	N/A	0.62 ^h	0.38
Sent Reminders				
Intercept	69.97 ^g	12.86	N/A	N/A
Month (L1) ^b	0.47	0.4	N/A	N/A
Intervention (L2) ^{d,e}	-17.28 ^h	7.82	N/A	N/A
Site (L2) ^{d,h}	-3.82	8.79	N/A	N/A
Facilitator	N/A	N/A	21.59 ^g	466.26
Month	N/A	N/A	1.62 ^g	2.63

^aN/A: not applicable.

^bL1: Level 1 predictor.

^c $P < .01$

^dL2: Level 2 predictor.

^eIntervention: 0=control group (reference), 1=intervention group.

^fSite: 0=Site 1 (reference) 1=Site 2.

^g $P < .001$

^h $P < .05$

Time trends and other results from the growth curve models for the facilitator-completed CAPER are found in [Table 6](#). Significant linear time trends were found for six items, with increasing proportions of facilitators reporting administering standardized assessments, giving feedback about assessments, systematically tracking outcomes, altering plans of care based on assessments, giving feedback on outcomes, and using assessments to choose services. Regarding between-group differences, facilitators in the EHR group reported sending reminders to a significantly smaller proportion of families compared to those in the SAU group. Facilitator reports did not differ by treatment group for the remaining subscales.

Discussion

Principal Results

Research on the use of EHR systems in behavioral healthcare has lagged behind research in general healthcare, resulting in a dearth of empirical guidance around issues such as software design and the impact of EHR adoption on services. The current study attempted to fill gaps in the research base by asking whether care coordinators serving children and youth with complex behavioral health needs who were randomly assigned to use an EHR would demonstrate differences in service processes and service quality compared to providers using paper records. Results indicated that there were few such impacts. No between-group differences were found for fidelity to the Wraparound practice model, an overall working alliance among

practitioners and families, Wraparound team climate, parent satisfaction with care, or practitioner satisfaction with services.

At the same time, practitioners in the EHR group reported spending significantly more time reviewing and applying client progress data in supervision, and significantly less time on administrative tasks. This finding provides support to the proposal that use of EHR systems can facilitate greater attention to client progress and subsequent problem solving and is consistent with prior research indicating that digital feedback technologies can effectively support assessment-related provider behavior change [63,64]. This is an encouraging result given that “treating to target” is a commonly-cited principle of effective behavioral healthcare, and has been found to account for substantial variance in positive outcomes [65]. Also, results from the CAPER found that facilitators in the EHR group were significantly less likely to send reminders to enrolled clients. Given that reminders around meetings and appointments can be automatically undertaken by the EHR, it may be that EHR use reduced the need for facilitators to do these tasks manually, potentially freeing time for other tasks.

Results from the CAPER also showed significant increases among facilitators in both groups for collecting and using assessment and outcomes data, altering plans of care based on assessments, and using assessments to choose services. Although between-group effects were not found, leaders in the two agencies suggested that these significant increases may have been a result of the EHR influencing supervisor behavior with

facilitators in both groups, and peer influence among practitioners within the agency.

Not all significant results supported positive impacts of the EHR. First, there was a pattern of poorer scores for the EHR group on the WAI, including a trend toward significance ($P=.10$) on the subscale focused on agreement on tasks to achieve identified goals. While such findings may have been spurious given the number of statistical tests conducted, they also may indicate that the time and effort needed to integrate a new EHR into workflow compromised engagement and alliance between EHR-assigned facilitators and families.

Second, perceptions of psychometric quality of standardized assessments improved over time for the SAU group but not the EHR group, resulting in a significant between-group difference for this subscale. Although a subtle effect, it may be that these provider organizations' increased attention to use of measurement and measurement scales was received more enthusiastically among staff waiting to be trained on the EHR than the initial adopters, who were exposed to the day-to-day realities and challenges of a rapid training and implementation process on the EHR, as well as shifts in how supervision was conducted.

All the above findings must be interpreted within the context of practitioners' perceptions of usability and acceptability of this particular software package. Results from surveys suggested that staff perceived the software to be reasonably well-aligned with the Wraparound practice model and the day-to-day workflow of facilitators. Mean ratings of overall usability were, however, lower, scoring in the "marginal" range on the SUS. Qualitative feedback from staff assigned to the EHR condition indicates that at the launch of the study, the EHR had some functionality issues (eg, frequent timing out, multiple clicks required to execute simple but frequently required tasks) that compromised its ease of use. Although the study (and the larger federally-funded project within which the study was conducted) allowed such issues to be identified and addressed by programmers, usability issues at the outset may have compromised the capacity for the EHR to achieve its full proposed impact.

Future studies would benefit from an examination of the impact of usability of EHR systems (or impact across multiple stages of development or implementation of a single system) on outcomes. It is important to note that successful application of EHRs—and HIT in general—requires strategic implementation supports to be successfully applied [66-68]. The rapid timeframe for the current study meant that EHR training and initiation of youth/family study enrollment happened very quickly and with less development of readiness and local implementation support than may have been ideal. Although service quality and fidelity were not assessed, it is worth noting that six months after initiation of the current study, the larger of the two provider organizations introduced a refined variant of the EHR in another site with all its staff at once and with local staff who had participated in the current study leading the roll-out and supporting implementation. Mean SUS scores for this cohort of facilitators was 63.7, nearly 10 points higher than for the EHR-assigned group in the current study.

Limitations

The current study has several major limitations. It focused only on short-term (4 months) outcomes, and these were limited to the provider, workflow, and service variables. The actual impact on outcomes such as residential placement, symptoms, and functioning and family outcomes such as family functioning or caregiver strain were not assessed. The sample size was small, and over one-third of the initial sample of youth/caregivers was lost to follow-up, limiting our ability to detect significant differences. As described above, randomization at the site or supervisor level was not possible, meaning that between-group contamination (eg, in areas such as supervision style or activities or use of standardized assessment) may have occurred. This may also have compromised the study's ability to detect impacts of the EHR.

Finally, as described above, the funding mechanism for the study only provided one year for development and refinement of the system followed by a single year to undertake a randomized pilot study, hindering the usability and implementation of the EHR. Although research and experience suggest that practitioner perceptions of EHR system usability and provision of training and implementation support are often poor in behavioral health, this situation may reduce the generalizability of results. Conducting the study in the context of Wraparound facilitation, which consists of a relatively unique set of practice activities, also may limit generalizability to other service types.

Implications

Along with other subtypes of HIT, EHR systems have been increasingly proposed as a method to support service quality implementation support, functioning as a practitioner-facing implementation strategy that can help organize plans of care, provide reminders, and structure workflow and supervision [69-71]. Results of the current study suggest that even when implemented under unideal circumstances (eg, a randomized study within an organization), the introduction of an EHR may indeed facilitate measurable and beneficial shifts in practice, such as greater attention to measurement-based care. At the same time, results suggest that EHRs may give rise—at least initially—to measurable, if subtle, negative impacts, such as less capacity for practitioners to nurture engagement and alliance.

These findings align with findings from other studies [15,72] that work tasks can be influenced positively by EHR adoption. At the same time, this research also supports conclusions by other researchers that productivity, and presumably quality of care, may decrease after initial implementation of an EHR, primarily as a result of the implementation effort typically required, and that no less than one month may be required after transition to a new EHR before practitioners return to baseline productivity [18,73]. This is also consistent with research on the implementation of new practices in general, where an initial decrease in competence might be expected before providers building mastery of the innovation [74].

Research is needed that provides more rigorous tests of these associations, and that can unpack the underlying causes. In this

study, for example, it is unclear whether greater benefits of EHR implementation would have been found had the system featured greater usability at the outset, rather than still undergoing improvement during the study. Fleming and colleagues, for example, found that it took 12 months for overall productivity to rebound to baseline levels after installation of an EHR [18]. Although logistical and methodological challenges may arise, researchers conducting more robust controlled tests of EHR systems in the future may be advised to wait up to one-year post-implementation before assessing impacts.

Similarly, it is unclear whether more robust implementation support (and agency-wide versus partial implementation) may have resulted in different outcomes. Future studies may focus on these issues by using “hybrid trial” approaches that simultaneously consider—or experimentally manipulate—EHR usability and contextual fit, implementation strategies, and outcomes [75]. Given the level of prioritization of HIT generally and EHR specifically in behavioral healthcare—and the number of system resources and human capital being invested in these technologies—continued expansion of the research base on these topics would seem to be a critically important investment.

Acknowledgments

This study was funded by the National Institute of Mental Health (R42-MH95516; PI EJB). The authors would like to express their deep gratitude to the participating Wraparound organizations for their willingness to partner in this research.

Conflicts of Interest

KLH is the founder, CEO, and majority owner of FidelityEHR, the company that created and maintains the EHR used in the trial presented in this manuscript. The other authors have no conflicts to disclose.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V.1.6.1).

[PDF File (Adobe PDF File), 757KB - [jmir_v20i6e10197_app1.pdf](#)]

References

1. Chase HS, Mitrani LR, Lu GG, Fulgieri DJ. Early recognition of multiple sclerosis using natural language processing of the electronic health record. *BMC Med Inform Decis Mak* 2017 Feb 28;17(1):24 [FREE Full text] [doi: [10.1186/s12911-017-0418-4](#)] [Medline: [28241760](#)]
2. Francis DB, Cates JR, Wagner KPG, Zola T, Fitter JE, Coyne-Beasley T. Communication technologies to improve HPV vaccination initiation and completion: A systematic review. *Patient Educ Couns* 2017 Jul;100(7):1280-1286. [doi: [10.1016/j.pec.2017.02.004](#)] [Medline: [28209248](#)]
3. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013 Jan 1;20(1):144-151 [FREE Full text] [doi: [10.1136/amiainl-2011-000681](#)] [Medline: [22733976](#)]
4. US Department of Health and Human Services, Office of the National Council of Health Information Technology. Using Information Technology to Integrate Behavioral Health and Primary Care. 2012. URL: https://www.healthit.gov/sites/default/files/bh-roundtable-findings-report_0.pdf [accessed 2017-12-07] [WebCite Cache ID 6z9H3pc6Y]
5. Bardach SH, Real K, Bardach DR. Perspectives of healthcare practitioners: An exploration of interprofessional communication using electronic medical records. *J Interprof Care* 2017 May;31(3):300-306. [doi: [10.1080/13561820.2016.1269312](#)] [Medline: [28151026](#)]
6. Elias B, Barginere M, Berry PA, Selleck CS. Implementation of an electronic health records system within an interprofessional model of care. *J Interprof Care* 2015;29(6):551-554. [doi: [10.3109/13561820.2015.1021001](#)] [Medline: [25955512](#)]
7. Jensen RE, Chan KS, Weiner JP, Fowles JB, Neale SM. Implementing electronic health record-based quality measures for developmental screening. *Pediatrics* 2009 Oct;124(4):e648-e654. [doi: [10.1542/peds.2008-3091](#)] [Medline: [19786425](#)]
8. Chiu S, Fitzgerald KM. Electronic medical/health record and pediatric behavioral health providers: progress and problems. *Arch Psychiatr Nurs* 2013 Apr;27(2):108-109. [doi: [10.1016/j.apnu.2013.01.001](#)] [Medline: [23540522](#)]
9. Lowes LP, Noritz GH, Newmeyer A, Embi PJ, Yin H, Smoyer WE, 'Learn from Every Patient' Study Group. 'Learn From Every Patient': implementation and early results of a learning health system. *Dev Med Child Neurol* 2017 Feb;59(2):183-191 [FREE Full text] [doi: [10.1111/dmcn.13227](#)] [Medline: [27545839](#)]
10. Ose D, Kunz A, Pohlmann S, Hofmann H, Qreini M, Krisam J, et al. A Personal Electronic Health Record: Study Protocol of a Feasibility Study on Implementation in a Real-World Health Care Setting. *JMIR Res Protoc* 2017 Mar 02;6(3):e33 [FREE Full text] [doi: [10.2196/resprot.6314](#)] [Medline: [28254735](#)]
11. The National Council for Community Behavioral Healthcare. 2012. HIT Adoption and Readiness for Meaningful Use in Community Behavioral Health: Report on the 2012 National Council Survey URL: <https://www.thenationalcouncil.org/wp-content/uploads/2012/10/HIT-Survey-Full-Report.pdf> [accessed 2017-12-05] [WebCite Cache ID 6z9DqXS0s]

12. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol* 2017 Jan;106(1):1-9 [FREE Full text] [doi: [10.1007/s00392-016-1025-6](https://doi.org/10.1007/s00392-016-1025-6)] [Medline: [27557678](#)]
13. Sakata KK, Stephenson LS, Mulanax A, Bierman J, McGrath K, Scholl G, et al. Professional and interprofessional differences in electronic health records use and recognition of safety issues in critically ill patients. *J Interprof Care* 2016 Sep;30(5):636-642. [doi: [10.1080/13561820.2016.1193479](https://doi.org/10.1080/13561820.2016.1193479)] [Medline: [27341177](#)]
14. Cellucci LW, Cellucci T, Stanton M, Kerrigan D, Madrake M. Current status and future directions of EMR use in psychology clinics. *Health Policy and Technology* 2015 Jun;4(2):91-99. [doi: [10.1016/j.hlpt.2015.02.005](https://doi.org/10.1016/j.hlpt.2015.02.005)]
15. Lau F, Price M, Boyd J, Partridge C, Bell H, Raworth R. Impact of electronic medical record on physician practice in office settings: a systematic review. *BMC Med Inform Decis Mak* 2012 Feb 24;12:10 [FREE Full text] [doi: [10.1186/1472-6947-12-10](https://doi.org/10.1186/1472-6947-12-10)] [Medline: [22364529](#)]
16. Holroyd-Leduc JM, Lorenzetti D, Straus SE, Sykes L, Quan H. The impact of the electronic medical record on structure, process, and outcomes within primary care: a systematic review of the evidence. *J Am Med Inform Assoc* 2011;18(6):732-737 [FREE Full text] [doi: [10.1136/amiainl-2010-000019](https://doi.org/10.1136/amiainl-2010-000019)] [Medline: [21659445](#)]
17. Larrison CR, Xiang X, Gustafson M, Lardiere MR, Jordan N. Implementation of Electronic Health Records Among Community Mental Health Agencies. *J Behav Health Serv Res* 2018 Jan;45(1):133-142. [doi: [10.1007/s11414-017-9556-9](https://doi.org/10.1007/s11414-017-9556-9)] [Medline: [28439789](#)]
18. Fleming NS, Becker ER, Culler SD, Cheng D, McCorkle R, da GB, et al. The impact of electronic health records on workflow and financial measures in primary care practices. *Health Serv Res* 2014 Feb;49(1 Pt 2):405-420 [FREE Full text] [doi: [10.1111/1475-6773.12133](https://doi.org/10.1111/1475-6773.12133)] [Medline: [24359533](#)]
19. Ser G, Robertson A, Sheikh A. A qualitative exploration of workarounds related to the implementation of national electronic health records in early adopter mental health hospitals. *PLoS One* 2014;9(1):e77669 [FREE Full text] [doi: [10.1371/journal.pone.0077669](https://doi.org/10.1371/journal.pone.0077669)] [Medline: [24454678](#)]
20. Lutterman T, Phelan B, Berhane A, Shaw R, Rana V. Substance Abuse and Mental Health Services Administration. Rockville, MD; 2008. Characteristics of state mental health agency data systems URL: <https://store.samhsa.gov/shin/content/SMA08-4361/SMA08-4361.pdf> [accessed 2018-05-03] [WebCite Cache ID 6z9EWFC8E]
21. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. *BMC Health Serv Res* 2010;10:231 [FREE Full text] [doi: [10.1186/1472-6963-10-231](https://doi.org/10.1186/1472-6963-10-231)] [Medline: [20691097](#)]
22. Cifuentes M, Davis M, Fernald D, Gunn R, Dickinson P, Cohen DJ. Electronic Health Record Challenges, Workarounds, and Solutions Observed in Practices Integrating Behavioral Health and Primary Care. *J Am Board Fam Med* 2015 Oct;28 Suppl 1:S63-S72 [FREE Full text] [doi: [10.3122/jabfm.2015.S1.150133](https://doi.org/10.3122/jabfm.2015.S1.150133)] [Medline: [26359473](#)]
23. Kobus AM, Harman JS, Do HD, Garvin RD. Challenges to depression care documentation in an EHR. *Fam Med* 2013 Apr;45(4):268-271. [Medline: [23553091](#)]
24. Lawlor T, Barrows E. Behavioral health electronic medical record. *Psychiatr Clin North Am* 2008 Mar;31(1):95-103. [doi: [10.1016/j.psc.2007.11.009](https://doi.org/10.1016/j.psc.2007.11.009)] [Medline: [18295041](#)]
25. Matthews EB. Integrating the Electronic Health Record into Behavioral Health Encounters: Strategies, Barriers, and Implications for Practice. *Adm Policy Ment Health* 2017 Jul;44(4):512-523. [doi: [10.1007/s10488-015-0676-3](https://doi.org/10.1007/s10488-015-0676-3)] [Medline: [26208693](#)]
26. Bruns EJ, Pullmann MD, Sather A, Denby BR, Ramey M. Effectiveness of wraparound versus case management for children and adolescents: results of a randomized study. *Adm Policy Ment Health* 2015 May;42(3):309-322 [FREE Full text] [doi: [10.1007/s10488-014-0571-3](https://doi.org/10.1007/s10488-014-0571-3)] [Medline: [24973891](#)]
27. Schurer Coldiron J, Bruns EJ, Quick H. A Comprehensive Review of Wraparound Care Coordination Research, 1986–2014. *J Child Fam Stud* 2017 Jan 28;26(5):1245-1265. [doi: [10.1007/s10826-016-0639-7](https://doi.org/10.1007/s10826-016-0639-7)]
28. McGregor B, Mack D, Wrenn G, Shim RS, Holden K, Satcher D. Improving Service Coordination and Reducing Mental Health Disparities Through Adoption of Electronic Health Records. *Psychiatr Serv* 2015 Sep;66(9):985-987 [FREE Full text] [doi: [10.1176/appi.ps.201400095](https://doi.org/10.1176/appi.ps.201400095)] [Medline: [25975885](#)]
29. Brooke J. System usability scale (SUS). Usability Evaluation in Industry. London: Taylor and Francis; 1996.
30. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 2009;4(3):114-123 [FREE Full text]
31. Cherry BJ, Ford EW, Peterson LT. Experiences with electronic health records: early adopters in long-term care facilities. *Health Care Manage Rev* 2011;36(3):265-274. [doi: [10.1097/HMR.0b013e31820e110f](https://doi.org/10.1097/HMR.0b013e31820e110f)] [Medline: [21646885](#)]
32. Bruns EJ, Hyde KL, Sather A, Hook AN, Lyon AR. Applying User Input to the Design and Testing of an Electronic Behavioral Health Information System for Wraparound Care Coordination. *Adm Policy Ment Health* 2016 May;43(3):350-368 [FREE Full text] [doi: [10.1007/s10488-015-0658-5](https://doi.org/10.1007/s10488-015-0658-5)] [Medline: [26060099](#)]
33. Matiz LA, Robbins-Milne L, Krause MC, Peretz PJ, Rausch JC. Evaluating the Impact of Information Technology Tools to Support the Asthma Medical Home. *Clin Pediatr (Phila)* 2016 Feb;55(2):165-170. [doi: [10.1177/0009922815596070](https://doi.org/10.1177/0009922815596070)] [Medline: [26195588](#)]
34. King J, Patel V, Jamoom E, DesRoches C. The role of health IT and delivery system reform in facilitating advanced care delivery. *Am J Manag Care* 2016 Apr;22(4):258-265 [FREE Full text] [Medline: [27143291](#)]

35. Hsiao C, King J, Hing E, Simon AE. The role of health information technology in care coordination in the United States. *Med Care* 2015 Feb;53(2):184-190. [doi: [10.1097/MLR.0000000000000276](https://doi.org/10.1097/MLR.0000000000000276)] [Medline: [25464164](#)]
36. Bruns EJ, Walker JS, Bernstein A, Daleiden E, Pullmann MD, Chorpita BF. Family voice with informed choice: coordinating wraparound with research-based treatment for children and adolescents. *J Clin Child Adolesc Psychol* 2014;43(2):256-269 [FREE Full text] [doi: [10.1080/15374416.2013.859081](https://doi.org/10.1080/15374416.2013.859081)] [Medline: [24325146](#)]
37. Bruns EJ, Hoagwood KE, Rivard JC, Wotring J, Marsenich L, Carter B. State implementation of evidence-based practice for youths, part II: recommendations for research and policy. *J Am Acad Child Adolesc Psychiatry* 2008 May;47(5):499-504. [doi: [10.1097/CHI.0b013e3181684557](https://doi.org/10.1097/CHI.0b013e3181684557)] [Medline: [18438187](#)]
38. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC: American Psychiatric Association; 2013.
39. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry* 2001 Nov;40(11):1337-1345. [doi: [10.1097/00004583-200111000-00015](https://doi.org/10.1097/00004583-200111000-00015)] [Medline: [11699809](#)]
40. Walker J, Bruns E, Penn M. Individualized services in systems of care: The wraparound process. In: Stroul BA, Blau GM, editors. *The system of care handbook: Transforming mental health services for children, youth, families*. Baltimore, MD: Paul H Brookes Publishing; 2008:127-153.
41. Walker JS, Bruns EJ. Building on practice-based evidence: using expert perspectives to define the wraparound process. *Psychiatr Serv* 2006 Nov;57(11):1579-1585. [doi: [10.1176/ps.2006.57.11.1579](https://doi.org/10.1176/ps.2006.57.11.1579)] [Medline: [17085605](#)]
42. Sauro J. *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. N/A: Measuring Usability LLC; 2011.
43. Brooke J. *System usability scale (SUS): a quick-and-dirty method of system evaluation user information*. Reading, UK: Digital Equipment Co Ltd; 1986.
44. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 2008 Jul 30;24(6):574-594. [doi: [10.1080/10447310802205776](https://doi.org/10.1080/10447310802205776)]
45. Peres SC, Pham T, Phillips R. Validation of the System Usability Scale (SUS). In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2013 Sep 30 Presented at: Proc Hum Factors Ergon Soc Annu Meet. ;57; 2013; Los Angeles, CA p. 192-196. [doi: [10.1177/1541931213571043](https://doi.org/10.1177/1541931213571043)]
46. Chafouleas S, Briesch A, Neugebauer S, Riley-Tillman T. *Usage Rating Profile-Intervention (Revised)*. Storrs, CT: University of Connecticut; 2011.
47. Reimers TM, Lee J. Parental acceptability of treatments for children's hypercholesterolemia. *J Behav Med* 1991 Jun;14(3):225-239. [Medline: [1875403](#)]
48. Martens BK, Witt JC, Elliott SN, Darveaux DX. Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice* 1985;16(2):191-198. [doi: [10.1037/0735-7028.16.2.191](https://doi.org/10.1037/0735-7028.16.2.191)]
49. Cook CR, Zhang Y. Merging measures to develop a more robust measure of perceptual predictors of implementation. 2014 Presented at: Annual Conference on Advancing School Mental Health; September 18-20, 2014; Pittsburgh, PA.
50. Accurso EC, Taylor RM, Garland AF. Evidence-based Practices Addressed in Community-based Children's Mental Health Clinical Supervision. *Train Educ Prof Psychol* 2011 May 01;5(2):88-96 [FREE Full text] [doi: [10.1037/a0023537](https://doi.org/10.1037/a0023537)] [Medline: [24761163](#)]
51. Dorsey S, Pullmann MD, Deblinger E, Berliner L, Kerns SE, Thompson K, et al. Improving practice in community-based settings: a randomized trial of supervision - study protocol. *Implement Sci* 2013 Aug 10;8:89 [FREE Full text] [doi: [10.1186/1748-5908-8-89](https://doi.org/10.1186/1748-5908-8-89)] [Medline: [23937766](#)]
52. Jensen-Doss A, Hawley KM. Understanding barriers to evidence-based assessment: clinician attitudes toward standardized assessment tools. *J Clin Child Adolesc Psychol* 2010;39(6):885-896 [FREE Full text] [doi: [10.1080/15374416.2010.517169](https://doi.org/10.1080/15374416.2010.517169)] [Medline: [21058134](#)]
53. Strating MMH, Nieboer AP. Psychometric test of the Team Climate Inventory-short version investigated in Dutch quality improvement teams. *BMC Health Serv Res* 2009 Jul 24;9:126 [FREE Full text] [doi: [10.1186/1472-6963-9-126](https://doi.org/10.1186/1472-6963-9-126)] [Medline: [19627621](#)]
54. Pullmann MD, Bruns EJ, Sather AK. Evaluating fidelity to the wraparound service model for youth: application of item response theory to the Wraparound Fidelity Index. *Psychol Assess* 2013 Jun;25(2):583-598 [FREE Full text] [doi: [10.1037/a0031864](https://doi.org/10.1037/a0031864)] [Medline: [23544392](#)]
55. Sather A, Hook A, Esposito I. The WFAS: A history of national means and other psychometrics. 2016 Presented at: Annual Research and Policy Conference for Child, Adolescent, and Youth Adult Behavioral Health; March 13-16, 2016; Tampa, FL.
56. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology* 1989;36(2):223. [doi: [10.1037/0022-0167.36.2.223](https://doi.org/10.1037/0022-0167.36.2.223)]
57. Hanson WE, Curry KT, Bandalos DL. Reliability Generalization of Working Alliance Inventory Scale Scores. *Educational and Psychological Measurement* 2016 Jul 02;62(4):659-673. [doi: [10.1177/0013164402062004008](https://doi.org/10.1177/0013164402062004008)]
58. Nguyen TD, Attkisson CC, Stegner BL. Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. *Eval Program Plann* 1983;6(3-4):299-313. [Medline: [10267258](#)]

59. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. *Eval Program Plann* 1982;5(3):233-237. [Medline: [10259963](#)]
60. Chorpita BF, Park A, Tsai K, Korathu-Larson P, Higa-McMillan CK, Nakamura BJ, Research Network on Youth Mental Health. Balancing effectiveness with responsiveness: Therapist satisfaction across different treatment designs in the Child STEPs randomized effectiveness trial. *J Consult Clin Psychol* 2015 Aug;83(4):709-718. [doi: [10.1037/a0039301](#)] [Medline: [25984802](#)]
61. Raudenbush S, Bryk A. Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences). Newbury Park, CA: Sage; 2002.
62. Chorpita BF, Reise S, Weisz JR, Grubbs K, Becker KD, Krull JL, Research Network on Youth Mental Health. Evaluation of the Brief Problem Checklist: child and caregiver interviews to measure clinical progress. *J Consult Clin Psychol* 2010 Aug;78(4):526-536. [doi: [10.1037/a0019602](#)] [Medline: [20658809](#)]
63. Lyon AR, Pullmann MD, Whitaker K, Ludwig K, Wasse JK, McCauley E. A Digital Feedback System to Support Implementation of Measurement-Based Care by School-Based Mental Health Clinicians. *J Clin Child Adolesc Psychol* 2017 Mar 02;1-12. [doi: [10.1080/15374416.2017.1280808](#)] [Medline: [28278597](#)]
64. Persons JB, Koerner K, Eidelman P, Thomas C, Liu H. Increasing psychotherapists' adoption and implementation of the evidence-based practice of progress monitoring. *Behav Res Ther* 2016 Jan;76:24-31 [FREE Full text] [doi: [10.1016/j.brat.2015.11.004](#)] [Medline: [26618237](#)]
65. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A Tipping Point for Measurement-Based Care. *Psychiatr Serv* 2017 Feb 01;68(2):179-188. [doi: [10.1176/appi.ps.201500439](#)] [Medline: [27582237](#)]
66. Cohen D. Assessing the Effect of an Electronic Decision Support System on Children's Mental Health Service Outcomes. *Journal of Technology in Human Services* 2015 Aug 12;33(3):225-240. [doi: [10.1080/15228835.2015.1039687](#)]
67. Ruud T. Routine outcome measures in Norway: Only partly implemented. *Int Rev Psychiatry* 2015;27(4):338-344. [Medline: [26256034](#)]
68. 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 May 10;19(5):e153 [FREE Full text] [doi: [10.2196/jmir.7725](#)] [Medline: [28490417](#)]
69. Lyon AR, Lewis CC, Melvin A, Boyd M, Nicodimos S, Liu FF, et al. Health Information Technologies-Academic and Commercial Evaluation (HIT-ACE) methodology: description and application to clinical feedback systems. *Implement Sci* 2016 Sep 22;11(1):128 [FREE Full text] [doi: [10.1186/s13012-016-0495-2](#)] [Medline: [27659426](#)]
70. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. *Med Care Res Rev* 2012 Apr;69(2):123-157 [FREE Full text] [doi: [10.1177/1077558711430690](#)] [Medline: [22203646](#)]
71. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci* 2015 Feb 12;10:21 [FREE Full text] [doi: [10.1186/s13012-015-0209-1](#)] [Medline: [25889199](#)]
72. McGeorge NM, Hegde S, Guarrera TK, Zhou Y, Lin L, Crane PW, et al. Studying the impact of interoperable electronic health records on workflow in ambulatory care. *International Journal of Industrial Ergonomics* 2015 Sep;49:144-155. [doi: [10.1016/j.ergon.2013.10.005](#)]
73. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006 May 16;144(10):742-752. [Medline: [16702590](#)]
74. Lyon AR, Stirman SW, Kerns SEU, Bruns EJ. Developing the mental health workforce: review and application of training approaches from multiple disciplines. *Adm Policy Ment Health* 2011 Jul;38(4):238-253 [FREE Full text] [doi: [10.1007/s10488-010-0331-y](#)] [Medline: [21190075](#)]
75. 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](#)]

Abbreviations

ASA: Attitudes toward Standardized Assessment Scale
BSPC: Brief Supervision Practice Checklist-Adapted
CAPER: Current Assessment Practice Evaluation-Revised
CSQ: Client Satisfaction Questionnaire
EHR: electronic health record
HIT: health information technology
HITECH Act: Health Information Technology for Economic and Clinical Health
ICC: intraclass correlation coefficient
SAAS: System Acceptability and Appropriateness Scale

SAU: services as usual

SPQ: Supervision Process Questionnaire

SUS: System Usability Scale

TCI: Team Climate Inventory, Short Version

WAI: Working Alliance Inventory

WFI-EZ: Wraparound Fidelity Index, Brief Version

Edited by G Eysenbach; submitted 16.03.18; peer-reviewed by G Strudwick, R Verheij; comments to author 09.04.18; revised version received 23.04.18; accepted 24.04.18; published 14.06.18.

Please cite as:

Bruns EJ, Hook AN, Parker EM, Esposito I, Sather A, Parigoris RM, Lyon AR, Hyde KL

Impact of a Web-Based Electronic Health Record on Behavioral Health Service Delivery for Children and Adolescents: Randomized Controlled Trial

J Med Internet Res 2018;20(6):e10197

URL: <http://www.jmir.org/2018/6/e10197/>

doi: [10.2196/10197](https://doi.org/10.2196/10197)

PMID: [29903701](https://pubmed.ncbi.nlm.nih.gov/29903701/)

©Eric J Bruns, Alyssa N Hook, Elizabeth M Parker, Isabella Esposito, April Sather, Ryan M Parigoris, Aaron R Lyon, Kelly L Hyde. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.06.2018. 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.

Original Paper

Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System

Seuli Bose-Brill¹, MD; Michelle Feeney², BS; Laura Prater³, MPH, MHA; Laura Miles¹, BSc; Angela Corbett¹, RN; Stephen Koesters¹, MD

¹Division of General Internal Medicine, Department of Internal Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States

²Medical Student Research Program, College of Medicine, The Ohio State University, Columbus, OH, United States

³College of Public Health, The Ohio State University, Columbus, OH, United States

Corresponding Author:

Seuli Bose-Brill, MD

Division of General Internal Medicine, Department of Internal Medicine

College of Medicine

The Ohio State University

895 Yard Street

Columbus, OH, 43212

United States

Phone: 1 614 293 7980

Email: seuli.brill@osumc.edu

Abstract

Background: Advance care planning allows patients to articulate their future care preferences should they no longer be able to make decisions on their own. Early advance care planning in outpatient settings provides benefits such as less aggressive care and fewer hospitalizations, yet it is underutilized due to barriers such as provider time constraints and communication complexity. Novel methods, such as patient portals, provide a unique opportunity to conduct advance care planning previsit planning for outpatient care. This follow-up to our pilot study aimed to conduct pragmatic testing of a novel electronic health record-tethered framework and its effects on advance care planning delivery in a real-world primary care setting.

Objective: Our intervention tested a previsit advance care planning workflow centered around a framework sent via secure electronic health record-linked patient portal in a real-world clinical setting. The primary objective of this study was to determine its impact on frequency and quality of advance care planning documentation.

Methods: We conducted a pragmatic trial including 2 sister clinical sites, one site implementing the intervention and the other continuing standard care. A total of 419 patients aged between 50 and 93 years with active portal accounts received intervention (n=200) or standard care (n=219). Chart review analyzed the presence of advance care planning and its quality and was graded with previously established scoring criteria based on advance care planning best practice guidelines from multiple nations.

Results: A total of 19.5% (39/200) of patients who received previsit planning responded to the framework. We found that the intervention site had statistically significant improvement in new advance care planning documentation rates ($P<.01$) and quality ($P<.01$) among all eligible patients. Advance care planning documentation rates increased by 105% (19/39 to 39/39) and quality improved among all patients who engaged in the previsit planning framework (n=39). Among eligible patients aged between 50 and 60 years at the intervention site, advance care planning documentation rates increased by 37% (27/96 to 37/96). Advance care planning documentation rates increased 34% among high users (27/67 to 36/67).

Conclusions: Advance care planning previsit planning using a secure electronic health record-supported patient portal framework yielded improvement in the presence of advance care planning documentation, with highest improvement in active patient portal users and patients aged between 50 and 60 years. Targeted previsit patient portal advance care planning delivery in these populations can potentially improve the quality of care in these populations.

(*J Med Internet Res* 2018;20(6):e208) doi:[10.2196/jmir.9203](https://doi.org/10.2196/jmir.9203)

KEYWORDS

advance care planning; electronic health records; patient portal

Introduction

Background

Advance care planning (ACP) is the formal process of outlining a patient's future care preferences should they lose the ability to make informed decisions for themselves [1,2]. ACP documentation provides guidance in accordance with patient care preferences to proxy decision makers and medical teams in times of medical crisis. Patients with documented ACP experience increased adherence to their desired medical preferences [3], higher rates of palliative management, fewer hospitalizations, and increased quality of life near death [1]. Similar studies have found that patients with a plan for future medical care spend less time in the hospital during their last year of life and have their wishes more frequently respected by family members [1,4-6]. ACP also has the potential to reduce expensive health care interventions not wanted by the patient, such as lengthy critical care stays at the end of life [7]. Despite these benefits, the Center for Disease Control and Prevention estimates that ACP completion rates are around 30%, even with advancements in the electronic health record (EHR) and ACP delivery [8].

The Institute of Medicine recommends conducting ACP early in a patient's chronic disease diagnosis, with periodic reassessment every several years or with change in prognosis (such as new diagnosis, hospitalization, or worsening of chronic disease) [9]. This ACP communication is best provided in the primary care setting [10]; yet, primary care remains ill-equipped to systematically conduct ACP discussions due to the competing care demands and fast pace of appointments [11]. As few as 1% of Medicare beneficiaries with an established primary care physician report having an ACP conversation with their health care provider [12]. Identified barriers include provider time constraints, uncertain patient prognosis, emotional complexity of ACP decisions, and difficulty in information sharing within and between health organizations [13]. The Centers for Medicare and Medicaid have recognized the need for early ACP by providing reimbursement for team-based primary care ACP discussions occurring under physician supervision [14]. These current gaps in care, juxtaposed with the urgent need for effective outpatient ACP care models, require development and rapid dissemination of innovative ACP strategies [15].

EHR-linked patient portals, first described in the 1990s and now ubiquitous due to EHR Meaningful Use guidelines [16], allow patients to electronically communicate with their medical providers within a secure platform. Patient portal communication has driven innovation in chronic disease management, population management, and previsit planning with strategies such as incorporating a patient portal refill button for hypertension patients; sharing top-priority problem list information with complex diabetes patients; and providing influenza vaccine outreach. These strategies allow providers to more effectively use time during an appointment by providing preparatory communication to patients before the appointment begins [17-21]. Patient portal-based outreach for health maintenance has been reported to marginally improve health maintenance behaviors such as flu vaccination (by 1-2%) [17].

Response rates to provider-initiated patient portal communication for chronic disease management have been reported at about 15% [22]. Patient portals are not without their own unique barriers; one study found that patients chose not to activate their portal due to a lack of sufficient instructions, privacy concerns, preference for face-to-face interaction, or connectivity obstacles [23-26]. Despite these limitations, patients and care teams are increasingly actively incorporating portals into medical management.

Framework and Objectives

In an earlier study, our research team developed and pilot-tested a concise EHR patient portal-linked, electronic ACP communication framework in a small randomized controlled trial [11,27]. The framework was developed through incorporating best practice guidelines, and then refined through focus group feedback and cognitive interviewing [11]. This developed framework consists of an introduction to ACP and key evidence-based questions that can be sent to patients for response outside of their office visit. By allowing patients to think and comment about their future wishes for care in advance of visits, we hoped to maximize patient-provider time in the office visit for advanced communication and documentation. The framework responses were automatically stored in a patient's medical record for retrieval by clinical staff or physicians at office visits. Patient responses to the framework could also be sent to the primary care provider for review [11]. Use of this framework in a small pilot study demonstrated improvement in ACP documentation rates and quality [11]. Even though our framework was the first to be piloted in the field [11], its feasibility and impact on outcomes in a real-world primary care setting when integrated into actual previsit planning algorithms in the course of actual clinical primary care were yet to be determined [28].

The aim of this study was to determine the impact of previsit ACP planning using a secure EHR-linked framework upon ACP documentation when incorporated into a real-world primary care environment.

Methods

This study was approved by the Ohio State University Institutional Review Board.

Sample

Patients 50 years or older, presenting for a preventive health or chronic disease follow-up visit, with an active MyChart account, at a participating clinical site were included in the study. Patients did not need prior MyChart experience. There were 2 clinical sites participating in the study. Sites were selected based on their demographic, size, and provider similarity, as well as their uniform clinical practices with respect to ACP delivery. Each clinical site used the same ACP practices before the study period, which included an institutional packet of information on ACP, state-issued documents about Advance Directives (ADs), and encouragement to discuss any ACP questions with their provider. The usual care site maintained these practices throughout the duration of the study. The intervention site incorporated an open-ended ACP framework (containing 4

questions), sent via a patient's EHR-tethered patient portal, into a clinical practice algorithm. Physicians, nurses, and other clinical staff at the intervention clinic were collectively involved in developing this ACP previsit planning algorithm that was rolled out practice wide over a 3 month period (Figure 1). The algorithm focused on promoting patient/provider communication surrounding ACP preferences, rather than intervening specifically on completed ADs scanned into the EHR. Furthermore, 219 patients in the control group and 200 patients in the intervention group were cared for during the study period, yielding 419 total participants. Further demographic characteristics are outlined in Table 1 for both the control and intervention sites. This study measured the impact of a clinical intervention rolled out practice wide. Our study team received a Waiver of Consent by the Institutional Review Board to assess the pragmatic impact of this clinical intervention. It was not possible to blind this study because chart review automatically revealed participants who received the ACP framework versus those who did not. We mitigated the inability to blind this study by using binomial metrics, such as documentation present/absent, and rigid scoring criteria for quality. We also conducted spot checking of reviews to ensure accuracy of dataset, as outlined above.

Participants did not know their intervention was the intervention of interest. Each clinical site agreed to participate in an ACP process study. However, providers and patient participants did not receive labels about whether they were receiving intervention or usual care. Practice workflow was implemented without labels.

Data security was paramount in this study. We used clinical staff (Institutional Review Board approved) routinely interacting with the patient record and completing previsit planning for clinical care to administer the intervention. Our research team was embedded in the clinical site. Data were housed within the secure institutional firewall and only accessed within the clinical site. Only de-identified datasets were shared with the statistical team for analysis using a secure, institutional drive. The delivery system developed by the practice providers had built-in safeguards for addressing clinical emergencies, such as patients responding to the secure message with medical complaints, by having a nurse and physician on call for urgent messages.

Measures

Charts were reviewed both 1 week and 4 weeks post appointment. Charts were reviewed by a member of the team who had received training and quality control checks on ACP chart review protocol. Chart review findings were spot checked by a second member of the team (one every 20 records) to ensure accuracy. The protocol outlined that training and education interventions would be used to respond to discrepancies in chart review rates and quality assessments. However, interventions were not needed because spot checking did not yield discrepancies. The participant's demographics, presence of ACP (including before and after the visit), quality of ACP if present, and number of MyChart messages sent in the last year were recorded. The intervention charts were reviewed to see if the patient had read the intervention on their portal and responded to any of the questions.

Figure 1. Intervention workflow that was implemented at the study practice during the 3-month trial period. ACP: advance care planning; PCP: primary care physician.

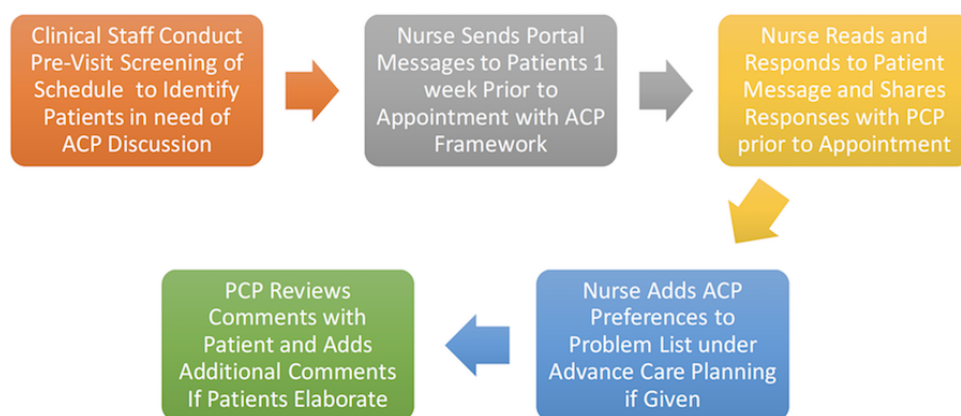


Table 1. Demographic information from both the control and intervention clinic.

Demographic information	Control (n=219)	Intervention (n=200)
Median age, years	61	61
Age range, years	50-93	50-91
Male, n (%)	76 (35)	101 (50.5)
Female, n (%)	143 (65)	99 (49.5)
Number of chronic diseases ^a	4	4
Number of medications ^a	7	6

^aMedian number per patient at each practice site.

Quality of ACP was measured using a 20-point scoring criteria entitled “Criteria for Scoring Quality of ACP Documentation” ([Multimedia Appendix 1](#)) that has been used in previous studies [11,27]. The rubric was created by our team after reviewing best practice reports on how to measure ACP quality, such as those from the United Kingdom National Health Service, the Australian Quality Advance Care Planning Board, and the National Hospice and Palliative Care Organization’s “Caring Connections” program [29-31]. This method of scoring the quality of ACP has not been validated, but assigned points for the presence of items in the patient’s ACP. The ability to quantify the quality of our patient’s ACP was a crucial component of the study and gave us more thorough feedback on the utility of the novel framework.

Analysis

Our study analyzed the documentation rates and quality of ACP across both study sites, especially focusing on new ACP documentation appearing in the EHR. To assess whether or not the increase in new documentation was significant between the 2 sites, a Fisher exact test was used. To analyze quality, a Mann-Whitney test was used to test the significance in new ACP quality between the 2 sites. The data were also analyzed by age and portal usage. Participants were separated into age

groups by decade and portal usage was defined as either high or low, with high usage being more than 10 portal messages in 1 year.

Results

Intervention

Of the 200 patients who were sent the intervention, 156 read the message (78.0% read rate) on their portal and 39 responded (19.5% response rate) to at least one question in the framework (see [Multimedia Appendix 2](#)). Of those who responded to our intervention, 49% (19/39) already had some form of ACP documented in their EHR and 51% (20/39) added ACP to their EHR for the first time, yielding a 105% (19/39 to 39/39) increase in ACP documentation rates. Responders with existing ACP had a mean quality score of 4.94, as compared with a mean score of 4.09 for all documented ACP at the intervention arm during our study period. For respondents without prior documented ACP, the intervention alone yielded a mean quality score of 3.7. Respondents sent a median 11 MyChart messages per year and had a median age of 63 years. MyChart usage did not increase due to the intervention at the site overall; patients had a median 5 messages per year at our intervention arm before and during the study period.

Table 2. Documented advance care planning (ACP) in electronic health record. Documentation rates represent the percentage of charts that had any form of ACP, and quality is rated by the 20-point scoring criteria.

Patient characteristic	Control	Intervention
All patients		
ACP^a documentation rate, N	219	200
Preintervention, n (%)	129 (58.9)	74 (37.0)
Postintervention, n (%)	130 (59.3)	94 (47.0)
Rate percentage increase, n (%)	0.7	27.0
Quality of all documented ACP, N	130	94
Quality rating postintervention, mean	3.26	4.09
Patients aged 50-60 years		
ACP documentation rate, N	109	96
Preintervention, n (%)	54 (49.5)	27 (28)
Postintervention, n (%)	55 (50.4)	37 (39)
Rate percentage increase, n (%)	1.8	37
Quality of all documented ACP, N	55	37
Quality rating postintervention, mean	2.81	3.75
Patients who are high portal users (>10 messages in 1 year)		
ACP documentation rate, N	82	67
Preintervention, n (%)	51 (62)	27 (40)
Postintervention, n (%)	51 (62)	36 (54)
Rate percentage increase, n (%)	0	33
Quality of ACP documentation, N	51	36
Quality rating postintervention, mean	3.25	4.19

^aACP: advance care planning.

Our intervention did not appear to affect the percentage of patients who had a scanned document in their EHR; both before and after the intervention, approximately 14% (28/200 and 7/47) of patients had a scanned directive at that practice. One patient brought in an Advance Directive to be scanned after responding to our framework.

Documentation and Quality Rates

ACP documentation in the EHR increased by 27.0% (74/200 to 94/200) during the study period at our intervention site, compared with a 0.7% (129/219 to 130/219) increase at our control site (Table 2; see Multimedia Appendix 1 for scoring criteria used for rating quality). A Fisher exact test was used to determine the significance of the differing increase in new documentation rates during the study period and yielded $P < .001$, indicating that patients exposed to our intervention were more likely to document ACP than those receiving usual care. A Mann-Whitney test was used to see if new ACP documented under our intervention was higher in quality and yielded $P < .001$, indicating that having the intervention led to a statistically significant ACP quality difference.

Age

Patients aged between 50 and 60 years saw the greatest increase in ACP completion rates. At our intervention site, documentation rose 37% (27/96 to 37/96) as compared with 1.8% (54/109 to 55/109) at the control site. Comparatively, the 61-70 age group saw a 31% (29/76 to 38/76) increase in documentation rates, and the 71-80 age group saw a 6% (17/26 to 18/26) increase at the intervention site. Our control site had a 0% increase in each of those 2 age groups, but higher baseline rates of ACP completion before the study period (64% and 81%). In the intervention arm, there was only 1 patient between 80 and 89 years, and 1 patient over 90 years, so there were insufficient data to analyze this group. Individuals in the 50-60 age group, however, had slightly lower ACP quality as compared with the study population as a whole.

MyChart Users

Those who sent more than 10 MyChart messages in 1 year were defined as "High Portal Users" and comprised approximately a third of the study group at each site. Documentation rose by over 33% (27/67 to 36/67) at the intervention site for this group. Comparatively, low portal users (10 messages or less) at our intervention site saw a 23.4% (47/133 to 58/133) increase in documentation rates.

Discussion

Patient Applications

In this study, we found that patients exposed to our framework were significantly more likely to have ACP documentation in the EHR and the quality of that documentation was better. This intervention benefits both the patient and the provider by providing another way for patients to think through the difficult decisions of how they envision their future care before their office visits. For patients who already used the patient portal, adding the framework would be a seamless integration into their usual part of their care. This tool was used most frequently by

patients in the 50-60 age group and already active on MyChart. Targeting patients who are high users or in this age demographic to receive this tool can be a strategy for providing high-yield individualized previsit planning for ACP using patient portals. This intervention did not capture many different demographics, including all nonportal users, so developing other strategies to improve ACP documentation against cultural, technological, and demographic barriers must be used to ensure that there are improved outcomes for all and to continue to address existing health disparities in ACP documentation [32]. Previous research has found that patients who are middle-aged, male, and have greater disease burden are more likely to use their patient portals, which is similar to the trend we found in our study [33]. Few studies report on response rates to practice level, provider-initiated patient portal interventions; however, response rates to our framework (20%) using this previsit planning system were higher than response rates to a similar intervention for primary care depression screening and management (15.4%) using a secure patient portal messaging system [22].

Office Workflow Applications

In terms of workflow, the framework requires a member of the care team to send out the MyChart message 3-5 days before the appointment. If a patient responded, answers were appropriately documented and sent to the patient's provider. In our study, the messages were sent out by a clinic nurse who could also answer any follow-up questions the patient had and then route the message to the appropriate provider. With increased team-based, patient-centered medical home patient outreach before appointments in primary care settings, as well as ubiquitous use of patient portals for practices to adhere to meaningful use guidelines, these interventions can be disseminated to a wide array of primary care practices. Higher rates of ACP documentation resulted, while reducing time needed to have a complete ACP discussion with the patient during the office visit, as existing answers have already been recorded and the patient had preparatory time to articulate their wishes.

Long-Term Clinical Applications

Previous studies have shown the benefits of previsit planning; if the provider has documentation of some of the patient's future care preferences beforehand, there can be a more productive discussion with the patient during their appointment [34]. As end-of-life discussions can be difficult for patients, having a standardized framework that is sent to all patients over 50 years helps normalize the discussion and better prepare the patient [34]. Physicians can also tailor their future ACP discussions based off the patient's documented answers from the framework, allowing them to be more effective conducting individualized ACP discussions [35]. For provider workflow, having the ACP process begin before the appointment can save time during the appointment and documentation afterwards, helping to ensure communication about end-of-life preferences occurs despite competing priorities [36].

The study was not designed to elicit qualitative feedback from patients and providers to promote its pragmatic implementation. However, the participating site liked the delivery system enough to implement it as a permanent intervention. Furthermore, the participating site shared the intervention, which has now been

disseminated to the wider net of associated primary care sites at the institution.

Limitations

Baseline rates of ACP documentation at each site were different, as noted in the results section, with the usual care site having higher rates of completed ACP documentation at baseline. However, preintervention chart review at both sites allowed assessment of typical documentation rates, to determine the change in documentation rates in intervention versus control before and after the intervention period. Additionally, patients

had to have an activated MyChart account to be included, which excluded a portion of the clinic population.

Conclusions

Incorporating the patient portal into ACP delivery is a promising way to increase completion rates and efficiently facilitate the conversation between the provider and the patient about their future wishes. This strategy may be more effective in patients familiar with patient portal use, who regularly use patient portal communication to access clinical care.

Acknowledgments

The project described was supported in part by the OSU College of Medicine Roessler Research Scholarship (MF).

Authors' Contributions

SB-B, MF, LM, LP, AC, and SK were involved in the concept and design; SB-B, MF, LM, and AC were involved in the acquisition of data; SB-B, MF, LM, and LP were involved in the analysis and interpretation of data; SB-B, MF, LM, and SK were involved in the drafting of the manuscript; SB-B, MF, LM, LP, AC, and SK were involved in the critical revision of the manuscript for important intellectual content; MF and LP were involved in the statistical analysis; and SB-B was involved in supervision.

Conflicts of Interest

SB-B was involved in developing the tested framework that is now available for nonproprietary use. She was not involved in administering the intervention or completing the primary chart review.

Multimedia Appendix 1

Advance Care Planning Quality Grading. Criteria for scoring quality of ACP documentation in the EHR.

[PDF File (Adobe PDF File), 25KB - [jmir_v20i6e208_app1.pdf](#)]

Multimedia Appendix 2

Framework Questionnaire.

[PDF File (Adobe PDF File), 28KB - [jmir_v20i6e208_app2.pdf](#)]

References

1. Mullick A, Martin J, Sallnow L. An introduction to advance care planning in practice. *BMJ* 2013 Oct 21;347:f6064. [Medline: [24144870](#)]
2. Brinkman-Stoppelenburg A, Rietjens JA, van der Heide A. The effects of advance care planning on end-of-life care: a systematic review. *Palliat Med* 2014 Sep;28(8):1000-1025. [doi: [10.1177/0269216314526272](#)] [Medline: [24651708](#)]
3. Hammes BJ, Rooney BL, Gundrum JD. A comparative, retrospective, observational study of the prevalence, availability, and specificity of advance care plans in a county that implemented an advance care planning microsystem. *J Am Geriatr Soc* 2010 Jul;58(7):1249-1255. [doi: [10.1111/j.1532-5415.2010.02956.x](#)] [Medline: [20649688](#)]
4. Wright AA, Zhang B, Ray A, Mack JW, Trice E, Balboni T, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. *JAMA* 2008 Oct 08;300(14):1665-1673 [FREE Full text] [doi: [10.1001/jama.300.14.1665](#)] [Medline: [18840840](#)]
5. O'Malley AJ, Caudry DJ, Grabowski DC. Predictors of nursing home residents' time to hospitalization. *Health Serv Res* 2011 Feb;46(1 Pt 1):82-104 [FREE Full text] [doi: [10.1111/j.1475-6773.2010.01170.x](#)] [Medline: [20849556](#)]
6. Molloy DW, Guyatt GH, Russo R, Goeree R, O'Brien BJ, Bédard M, et al. Systematic implementation of an advance directive program in nursing homes: a randomized controlled trial. *JAMA* 2000 Mar 15;283(11):1437-1444. [Medline: [10732933](#)]
7. Khandelwal N, Kross EK, Engelberg RA, Coe NB, Long AC, Curtis JR. Estimating the effect of palliative care interventions and advance care planning on ICU utilization: a systematic review. *Crit Care Med* 2015 May;43(5):1102-1111 [FREE Full text] [doi: [10.1097/CCM.0000000000000852](#)] [Medline: [25574794](#)]
8. Koeppen E, Burkard J, Ward E. Improving Advance Directive Completion Rates through Advance Care Planning. *Doctor of Nursing Practice Final Manuscripts* 2017:57 [FREE Full text]

9. Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life. Washington, D.C: The National Academies Press; 2015.
10. Ramsaroop SD, Reid MC, Adelman RD. Completing an advance directive in the primary care setting: what do we need for success? *J Am Geriatr Soc* 2007 Feb;55(2):277-283. [doi: [10.1111/j.1532-5415.2007.01065.x](https://doi.org/10.1111/j.1532-5415.2007.01065.x)] [Medline: [17302667](#)]
11. Bose-Brill S, Kretovics M, Ballenger T, Modan G, Lai A, Belanger L, et al. Testing of a tethered personal health record framework for early end-of-life discussions. *Am J Manag Care* 2016 Jul 01;22(7):e258-e263 [[FREE Full text](#)] [Medline: [27442309](#)]
12. Keary S, Moorman SM. Patient-Physician End-of-Life Discussions in the Routine Care of Medicare Beneficiaries. *J Aging Health* 2015 Sep;27(6):983-1002. [doi: [10.1177/0898264315569458](https://doi.org/10.1177/0898264315569458)] [Medline: [25649675](#)]
13. Bernacki RE, Block SD, American College of Physicians High Value Care Task Force. Communication about serious illness care goals: a review and synthesis of best practices. *JAMA Intern Med* 2014 Dec;174(12):1994-2003. [doi: [10.1001/jamainternmed.2014.5271](https://doi.org/10.1001/jamainternmed.2014.5271)] [Medline: [25330167](#)]
14. Centers for Medicare and Medicaid Services. 2015. Advance Care Planning URL: <https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/AdvanceCarePlanning.pdf> [accessed 2018-02-26] [[WebCite Cache ID 6xWmUluqO](#)]
15. Lund S, Richardson A, May C. Barriers to advance care planning at the end of life: an explanatory systematic review of implementation studies. *PLoS One* 2015 Feb 13;10(2):e0116629 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0116629](https://doi.org/10.1371/journal.pone.0116629)] [Medline: [25679395](#)]
16. Centers for Disease Control and Prevention. 2017. Meaningful Use URL: <https://www.cdc.gov/ehrmeaningfuluse/introduction.html> [accessed 2018-02-26] [[WebCite Cache ID 6xWmf7pQU](#)]
17. Cutrona SL, Sreedhara M, Goff SL, Fisher LD, Preusse P, Jackson M, et al. Improving Rates of Influenza Vaccination Through Electronic Health Record Portal Messages, Interactive Voice Recognition Calls and Patient-Enabled Electronic Health Record Updates: Protocol for a Randomized Controlled Trial. *JMIR Res Protoc* 2016 May 06;5(2):e56 [[FREE Full text](#)] [doi: [10.2196/resprot.5478](https://doi.org/10.2196/resprot.5478)] [Medline: [27153752](#)]
18. Grant RW, Wald JS, Poon EG, Schnipper JL, Gandhi TK, Volk LA, et al. Design and implementation of a web-based patient portal linked to an ambulatory care electronic health record: patient gateway for diabetes collaborative care. *Diabetes Technol Ther* 2006 Oct;8(5):576-586 [[FREE Full text](#)] [doi: [10.1089/dia.2006.8.576](https://doi.org/10.1089/dia.2006.8.576)] [Medline: [17037972](#)]
19. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. *J Med Internet Res* 2015 Jun 23;17(6):e148 [[FREE Full text](#)] [doi: [10.2196/jmir.4255](https://doi.org/10.2196/jmir.4255)] [Medline: [26104044](#)]
20. Grant RW, Uratsu CS, Estacio KR, Altschuler A, Kim E, Fireman B, et al. Pre-Visit Prioritization for complex patients with diabetes: Randomized trial design and implementation within an integrated health care system. *Contemp Clin Trials* 2016 Mar;47:196-201 [[FREE Full text](#)] [doi: [10.1016/j.cct.2016.01.012](https://doi.org/10.1016/j.cct.2016.01.012)] [Medline: [26820612](#)]
21. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. *Med Care* 2014 Mar;52(3):194-201 [[FREE Full text](#)] [doi: [10.1097/MLR.0000000000000069](https://doi.org/10.1097/MLR.0000000000000069)] [Medline: [24374412](#)]
22. Pecina J, North F, Williams MD, Angstman KB. Use of an on-line patient portal in a depression collaborative care management program. *J Affect Disord* 2017 Jan 15;208:1-5. [doi: [10.1016/j.jad.2016.08.034](https://doi.org/10.1016/j.jad.2016.08.034)] [Medline: [27736688](#)]
23. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. *J Am Med Inform Assoc* 2011 Dec;18(Suppl 1):i8-12 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000473](https://doi.org/10.1136/amiajnl-2011-000473)] [Medline: [22071530](#)]
24. Ronda MC, Dijkhorst-Oei LT, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. *J Med Internet Res* 2014 Nov 25;16(11):e263 [[FREE Full text](#)] [doi: [10.2196/jmir.3457](https://doi.org/10.2196/jmir.3457)] [Medline: [25424228](#)]
25. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. *J Med Internet Res* 2015 Dec 03;17(12):e275 [[FREE Full text](#)] [doi: [10.2196/jmir.4847](https://doi.org/10.2196/jmir.4847)] [Medline: [26681155](#)]
26. Uhler LM, Pérez Figueroa RE, Dickson M, McCullagh L, Kushniruk A, Monkman H, et al. InformedTogether: Usability Evaluation of a Web-Based Decision Aid to Facilitate Shared Advance Care Planning for Severe Chronic Obstructive Pulmonary Disease. *JMIR Hum Factors* 2015 Feb 25;2(1):e2 [[FREE Full text](#)] [doi: [10.2196/humanfactors.3842](https://doi.org/10.2196/humanfactors.3842)] [Medline: [27025896](#)]
27. Bose-Brill S, Kretovics M, Ballenger T, Modan G, Lai A, Belanger L, et al. Development of a tethered personal health record framework for early end-of-life discussions. *Am J Manag Care* 2016 Jun;22(6):412-418 [[FREE Full text](#)] [Medline: [27355808](#)]
28. Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials* 2012 Aug;9(4):436-446. [doi: [10.1177/1740774512450097](https://doi.org/10.1177/1740774512450097)] [Medline: [22752634](#)]
29. Australian Government Department of Health and Ageing. Canberra: Commonwealth of Australia; 2010. A National framework for advance care directives URL: <https://www.dementia.org.au/files/start2talk/5.0.4.1%20AHMAC%20framework.pdf> [accessed 2018-02-26] [[WebCite Cache ID 6xWmn6KIH](#)]

30. Henry C, Seymour J. NCCP. 2007. Advance care planning: a guide for health and social care staff URL: <http://www.nccp.org.uk/sites/default/files/AdvanceCarePlanning.pdf> [accessed 2018-02-26] [WebCite Cache ID 6xWn3zDmq]
31. CaringInfo, National Hospice and Palliative Care Organization. Advance Care Planning URL: <http://www.caringinfo.org/i4a/pages/index.cfm?pageid=3277> [accessed 2018-02-26] [WebCite Cache ID 6xWnGWLgP]
32. Li IF, Hsiung Y. Exploring Advance Care Planning in Taiwanese Indigenous Cancer Survivors: Proposal for a Pilot Case-Control Study. *JMIR Res Protoc* 2017 Dec 21;6(12):e238 [FREE Full text] [doi: [10.2196/resprot.5428](https://doi.org/10.2196/resprot.5428)] [Medline: [29269340](https://pubmed.ncbi.nlm.nih.gov/29269340/)]
33. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. *Ann Intern Med* 2013 Nov 19;159(10):677-687. [doi: [10.7326/0003-4819-159-10-201311190-00006](https://doi.org/10.7326/0003-4819-159-10-201311190-00006)] [Medline: [24247673](https://pubmed.ncbi.nlm.nih.gov/24247673/)]
34. Au DH, Udris EM, Engelberg RA, Diehr PH, Bryson CL, Reinke LF, et al. A randomized trial to improve communication about end-of-life care among patients with COPD. *Chest* 2012 Mar;141(3):726-735 [FREE Full text] [doi: [10.1378/chest.11-0362](https://doi.org/10.1378/chest.11-0362)] [Medline: [21940765](https://pubmed.ncbi.nlm.nih.gov/21940765/)]
35. Modan G, Brill SB. Engaging death: Narrative and constructed dialogue in Advance Care Planning discussions. *Commun Med* 2014;11(2):153-165. [Medline: [26596123](https://pubmed.ncbi.nlm.nih.gov/26596123/)]
36. Sinsky CA, Sinsky TA, Rajcevic E. Putting Pre-Visit Planning Into Practice. *Fam Pract Manag* 2015;22(6):34-38 [FREE Full text] [Medline: [26761083](https://pubmed.ncbi.nlm.nih.gov/26761083/)]

Abbreviations

ACP: advance care planning

EHR: electronic health record

Edited by G Eysenbach; submitted 17.10.17; peer-reviewed by K Powell, T Irizarry; comments to author 07.12.17; revised version received 27.02.18; accepted 14.03.18; published 26.06.18.

Please cite as:

Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S

Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System

J Med Internet Res 2018;20(6):e208

URL: <http://www.jmir.org/2018/6/e208/>

doi: [10.2196/jmir.9203](https://doi.org/10.2196/jmir.9203)

PMID: [29945860](https://pubmed.ncbi.nlm.nih.gov/29945860/)

©Seuli Bose-Brill, Michelle Feeney, Laura Prater, Laura Miles, Angela Corbett, Stephen Koesters. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.06.2018. 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.

Original Paper

Pokémon GO and Physical Activity in Asia: Multilevel Study

Ben D Ma^{1,2}, MSc; Sai Leung Ng¹, PhD; Tim Schwanen³, PhD; John Zacharias⁴, PhD; Mudi Zhou⁵, BSc; Ichiro Kawachi⁶, MBChB, PhD; Guibo Sun⁷, PhD

¹Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

²Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

³Transport Studies Unit, School of Geography and Environment, University of Oxford, Oxford, United Kingdom

⁴College of Architecture and Landscape, Peking University, Beijing, China

⁵Department of Mathematics, Imperial College London, London, United Kingdom

⁶Department of Social and Behavioral Sciences, TH Chan School of Public Health, Harvard University, Boston, MA, United States

⁷Department of Urban Planning and Design, Faculty of Architecture, The University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:

Sai Leung Ng, PhD

Department of Geography and Resource Management

The Chinese University of Hong Kong

2nd Floor, Wong Foo Yuan Building

Hong Kong, 999077

China (Hong Kong)

Phone: 852 39436527

Fax: 852 26035006

Email: sln@cuhk.edu.hk

Abstract

Background: Physical activity has long been considered as an important component of a healthy lifestyle. Although many efforts have been made to promote physical activity, there is no effective global intervention for physical activity promotion. Some researchers have suggested that Pokémon GO, a location-based augmented reality game, was associated with a short-term increase in players' physical activity on a global scale, but the details are far from clear.

Objective: The objective of our study was to study the relationship between Pokémon GO use and players' physical activity and how the relationship varies across players with different physical activity levels.

Methods: We conducted a field study in Hong Kong to investigate if Pokémon GO use was associated with physical activity. Pokémon GO players were asked to report their demographics through a survey; data on their Pokémon GO behaviors and daily walking and running distances were collected from their mobile phones. Participants (n=210) were Hong Kong residents, aged 13 to 65 years, who played Pokémon GO using iPhone 5 or 6 series in 5 selected types of built environment. We measured the participants' average daily walking and running distances over a period of 35 days, from 14 days before to 21 days after game installation. Multilevel modeling was used to identify and examine the predictors (including Pokémon GO behaviors, weather, demographics, and built environment) of the relationship between Pokémon GO use and daily walking and running distances.

Results: The average daily walking and running distances increased by 18.1% (0.96 km, approximately 1200 steps) in the 21 days after the participants installed Pokémon GO compared with the average distances over the 14 days before installation ($P<.001$). However, this association attenuated over time and was estimated to disappear 24 days after game installation. Multilevel models indicated that Pokémon GO had a stronger and more lasting association among the less physically active players compared with the physically active ones ($P<.001$). Playing Pokémon GO in green space had a significant positive relationship with daily walking and running distances ($P=.03$). Moreover, our results showed that whether Pokémon GO was played, the number of days played, weather (total rainfall, bright sunshine, mean air temperature, and mean wind speed), and demographics (age, gender, income, education, and body mass index) were associated with daily walking and running distances.

Conclusions: Pokémon GO was associated with a short-term increase in the players' daily walking and running distances; this association was especially strong among less physically active participants. Pokémon GO can build new links between humans and green space and encourage people to engage in physical activity. Our results show that location-based augmented reality games, such as Pokémon GO, have the potential to be a global public health intervention tool.

KEYWORDS

physical activity; Pokémon Go; public health intervention; exergame; weather

Introduction

As a Pokémon GO player, the first author of this paper walked 1526 km over 5 continents (Asia, Africa, Europe, North America, and South America) during the course of the game. This substantially increased his walking activity. His case may be that of an outlier, but increased walking is very common among Pokémon GO players. Pokémon GO is a location-based augmented reality (AR) game, which was launched in July 2016. Impressively, it receives 65 million monthly active players and has recorded more than 650 million app downloads within 6 months [1]. Pokémon are virtual creatures that inhabit the fictional Pokémon World. Pokémon GO allows players to locate, capture, and battle Pokémon on their mobile devices, as if the Pokémon were in the same real-world locations as the players. As players move around their real world, their avatars in the game move within the in-game map based on real world geographical locations. The in-game map contains many features such as “Gyms” and “Pokéstops,” where players can get Poké balls and other items to catch and battle with Pokémon. To play this game, players need to move around in the real world and not be sedentary. By May 2017, 9% of the global population installed Pokémon GO and traveled more than 15.8 billion km while playing Pokémon GO [2].

In this study, we sought to measure the relationship between Pokémon GO use and the players’ physical activity and how this relationship varied across players. Physical activity has long been considered an essential component of a healthy lifestyle [3-5]. Although many efforts to promote physical activity [5-7] have been undertaken, there is no effective global intervention for physical activity promotion [8]. Physical inactivity is the second-leading cause of preventable death worldwide, next to smoking [5].

The literature indicates that physical activity is affected by various variables, such as demographics [9-11], body mass index (BMI) [12,13], weather [14-16], and built environment [17-19]. Recently, increasing research suggested that new technologies, especially “gamified” physical activity interventions, are used to promote physical activity [20]; however, it was not until 2016 that Pokémon GO provided the first glimpse of how to intervene for the promotion of physical activity on a global scale.

Past studies found that Pokémon GO was associated with short-term increases in players’ physical activity levels. The existing literature can be mainly divided into 2 groups. Objectively-collected physical activity data indicated that there was an increase in the daily number of walking steps among the Pokémon GO players after game installation [21,22]. Studies using self-reported data indicated that players spent more time outside and performed more physical exercise because of Pokémon GO [23-26]. Other research included studies on Pokémon GO players’ motivations [27,28], experiences while

playing Pokémon GO [29], and potential adverse effects associated with playing Pokémon GO [30,31].

However, there remain 2 major research gaps. First, physical activity on each day is not only determined by players’ demographics and Pokémon GO behaviors (such as whether Pokémon GO was played, and the number of days played) but also by weather and built environment. Inclement or extreme weather can be physical activity barriers [15]. Further, built environment is the essential element for location-based AR games because the games are built based on the actual physical world environment. However, existing research failed to take them into account when investigating the relationship between Pokémon GO use and physical activity.

Second, although some researchers suggested that the association between Pokémon GO and physical activity varies among different populations [24,25,32,33], to date, little research has been performed on these potential differences among players. Howe et al focused on young adults [21]; Althoff et al studied wearable Microsoft product users [22]; Kogan et al focused their study on dog owners [34]; and Wong et al researched university students [26]. Asia is home to 60% of the world’s population [35]; although Pokémon GO has been suspended in Mainland China, there are still considerable numbers of Pokémon GO players in other parts of Asia. However, most studies were conducted in Western countries, and there are no objective studies of Pokémon GO and physical activity in Asia.

Therefore, we used a multilevel modeling approach to answer 2 main research questions. They are as follows:

- What is the relationship between Pokémon GO use and players’ physical activity levels?
- How does the association between Pokémon GO and physical activity vary across players with different physical activity levels?

Methods

Research Design

We conducted a field survey in August 2016 during the first month after Pokémon GO was launched in Hong Kong. Researchers approached Pokémon GO players at 5 study sites that represented 5 typical types of built environment. Participants’ data, including sex, age, income, education level, BMI (kg/m²), and the start date of Pokémon GO, were collected using a questionnaire. Participants’ physical activity data, specifically daily walking and running distances, were captured from their iPhone “health” app pages by taking pictures of their screens. Eventually, we compiled a data set covering the period from 14 days before and 21 days after the installation of Pokémon GO. Weather data were collected from Hong Kong Observatory, including total rainfall (mm), bright sunshine

(hours), mean air temperature (°C), and mean wind speed (km/h).

Table 1. Descriptions of study sites.

Study site	Type of built environment	Description
Central	Office area	Land used for administration, or clerical, technical, professional, or other like business activity
Wan Chai	Mixed use area	Land used for mixed uses
Victoria Park	Green space	Integrated park consisting of playgrounds, sitting-out areas or public/mini sports grounds
Wong Tai Sin	Residential area	Land used to for residential accommodations
Causeway Bay	Retail premises	Land used to: (a) sell goods by retail or by retail and wholesale; (b) sell services; or (c) hire goods

The survey was conducted by our researchers on a face to face basis with each participant in the study areas. All participants provided written informed consent to participate prior to survey conduction. Participants received a Pokémon toy as a token of appreciation for completing the study. All data collected were anonymized, and the participants' names and home addresses were not collected.

Study Population

We approached Pokémon GO players while they were playing Pokémon GO within the study sites. To ensure the reliability and consistency of data, only iPhone 5s or 6 series users with qualified daily walking and running distances in their iPhone "health" app were included in this study. We excluded individuals who did not complete the questionnaire, were not iPhone 5s or 6 users, and those who were iPhone users but were unable to provide their daily walking and running distances data. We also excluded respondents with insufficient Pokémon GO levels to unlock the primary functions of the game. The primary function of the game was to reach level 5 and gain access to the Gyms.

Study Sites

We used 5 study sites, which represented 5 typical types of built environment in Hong Kong, namely, a green space, an office area, a residential area, a mixed-use area, and a retail location (Table 1). For each built environment, we selected a 200-m radial zone based on the surrounding land use [36] and set ≥60% of the surrounding land use belonging to that type of built environment as the criterion for the site selection through a GIS platform [37].

Daily Walking and Running Distances

Daily walking and running distance data represented the daily physical activities of Pokémon GO players. To compare players' physical activities before and after installation, we used the average walking and running distances each day over a period of 35 days (14 days before to 21 days after game installation). We estimated 95% CI 5.8-6.0 through a bootstrap with 500 resamples of daily walking and running distances. If there were no data on a given day, which meant the player probably did not touch his or her phone at all, or there were problems with the phone, we ignored the data recorded on that day. We excluded 15 observations from the analysis because the data values were empty.

Multilevel Modeling

Multilevel modeling was employed to investigate the difference of association between Pokémon GO and daily walking and running distances across players. We have multiple observations of daily walking and running distances of the same individual (from 14 days before to 21 days after the game installation); based on this cluster data structure, multilevel modeling was applied to investigate the relationship between Pokémon GO use and physical activity. At observation level, each observation has its own observation-level attributes, including whether Pokémon GO was played, the number of days played, total rainfall (mm), bright sunshine (hours), mean air temperature (degree Celsius), and mean wind speed (km/h). At player level, age, income, gender, education, BMI (kg/m²), and built environment variables were added. Further, interdependencies among different player levels were taken into account through multilevel modeling. We transferred parts of variables into dummy variables based on questions that we intended to investigate. MLwiN V.3.0 was used to conduct the multilevel modeling analysis.

Funding, Ethical Approval, and Data Sharing

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. This study was approved by Survey and Behavioral Research Ethics Committee at The Chinese University of Hong Kong on August 4, 2016. The statistical code and dataset are available from the corresponding author.

Results

Study Populations and Sites

Overall, 210 Pokémon GO players were included in this study; Table 2 shows the profile of the participants. We identified 2964 individuals as Pokémon GO players while they were playing Pokémon GO within the study sites. Of these, 1248 answered our questionnaire and 1028 players were excluded because they were not iPhone 5s or 6 users or were iPhone users but unable to provide their daily walking and running distance data. Additionally, we excluded those who did not complete the questionnaire (n=7) or who were unable to unlock the primary functions of the game (n=3).

Table 2. Characteristics of Pokémon GO players (N=210).

Characteristic	Value
Age (years), mean (SD)	26.1 (8.7)
Age group (years), n (%)	
13-17	25 (11.9)
18-23	67 (31.9)
24-29	43 (20.5)
30-35	35 (16.7)
36-50	21 (10.0)
>51	2 (1.0)
Female, n (%)	71 (33.8)
Education, n (%)	
High school or lower	70 (33.3)
College or higher	118 (56.1)
Monthly income (HKD), n (%)	
<5000	65 (31.0)
5000-10000	12 (5.7)
10001-15000	41 (19.5)
15001-20000	30 (14.3)
20001-30000	18 (8.6)
>30000	18 (8.6)
Site, n (%)	
Office (Central)	33 (15.7)
Mixed use (Wan Chai)	48 (22.9)
Green space (Victoria Park)	46 (21.9)
Residential (Wong Tai Sin)	55 (26.2)
Retail premises (Causeway Bay)	28 (13.3)
Body mass index (BMI; kg/m²), n (%)	
Underweight (BMI<18.5)	57 (27.1)
Normal (18.5≤BMI<25)	121 (57.6)
Overweight/obese (25≤BMI)	16 (7.6)
Daily walking and running distances in 2 weeks before installation of Pokémon GO (km), mean (SD)	5.4 (3.5)

Daily Walking and Running Distances Before and After the Game Installation

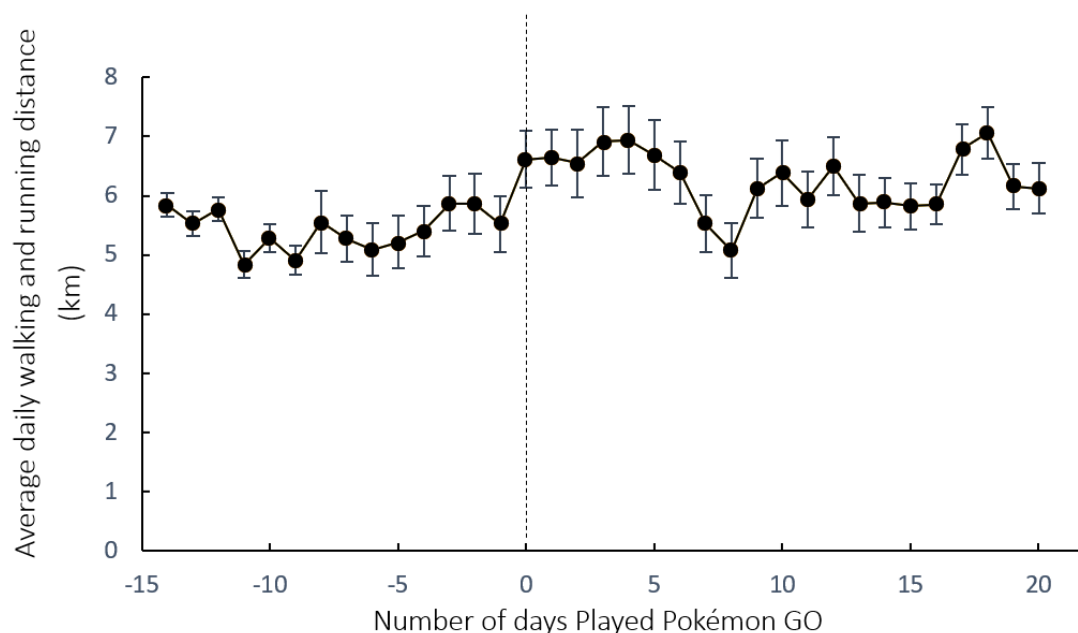
Figure 1 shows the change in average daily walking and running distances of the Pokémon GO players in the period from 14 days before to 21 days after game installation. The average distance increased by 18.1% (0.96 km), from 5.30 km (SD 2.12) (before installation) to 6.26 km (SD 2.45) (after installation). A comparison of means shows that Pokémon GO was associated with increases in daily walking and running distances ($F=33.825$, $P<.001$). We observed a decrease in daily walking

and running distances during the period from the 5th to 8th day after installation. During that period, Hong Kong's weather deteriorated because of the influence of the typhoon Nida.

Pokémon GO and Physical Activity across Players

The results of multilevel models are shown in Table 3. A null model, a model without any input variances, which simply describes the variance at each level, was conducted first. We found that 22.5% of the variance was explained at the observation level, and this provided justification for proceeding to the subsequent analysis.

Figure 1. Average daily walking and running distance before and after the installation of Pokémon GO. Error bars correspond to bootstrapped 95% confidence intervals.



Model 1 contains observation-level variables. In model 2, the player-level variables were added and analyzed together with the observation-level variables (Table 3). In model 3, assuming that all other variables in the model were held constant, the relation between Pokémon GO behaviors and dependent variables were investigated. In model 4, on the basis of model 3, the interactions of Pokémon GO use (or no Pokémon GO use) and built environment were included (Table 4).

The results of multilevel models showed that the observation-level variables (including whether Pokémon GO was played, the number of days played, total rainfall, bright sunshine duration, mean air temperature, and mean wind speed) were significantly related to daily walking and running distances. Whether Pokémon GO was played ($P=.005$) and bright sunshine duration ($P=.01$) were significantly associated with the increased daily walking and running distances. Conversely, the number of days played ($P<.001$), total rainfall ($P=.003$), mean air temperature ($P=.01$), and mean wind speed ($P=.04$) were negatively correlated with daily walking and running distances. At the player level, our results indicated that players who were male, had high school or lower education, and exhibited normal BMI (kg/m^2) were likely to have higher daily walking and running distances (Table 4). Furthermore, playing Pokémon

GO in green space was associated with higher daily walking and running distances ($P=.03$).

Playing Pokémon GO was significantly associated with increased daily walking and running distances ($P=.005$), whereas the number of days played was significantly associated with decreased daily walking and running distances ($P<.001$). The results showed that the association between Pokémon GO and daily walking and running distances attenuated over time and may have disappeared after players played Pokémon GO for 24 days (Table 4).

Most importantly, our results indicate that the effects of playing Pokémon GO are somewhat stronger for players with lower levels of daily physical activity and slightly stronger for those who have played the game for fewer days. These findings are obtained when the estimated effects of playing Pokémon GO and number of days played are allowed to vary across individuals in the sample. The negative estimate of the covariance of playing Pokémon GO and intercept (-0.008 in Model 4) captures the stronger health benefits of playing among respondents with shorter walking and running distances. The estimated covariance between days played and whether Pokémon GO was played of -0.00057 shows that those playing the game for a shorter period of time have, all else equal, experienced slightly less health benefits.

Table 3. Multilevel models of Pokémon GO and daily walking and running distances (km) models 1 and 2.

Variable	Model 1		Model 2	
	β	<i>P</i> value	β	<i>P</i> value
Observation level				
Pokémon GO or not ^a (reference: not)	0.085	<.001	0.084	<.001
Days played ^b	−0.003	<.001	−0.004	<.001
Weather				
Total rainfall	−0.001	.013	−0.001	.012
Bright sunshine duration	0.003	.013	0.003	.013
Mean air temperature	−0.012	.020	−0.013	.016
Mean wind speed	−0.001	.010	−0.001	.009
Player level				
Age (reference >35)				
13-17	—	—	−0.087	.019
18-23	—	—	−0.016	.320
24-29	—	—	−0.016	.316
30-35	—	—	−0.065	.028
Income (HKD; reference >30000)				
<5000	—	—	−0.011	.385
5000-10000	—	—	−0.111	.012
10001-15000	—	—	−0.028	.223
15001-20000	—	—	0.050	.096
20001-30000	—	—	0.030	.223
Gender (reference: female)				
Male			0.039	.026
Education (reference: high school or lower)				
College or higher	—	—	−0.077	.036
BMI^c (reference: normal; kg/m²)				
Underweight (BMI<18.5)	—	—	−0.069	.004
Overweight and obese (25≤BMI)	—	—	−0.050	.045
Built environment (reference: office)				
Mixed use area	—	—	0.026	.207
Green space	—	—	0.041	.094
Residential	—	—	0.001	.491
Retail premises	—	—	−0.003	.469
Intercept				
Mean	0.714	<.001	0.735	<.001
Variance (observation level)	0.058	<.001	0.058	<.001
Variance (player level)	0.017	<.001	0.014	<.001

^aIf the data belonged to the date before the participant played Pokémon GO, this value would be 0. Otherwise, it would be 1.

^bIf the data belonged to the date before the participant played Pokémon GO, this value would be 0. Otherwise, it would be the number of days he or she played.

^cBMI: body mass index.

Table 4. Multilevel models of Pokémon GO and daily walking and running distances (km) models 3 and 4.

Variable	Model 3		Model 4	
	β	<i>P</i> value	β	<i>P</i> value
Observation level				
Pokémon GO or not^a (reference: not)				
Mean	0.093	<.001	0.072	.005
Variance (player level)	0.018	<.001	0.017	<.001
Covariance with intercept (player level)	−0.009	<.001	−0.008	<.001
Days played^b				
Mean	−0.004	<.001	−0.005	<.001
Variance (player level)	0.000096	<.001	0.000097	<.001
Covariance with intercept (player level)	0.000068	.346	0.000059	.346
Covariance with Pokémon GO (player level)	−0.00058	.004	−0.00057	.004
Weather				
Total rainfall	−0.001	.003	−0.001	.003
Bright sunshine duration	0.003	.011	0.003	.012
Mean air temperature	−0.013	.010	−0.013	.011
Mean wind speed	−0.001	.037	−0.001	.036
Players level				
Age (reference >35)				
13-17	−0.068	.048	−0.067	.047
18-23	−0.007	.414	−0.007	.415
24-29	−0.010	.380	−0.009	.389
30-35	−0.061	.031	−0.061	.031
Income (HKD; reference >30000)				
<5000	−0.024	.263	−0.024	.265
5000-10000	−0.122	.005	−0.122	.005
10001-15000	−0.041	.123	−0.040	.125
15001-20000	0.044	.112	0.046	.103
20001-30000	0.027	.239	0.028	.227
Gender (reference: female)				
Male	0.040	.019	0.041	.017
Education (reference: high school or lower)				
College or higher	−0.075	.037	−0.073	.040
BMI^c (reference: normal; kg/m²)				
Underweight (BMI<18.5)	−0.074	.002	−0.073	.002
Overweight and obese (25≤BMI)	−0.052	.036	−0.051	.038
Built environment (reference: office)				
Mixed use area	0.032	.195	0.028	.191
Green space	0.044	.068	0.001	.494
Residential	−0.001	.490	−0.018	.309
Retail premises	0.003	.460	0.001	.491
Interactions^d (reference: Pokémon GO x Office)				

Variable	Model 3		Model 4	
	β	<i>P</i> value	β	<i>P</i> value
Pokémon GO * Mixed use	–	–	0.006	.437
Pokémon GO * Green space	–	–	0.068	.028
Pokémon GO * Residential	–	–	0.027	.210
Pokémon GO * Retail premises	–	–	–0.010	.406
Intercept				
Mean	0.733	<.001	0.733	<.001
Variance (observation level)	0.053	<.001	0.053	<.001
Variance (player level)	0.016	<.001	0.016	<.001

^aIf the data belonged to the date before the participant played Pokémon GO, this value would be 0. Otherwise, it would be 1.

^bIf the data belonged to the date before the participant played Pokémon GO, this value would be 0. Otherwise, it would be the number of days he or she played.

^cBMI: body mass index.

^dThe value is equal to the value of Pokémon GO or not multiplied by built environment.

Discussion

Principal Results and Comparison With Prior Work

This study reported that after the game installation, average daily walking and running distances of Hong Kong Pokémon GO players increased by 18.1% (0.96 km, approximately 1200 steps) compared with the period before installation, which spanned for over a period of 35 days. These results were largely consistent with those of previous studies that were conducted in USA. For example, Howe et al found that the average number of daily steps for Pokémon GO players during the first week of installation increased by 955 steps [21]. Althoff and his colleagues indicated that Pokémon GO led to significant increases of more than 25% (1473 steps per day over a period of 30 days) [22]. These studies conducted in USA and the study conducted in Asia study showed that playing Pokémon GO was associated with an increase in physical activity in the first month players played the game. These results supported the idea that Pokémon GO may improve public health by promoting physical activity [32,33,38,39].

Our results showed that the increases in daily walking and running distances attenuated over time. This finding is also congruent with those of the previous studies conducted in USA [21,22]. However, the speed of attenuation was different between the study conducted by us, in Asia, and those conducted in USA. Although Howe et al [21] reported that the association was no longer observed after 6 weeks (42 days), our results indicated that the association disappeared faster, in 24 days, with Hong Kong players. This may be because Hong Kong is a relatively fast-paced society [40], which may make it easier for people to lose interest in the game. The high-density built environment, which lacks spacious places for playing location-based AR games, could also be another reason for this difference. If we wish to maximize the positive benefit from the game, initiatives should be taken to find ways to slow down attenuation.

From multilevel models, we found that total rainfall, bright sunshine duration, mean air temperature, and mean wind speed

were significantly related to Pokémon GO players' physical activity levels. The link between the weather and physical activity is well established. An increased duration of daily bright sunshine could increase daily walking activity [41] and participation in outdoor activities [42]; higher rainfall and wind speed are likely to be associated with lower physical activity levels [43–46]; the temperature could have different impacts on physical activities across areas and seasons [47]. For the season during which we conducted our study in Hong Kong, we found that the duration of bright sunshine was significantly associated with increased daily walking and running distances, whereas total rainfall, mean air temperature, and mean wind speed were negatively correlated with the distances. To the best of our knowledge, this is the first study that controlled the weather variables to provide a more precise relationship between Pokémon GO behaviors and physical activity levels. On December 7, 2017, the new version of Pokémon GO was introduced and included weather features, in addition to an in-game weather visual map, weather near Pokémon players will impact Pokémon in a variety of ways. This update made weather an even more important factor for the players [48].

The relationship between physical activity and built environment has been well investigated in recent decades [17]. As a location-based AR game, Pokémon GO is highly connected with built environment and both of these variables could be associated with players' physical activity levels. In previous studies, however, built environment was not taken into consideration or they showed little impact on physical activity levels [21,22]. We examined Pokémon GO together with built environment and found that green space had a significant positive relationship with daily walking and running distances (Table 4). This result indicated that Pokémon GO may have encouraged players to use the green space around them to engage in physical activity. Nature-based recreation has decreased by 25% in the last 40 years [49], endangering the health benefits associated with nature and green space [50]. To solve this problem, Pokémon GO shows great potential to build connections between green space and opportunities for physical activity.

By controlling for the weather variable at the observation level and for demographics and built environment at the player level, our results showed that players with shorter daily walking and running distances exhibited stronger associations with Pokémon GO. These associations also lasted longer than players with relatively longer distances before installation. This result confirmed the previous findings from the studies conducted in USA that Pokémon GO is able to reach low activity populations [22,26]. Governments and public health agencies may consider the possibility of using location-based AR technology to improve physical activity and public health.

Contributions

To the best of our knowledge, this study was one of the earliest and the only studies conducted during the Pokémon GO craze (from July 2016 to August 2016). This work offers 3 main contributions. First, we combined field surveys with automatically recorded physical activity data from mobile devices. The field study gave us an idea of who was playing the game outside and engaging in physical activity. Second, this paper was the first paper that investigated the relationship between Pokémon GO and physical activity with the weather and built environment added as covariates. Third, the difference in associations between Pokémon GO use and physical activity among different players was investigated. Such information is important to enhance the positive impact brought by future games or interventions. Further, as the first objective physical activity study of Pokémon GO in Asia, this study provided evidence together with other previous studies, which indicated that Pokémon GO was associated with short-term increases in physical activity levels.

Limitations

Our study had some limitations. First, our study population was a representative of active Pokémon GO players; it was not a random sample of the Hong Kong population. It could not represent all the Pokémon GO players in Hong Kong. It is

possible that some players may have only played at midnight (or during other irregular times), which was not within our survey time, or some people may have only played Pokémon GO in a place that was dissimilar to our study areas.

Second, this was a one group pretest-posttest design study. We acknowledge that we lack a control group; however, our data indicate that Pokémon GO alone was associated with physical activity changes in participants. During the study time, there was no campaign, parade, or other large-scale physical activity-related event that happened in Hong Kong. There were also no dramatic weather changes that could have led to increases in physical activity levels.

Further, we used daily walking and running distances as a proxy measurement of physical activity. We do acknowledge the possibility that other physical activities (eg, swimming and basketball) were likely to be not recorded if the players did not carry their phones during those activities.

Conclusions

We studied 210 Pokémon GO players in Hong Kong during the Pokémon GO craze of 2016. Our results indicated that after the installation of the game, the average daily walking and running distances increased by 18.1% (0.96 km, approximately 1200 steps) compared with the levels measured before the installation. However, this association attenuated over time and was estimated to disappear after 24 days. The results of multilevel models indicated that the weather should be considered in this kind of research and the association between Pokémon GO use and physical activity was stronger among less physically active people compared with physically active people. We found that a game like Pokémon GO has the potential to build new links between humans and green space and to encourage people to engage in physical activity outdoor. Having a better understanding of the relationship between Pokémon GO use and physical activity may cast light on the future efforts needed to promote public health on a global scale.

Acknowledgments

The authors thank Hui Lin, PhD, of The Chinese University of Hong Kong for advice on the field study.

Authors' Contributions

BDM, GS, and SLN designed the project. BDM, SLN, and GS collected data. BDM, TS, and MZ analyzed the data. BDM and SLN wrote the manuscript. All authors contributed to the manuscript. TS, GS, IK, and JZ provided methodological guidance. SLN provided financial support for the data collection.

Conflicts of Interest

All authors declare no financial relationships with any organizations that might have an interest in the submitted work and no other relationships or activities that could appear to have influenced the submitted work.

References

1. Smith C. DMR. 80 Amazing pokemon go statistics URL: <https://expandedramblings.com/index.php/pokemon-go-statistics/> [accessed 2017-11-30] [WebCite Cache ID 6vMh1O2dB]
2. The Pokémon GO team. 2017. Adventures Await! Pokémon GO URL: <https://pokemongolive.com/en/post/adventureweek2017> [accessed 2017-11-30] [WebCite Cache ID 6vMh7JdtX]
3. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985 Apr;100(2):126-131 [FREE Full text] [Medline: 3920711]

4. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA* 1995 Feb 1;273(5):402-407. [Medline: [7823386](#)]
5. World Health Organization. 2010. Global recommendations on physical activity for health URL: http://www.who.int/dietphysicalactivity/factsheet_recommendations/en/ [accessed 2017-11-30] [WebCite Cache ID 6vMhN2meJ]
6. Sallis JF, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. *Am J Prev Med* 1998 Nov;15(4):379-397. [Medline: [9838979](#)]
7. Marcus BH, Nigg CR, Riebe D, Forsyth LH. Interactive communication strategies: implications for population-based physical-activity promotion. *Am J Prev Med* 2000 Aug;19(2):121-126. [Medline: [10913903](#)]
8. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet* 2012 Jul 21;380(9838):247-257. [doi: [10.1016/S0140-6736\(12\)60646-1](#)] [Medline: [22818937](#)]
9. Powell LM, Slater S, Chaloupka FJ, Harper D. Availability of physical activity-related facilities and neighborhood demographic and socioeconomic characteristics: a national study. *Am J Public Health* 2006 Sep;96(9):1676-1680. [doi: [10.2105/AJPH.2005.065573](#)] [Medline: [16873753](#)]
10. Azevedo MR, Araújo CLP, Reichert FF, Siqueira FV, da SMC, Hallal PC. Gender differences in leisure-time physical activity. *Int J Public Health* 2007;52(1):8-15 [FREE Full text] [Medline: [17966815](#)]
11. Borodulin K, Sipilä N, Rahkonen O, Leino-Arjas P, Kestilä L, Jousilahti P, et al. Socio-demographic and behavioral variation in barriers to leisure-time physical activity. *Scand J Public Health* 2016 Feb;44(1):62-69. [doi: [10.1177/1403494815604080](#)] [Medline: [26392420](#)]
12. Frank LD, Andresen MA, Schmid TL. Obesity relationships with community design, physical activity, and time spent in cars. *Am J Prev Med* 2004 Aug;27(2):87-96. [doi: [10.1016/j.amepre.2004.04.011](#)] [Medline: [15261894](#)]
13. Blair SN, Church TS. The fitness, obesity, and health equation: is physical activity the common denominator? *JAMA* 2004 Sep 08;292(10):1232-1234. [doi: [10.1001/jama.292.10.1232](#)] [Medline: [15353537](#)]
14. Lewis LK, Maher C, Belanger K, Tremblay M, Chaput J, Olds T. At the Mercy of the Gods: Associations Between Weather, Physical Activity, and Sedentary Time in Children. *Pediatr Exerc Sci* 2016 Feb;28(1):152-163. [doi: [10.1123/pes.2015-0076](#)] [Medline: [26098393](#)]
15. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. *Public Health* 2007 Dec;121(12):909-922. [doi: [10.1016/j.puhe.2007.04.009](#)] [Medline: [17920646](#)]
16. Wu Y, Luben R, Wareham N, Griffin S, Jones AP. Weather, day length and physical activity in older adults: Cross-sectional results from the European Prospective Investigation into Cancer and Nutrition (EPIC) Norfolk Cohort. *PLoS One* 2017;12(5):e0177767 [FREE Full text] [doi: [10.1371/journal.pone.0177767](#)] [Medline: [28562613](#)]
17. Handy SL, Boarnet MG, Ewing R, Killingsworth RE. How the built environment affects physical activity: views from urban planning. *Am J Prev Med* 2002 Aug;23(2 Suppl):64-73. [Medline: [12133739](#)]
18. Gebel K, Ding D, Foster C, Bauman AE, Sallis JF. Improving current practice in reviews of the built environment and physical activity. *Sports Med* 2015 Mar;45(3):297-302. [doi: [10.1007/s40279-014-0273-8](#)] [Medline: [25315457](#)]
19. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am J Prev Med* 2009 Apr;36(4 Suppl):S99-123.e12 [FREE Full text] [doi: [10.1016/j.amepre.2009.01.005](#)] [Medline: [19285216](#)]
20. Baranowski T, Blumberg F, Buday R, DeSmet A, Fiellin LE, Green CS, et al. Games for Health for Children-Current Status and Needed Research. *Games Health J* 2016 Feb;5(1):1-12 [FREE Full text] [doi: [10.1089/g4h.2015.0026](#)] [Medline: [26262772](#)]
21. Howe KB, Suharlim C, Ueda P, Howe D, Kawachi I, Rimm EB. Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study. *BMJ* 2016 Dec 13;355:i6270 [FREE Full text] [Medline: [27965211](#)]
22. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on Physical Activity: Study and Implications. *J Med Internet Res* 2016 Dec 06;18(12):e315 [FREE Full text] [doi: [10.2196/jmir.6759](#)] [Medline: [27923778](#)]
23. Mosendz P. Bloomberg. Counting the extra footsteps from Pokémon Go mania URL: <https://www.bloomberg.com/news/articles/2016-07-15/counting-the-extra-footsteps-from-pokemon-go-mania> [accessed 2017-11-30] [WebCite Cache ID 6vMhPIPUM]
24. Nigg CR, Mateo DJ, An J. Pokémon GO May Increase Physical Activity and Decrease Sedentary Behaviors. *Am J Public Health* 2017 Jan;107(1):37-38. [doi: [10.2105/AJPH.2016.303532](#)] [Medline: [27854536](#)]
25. Wagner-Greene V, Wotring A, Castor T, Kruger J, Mortemore S, Dake J. Pokémon GO: Healthy or Harmful? *Am J Public Health* 2017 Jan;107(1):35-36. [doi: [10.2105/AJPH.2016.303548](#)] [Medline: [27925813](#)]
26. Wong FY. Influence of Pokémon Go on physical activity levels of university players: a cross-sectional study. *Int J Health Geogr* 2017 Feb 22;16(1):8 [FREE Full text] [doi: [10.1186/s12942-017-0080-1](#)] [Medline: [28228102](#)]
27. Rasche P, Schlomann A, Mertens A. Who Is Still Playing Pokémon Go? A Web-Based Survey. *JMIR Serious Games* 2017 Apr 05;5(2):e7 [FREE Full text] [doi: [10.2196/games.7197](#)] [Medline: [28381393](#)]

28. Marquet O, Alberico C, Adlakha D, Hipp JA. Examining Motivations to Play Pokémon GO and Their Influence on Perceived Outcomes and Physical Activity. *JMIR Serious Games* 2017 Oct 24;5(4):e21 [FREE Full text] [doi: [10.2196/games.8048](https://doi.org/10.2196/games.8048)] [Medline: [29066423](https://pubmed.ncbi.nlm.nih.gov/29066423/)]
29. Lindqvist A, Castelli D, Hallberg J, Rutberg S. The Praise and Price of Pokémon GO: A Qualitative Study of Children's and Parents' Experiences. *JMIR Serious Games* 2018 Jan 03;6(1):e1 [FREE Full text] [doi: [10.2196/games.8979](https://doi.org/10.2196/games.8979)] [Medline: [29298750](https://pubmed.ncbi.nlm.nih.gov/29298750/)]
30. Barbieri S, Vettore G, Pietrantonio V, Snenghi R, Tredese A, Bergamini M, et al. Pedestrian Inattention Blindness While Playing Pokémon Go as an Emerging Health-Risk Behavior: A Case Report. *J Med Internet Res* 2017 Apr 01;19(4):e86 [FREE Full text] [doi: [10.2196/jmir.6596](https://doi.org/10.2196/jmir.6596)] [Medline: [28365563](https://pubmed.ncbi.nlm.nih.gov/28365563/)]
31. Ono S, Ono Y, Michihata N, Sasabuchi Y, Yasunaga H. Effect of Pokémon GO on incidence of fatal traffic injuries: a population-based quasi-experimental study using the national traffic collisions database in Japan. *Inj Prev* 2017 Oct 12:042503. [doi: [10.1136/injuryprev-2017-042503](https://doi.org/10.1136/injuryprev-2017-042503)] [Medline: [29025874](https://pubmed.ncbi.nlm.nih.gov/29025874/)]
32. LeBlanc AG, Chaput J. Pokémon Go: A game changer for the physical inactivity crisis? *Prev Med* 2016 Nov 14:235-237. [doi: [10.1016/j.ypmed.2016.11.012](https://doi.org/10.1016/j.ypmed.2016.11.012)] [Medline: [27856340](https://pubmed.ncbi.nlm.nih.gov/27856340/)]
33. McCartney M. Margaret McCartney: Game on for Pokémon Go. *BMJ* 2016 Aug 09;354:i4306 [FREE Full text] [doi: [10.1136/bmj.i4306](https://doi.org/10.1136/bmj.i4306)]
34. Kogan L, Hellyer P, Duncan C, Schoenfeld-Tacher R. A pilot investigation of the physical and psychological benefits of playing Pokémon GO for dog owners. *Computers in Human Behavior* 2017 Nov;76:431-437 [FREE Full text] [doi: [10.1016/j.chb.2017.07.043](https://doi.org/10.1016/j.chb.2017.07.043)]
35. United Nations. 2011. World Population Prospects The 2010 Revision URL: http://www.un.org/en/development/desa/population/publications/pdf/trends/WPP2010/WPP2010_Volume-I_Comprehensive-Tables.pdf [WebCite Cache ID [6y5W4w39P](https://www.webcitation.org/6y5W4w39P)]
36. Giridharan R, Ganesan S, Lau S. Daytime urban heat island effect in high-rise and high-density residential developments in Hong Kong. *Energy and Buildings* 2004 Jun;36(6):525-534 [FREE Full text] [doi: [10.1016/j.enbuild.2003.12.016](https://doi.org/10.1016/j.enbuild.2003.12.016)]
37. Sun G, Zacharias J, Ma B, Oreskovic N. How do metro stations integrate with walking environments? Results from walking access within three types of built environment in Beijing. *Cities* 2016 Jul;56:91-98 [FREE Full text] [doi: [10.1016/j.cities.2016.03.001](https://doi.org/10.1016/j.cities.2016.03.001)]
38. Tateno M, Skokauskas N, Kato TA, Teo AR, Guerrero APS. New game software (Pokémon Go) may help youth with severe social withdrawal, hikikomori. *Psychiatry Res* 2016 Dec 30;246:848-849. [doi: [10.1016/j.psychres.2016.10.038](https://doi.org/10.1016/j.psychres.2016.10.038)] [Medline: [27817905](https://pubmed.ncbi.nlm.nih.gov/27817905/)]
39. Wise J. Pokémon Go's health benefits seem short lived. *BMJ* 2016 Dec 13:i6684 [FREE Full text] [doi: [10.1136/bmj.i6684](https://doi.org/10.1136/bmj.i6684)]
40. Levine R, Norenzayan A. The Pace of Life in 31 Countries. *Journal of Cross-Cultural Psychology* 2016 Jul 26;30(2):178-205 [FREE Full text] [doi: [10.1177/0022022199030002003](https://doi.org/10.1177/0022022199030002003)]
41. Klenk J, Büchele G, Rapp K, Franke S, Peter R, ActiFE Study Group. Walking on sunshine: effect of weather conditions on physical activity in older people. *J Epidemiol Community Health* 2012 May;66(5):474-476. [doi: [10.1136/jech.2010.128090](https://doi.org/10.1136/jech.2010.128090)] [Medline: [21325149](https://pubmed.ncbi.nlm.nih.gov/21325149/)]
42. Lindsey G, Han Y, Wilson J, Yang J. Neighborhood Correlates of Urban Trail Use. *J Phys Act Health* 2006 Feb;3(s1):S139-S157. [doi: [10.1123/jpah.3.s1.s139](https://doi.org/10.1123/jpah.3.s1.s139)] [Medline: [28834516](https://pubmed.ncbi.nlm.nih.gov/28834516/)]
43. Brodersen NH, Steptoe A, Williamson S, Wardle J. Sociodemographic, developmental, environmental, and psychological correlates of physical activity and sedentary behavior at age 11 to 12. *Ann Behav Med* 2005 Feb;29(1):2-11. [doi: [10.1207/s15324796abm2901_2](https://doi.org/10.1207/s15324796abm2901_2)] [Medline: [15677295](https://pubmed.ncbi.nlm.nih.gov/15677295/)]
44. Harrison F, Jones AP, Bentham G, van SEMF, Cassidy A, Griffin SJ. The impact of rainfall and school break time policies on physical activity in 9-10 year old British children: a repeated measures study. *Int J Behav Nutr Phys Act* 2011 May 24;8:47 [FREE Full text] [doi: [10.1186/1479-5868-8-47](https://doi.org/10.1186/1479-5868-8-47)] [Medline: [21609441](https://pubmed.ncbi.nlm.nih.gov/21609441/)]
45. Currie JL, Develin E. Stroll your way to well-being: a survey of the perceived benefits, barriers, community support, and stigma associated with pram walking groups designed for new mothers, Sydney, Australia. *Health Care Women Int* 2002 Dec;23(8):882-893. [doi: [10.1080/07399330290112380](https://doi.org/10.1080/07399330290112380)] [Medline: [12487703](https://pubmed.ncbi.nlm.nih.gov/12487703/)]
46. Tu W, Stump TE, Damush TM, Clark DO. The effects of health and environment on exercise-class participation in older, urban women. *J Aging Phys Act* 2004 Oct;12(4):480-496. [Medline: [15851821](https://pubmed.ncbi.nlm.nih.gov/15851821/)]
47. Chan CB, Ryan DAJ, Tudor-Locke C. Relationship between objective measures of physical activity and weather: a longitudinal study. *Int J Behav Nutr Phys Act* 2006 Aug 07;3:21 [FREE Full text] [doi: [10.1186/1479-5868-3-21](https://doi.org/10.1186/1479-5868-3-21)] [Medline: [16893452](https://pubmed.ncbi.nlm.nih.gov/16893452/)]
48. Pokemongolive.com. Developer Insights: Introducing Dynamic Weather Gameplay and Battle Parties URL: <https://pokemongolive.com/en/post/decdevupdate-weather/> [accessed 2018-03-22] [WebCite Cache ID [6y7KhYkH2](https://www.webcitation.org/6y7KhYkH2)]
49. Pergams ORW, Zaradic PA. Evidence for a fundamental and pervasive shift away from nature-based recreation. *Proc Natl Acad Sci U S A* 2008 Feb 19;105(7):2295-2300 [FREE Full text] [doi: [10.1073/pnas.0709893105](https://doi.org/10.1073/pnas.0709893105)] [Medline: [18250312](https://pubmed.ncbi.nlm.nih.gov/18250312/)]
50. Lee ACK, Maheswaran R. The health benefits of urban green spaces: a review of the evidence. *J Public Health (Oxf)* 2011 Jun;33(2):212-222. [doi: [10.1093/pubmed/fdq068](https://doi.org/10.1093/pubmed/fdq068)] [Medline: [20833671](https://pubmed.ncbi.nlm.nih.gov/20833671/)]

Abbreviations**BMI:** body mass index**AR:** augmented reality

Edited by G Eysenbach; submitted 16.12.17; peer-reviewed by A Henriksen, R Kretschmann, T Jiang, P Wark; comments to author 19.03.18; revised version received 22.04.18; accepted 08.05.18; published 15.06.18.

Please cite as:

Ma BD, Ng SL, Schwanen T, Zacharias J, Zhou M, Kawachi I, Sun G

Pokémon GO and Physical Activity in Asia: Multilevel Study

J Med Internet Res 2018;20(6):e217

URL: <http://www.jmir.org/2018/6/e217/>

doi: [10.2196/jmir.9670](https://doi.org/10.2196/jmir.9670)

PMID: [29907559](https://pubmed.ncbi.nlm.nih.gov/29907559/)

©Ben D Ma, Sai Leung Ng, Tim Schwanen, John Zacharias, Mudi Zhou, Ichiro Kawachi, Guibo Sun. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 15.06.2018. 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.

Original Paper

Sharing and Empathy in Digital Spaces: Qualitative Study of Online Health Forums for Breast Cancer and Motor Neuron Disease (Amyotrophic Lateral Sclerosis)

Sarah Hargreaves¹, BA (Hons), MSc, MA, PhD; Peter A Bath¹, BSc (Technical), MSc, PhD; Suzanne Duffin¹, BMedSci (Hons), PG Cert, MSc; Julie Ellis², BA (Hons), MA, PhD

¹Health Informatics Research Group, Information School, University of Sheffield, Sheffield, United Kingdom

²Department 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, S1 4DP

United Kingdom

Phone: 44 114 222 2636

Email: p.a.bath@sheffield.ac.uk

Abstract

Background: The availability of an increasing number of online health forums has altered the experience of living with a health condition, as more people are now able to connect and support one another. Empathy is an important component of peer-to-peer support, although little is known about how empathy develops and operates within online health forums.

Objective: The aim of this paper is to explore how empathy develops and operates within two online health forums for differing health conditions: breast cancer and motor neuron disease (MND), also known as amyotrophic lateral sclerosis.

Methods: This qualitative study analyzed data from two sources: interviews with forum users and downloaded forum posts. Data were collected from two online health forums provided by UK charities: Breast Cancer Care and the Motor Neurone Disease Association. We analyzed 84 threads from the breast cancer forum and 52 from the MND forum. Threads were purposively sampled to reflect varied experiences (eg, illness stages, topics of conversation, and user characteristics). Semistructured interviews were conducted with 14 Breast Cancer Care forum users and five users of the MND forum. All datasets were analyzed thematically using Braun and Clarke's six-phase approach and combined to triangulate the analysis.

Results: We found that empathy develops and operates through shared experiences and connections. The development of empathy begins outside the forum with experiences of illness onset and diagnosis, creating emotional and informational needs. Users came to the forum and found their experiences and needs were shared and understood by others, setting the empathetic tone and supportive ethos of the forum. The forum was viewed as both a useful and meaningful space in which they could share experiences, information, and emotions, and receive empathetic support within a supportive and warm atmosphere. Empathy operated through connections formed within this humane space based on similarity, relationships, and shared feelings. Users felt a need to connect to users who they felt were like themselves (eg, people sharing the same specific diagnosis). They formed relationships with other users. They connected based on the emotional understanding of ill health. Within these connections, empathic communication flourished.

Conclusions: Empathy develops and operates within shared experiences and connections, enabled by structural possibilities provided by the forums giving users the opportunity and means to interact within public, restricted, and more private spaces, as well as within groups and in one-to-one exchanges. The atmosphere and feeling of both sites and perceived audiences were important facilitators of empathy, with users sharing a perception of virtual communities of caring and supportive people. Our findings are of value to organizations hosting health forums and to health professionals signposting patients to additional sources of support.

KEYWORDS

online health forum; sharing; breast cancer; motor neuron disease; amyotrophic lateral sclerosis; empathy

Introduction

Background

The availability of an increasing number of online health forums has altered the experience of living with a health condition because more people are now able to connect and support one another [1]. This relatively accessible means of communication is particularly important when someone is living with a health condition that is rare or which inhibits communication (eg, motor neuron disease [MND]) [2]. Online support communities provide various types of support, such as emotional [3-5] and informational support [2]. One aspect of human emotional intelligence, which may, in part, shape online communication within forums, is empathy. Previous research has highlighted the role of empathy in encouraging helpful behaviors, such as motivating users to share knowledge [6], and the potential benefits users derive from empathetic communications [7]. However, less is known about the processes which facilitate social and relational connections, such as empathy, within online forums [1]. The study reported in this paper addressed this by exploring how empathy develops within these spaces, is enacted, and operates within two online health forums—one for individuals living with or affected by breast cancer and another for people living with or affected by MND. Although most research on forums and online communities tends to focus on a single health condition [1,8], we sought to enrich our understanding of empathic processes by exploring the lived experience of individuals engaging with two different online health forums.

Definitions of Empathy

There are many definitions of empathy and these have been applied inconsistently within research [9]. However, there are aspects of empathy that are commonly discussed, such as *cognitive empathy* (the capacity to understand another person's feelings) and *affective empathy* (the capacity to respond with appropriate emotion). A less commonly discussed aspect is *congruence* in empathy, which is the requirement for empathetic emotion of the observer to correspond with that of the observed. For the purposes of this paper, we define empathy as “(1) knowing what the other person is feeling, (2) feeling what the other person is feeling, and (3) responding compassionately to another person's distress” (p 234 [10]). This definition was adopted for the purposes of this paper because it resonated with participants' understandings of empathy, which had emerged during the interviews. Participants with both health conditions understood empathy chiefly as the ability to understand what another is feeling and to stand in another person's shoes. Many noted the ease with which they were able to achieve this with forum members because of their shared illness experiences. This enabled them to both (1) know and (2) feel. Participants linked thought with action, requiring empathy to be demonstrated with (3) a compassionate response. Within previous studies on online health forums, focusing on a variety

of topics, there are elements within the findings that correspond with this definition:

1. Knowing what another person is feeling, such as users with shared very specific understandings of living with particular conditions (eg, [1,11,12]).
2. Feeling what another person is feeling, such as users recognizing emotions and feeling a resonance (eg, [13]).
3. Responding compassionately to another person's distress, such as written responses to problems expressed on the forum, demonstrating compassion by offering mutual understanding [14], sharing personal experience [14], validating feelings [8,15], or responding promptly to requests for support [14,16].

This paper, therefore, makes a new contribution to understanding of empathy in eHealth in that it explores the importance of these three aspects of empathy for people with life-threatening conditions using online health forums and their communication online. In the following section, we outline previous research that has examined the role of empathy in online communication and in online health forums.

Empathic Communication Online

For online health forums to flourish, users must be prepared to share knowledge and experiences with others in the community [6] and to develop trust in how the forum operates and with others who use it [17]. Empathy is thought to encourage and motivate sharing [6], and thus has been studied by researchers seeking both to understand the mechanisms that facilitate a forum's usefulness and to consider how software might facilitate empathetic communication [18]. To date, research on empathy has focused on assessing whether empathy is present within messages posted on online communities [19] and on how empathy is expressed and structured within interactions among users [19,20]. Studies have examined differing levels of support provided by users to group members in need, with empathic exchanges thought to offer a greater depth of support [19,21]. The emotional depth conveyed through empathic interactions emanates from an understanding gained from others having shared their experiences of being in a similar situation and the ability to imagine and work oneself into the emotions and situation of others [21]. Empathetic support requires effort, understanding, and caring [19].

Studies have identified the factors that may increase the likelihood of empathetic exchanges or strengthen empathetic bonds. Empathy is stronger or more evident when users share similarities, such as going through the same experience. Similarity provides a basis for greater identification between users, increases the likelihood of an accurate understanding of the situation (empathetic accuracy), and fosters greater intimacy [12,14,19,22]. Gender also influences these processes, with more empathy present in communities with a female membership [23]. Communications among women tend to have greater empathic content and lower factual content, and

communications between men typically have lower empathic content and higher factual content [18]. Age is also a factor in influencing emotional and empathetic communication online, with one study showing that teenagers send more emotional messages in a personal style, in contrast to the formal style of older adults [24]. However, older adults within this study expressed higher levels of concern for others than their teenage counterparts. Crucially, empathy is influenced greatly by the topic of interest, and is found to be most present within support communities [25,26]. Messages conveying emotions and feelings are the most common trigger for an empathic response [19,20]. Empathy is also encouraged by the altruistic sharing behaviors of users, such as sharing personal knowledge, feelings, and experiences [6], and moving the focus from self to offering peer support [13]. Empathy is commonly expressed in responses to requests for support [27]. Self-disclosure is an important element of online communication [19], which informs empathic interactions among users.

Empathy and Online Health Forums

Studies focusing on online health forums have explored the context in which online interactions take place as users work through illness experiences, interact with health services, and experience the impacts of ill health on offline relationships and everyday lives [1,8]. This context influences empathy building within health forums as users live with the ongoing stresses and challenges of ill health. For people living with health conditions that may or will shorten their life, the need for empathetic support and understanding can be particularly important [28,29]. Shared understanding and experiences of living with ill health found within condition-specific health forums have been shown to help foster a sense of community and build friendships [1]. These factors encourage empathetic behaviors because users seek to nurture and support one another [30]. However, this relatively limited research on empathy in online health forums has not considered the processes by which empathy is developed in online health forums and this study sought to fill this gap.

Study Aims

The aim of our study was to develop a better understanding of the role of empathy among people who share information, experiences, and emotions in online health forums. More specifically, we were interested in exploring the processes that underpin the development of empathy and influence sharing among users of two online health forums for two very different conditions. In this paper, we explore how empathy is developed through shared experiences and via connections formed between users. We used forum data to explore how empathy is narrated within the forums and the interviews to explore, and gain a deeper understanding of, the human processes behind it.

Methods

Study Design

This study is part of a wider project exploring how and why people living in a range of extreme circumstances share information, emotions, and experience in online spaces and communities. The key focus was on understanding the role of trust and empathy in encouraging and shaping sharing behaviors.

The study recently reported on the processes by which trust develops on a breast cancer forum [17], demonstrating how it operates within structural, relational, and temporal dimensions. This paper explores the development of empathy through sharing in online environments.

The paper focuses on the experiences of people living with two very different health conditions: breast cancer and MND, also known as amyotrophic lateral sclerosis. Breast cancer is both a life-threatening and a long-term condition given that many people survive for 10 years after diagnosis [31], whereas others may die of it or associated conditions. It is not a single condition; there are different stages and types of cancer, with differing prognoses, treatment options, and choices. People with breast cancer face differing trajectories and, for some, the condition can be terminal. In contrast, MND is a relatively rare and incurable condition, and people diagnosed with one of the diseases within this group can expect to have their life severely shortened [32]. MND causes damage to nerves and muscles, and as the disease progresses people may experience difficulties with mobility, speech, eating, drinking, and breathing. The lack of a curative treatment means that people face an inevitable decline. These two health conditions were chosen to explore aspects of living in extreme circumstances with two quite different conditions: one that is potentially life threatening (breast cancer) and the other which is life shortening and for which there is no cure (MND).

We explored how empathy develops within two online health forums that provide support for people living with breast cancer and with MND. We utilized a case-study approach to gain an in-depth understanding of the phenomenon, undertaking qualitative analysis of a sample of threads from the online forum provided by Breast Cancer Care (a UK-based charitable organization) and a forum hosted by the Motor Neurone Disease Association (a charity providing support in England, Wales, and Northern Ireland). We also conducted semistructured interviews with 20 users of these forums. The combined datasets broadened our understanding of the issues raised because each source illuminated particular aspects of the forum experience and additional insights were gained from analyzing connections between the threads and interview material. Although analysis of combined data sources is not commonly undertaken [8], in this instance, it provided the opportunity to triangulate analyses and reach a better understanding of how empathy develops in the forums [17]. The forum posts demonstrated how empathy is presented within postings and the interviews provided insights on participants' thoughts and feelings about empathy within the forums, and how this informed their interactions and sharing practices within these online spaces.

Study Setting

We approached and gained permission from two leading UK charities to access and analyze forum posts and to invite forum users to participate as interviewees in the study: Breast Cancer Care and the Motor Neurone Disease Association. These particular forums were chosen for a number of reasons. Both forums met the inclusion criteria, they are open access (in that anyone may view and read the messages without having a user ID and password), and the terms and conditions stated that it is

permissible to use the forum data for research purposes. The two forums were also chosen to explore differing research settings. The two forums differ in the following ways:

1. Size of membership: Breast Cancer Care is a relatively large forum—at the time of the study, it was estimated there were 200,000 registered users (Breast Cancer Care, personal communication). MND is a smaller forum, with more than 3000 registered users, although approximately 50 to 100 of these are active users (MND, personal communication).
2. Length of time established: Breast Cancer Care is one of the longest established Web-based forums; it was established in 1999-2000. The MND forum is more recent; it was established in 2013-2014.
3. The health conditions are distinctive and different (as previously described).
4. Gender differences: the Breast Cancer Care forum is predominantly female and the active members of the MND forum are predominantly male.

Both sites are moderated by staff employed by the charities. Both forums were structured to reflect both stages within user journeys (eg, diagnosis), user characteristics (eg, carers), and aspects of living with the condition. The Breast Cancer Care forum provides a wider range of boards in which to share experiences, reflecting journeys across differing trajectories.

Forum Threads

The researcher (JE) spent time reading through the posts to familiarize herself with both sites and gain an understanding of the context and feel of the online environment and characteristics

of the differing boards [17]. We undertook a two-stage process to download a sample of forum threads from the two forums. In the first stage, we purposively selected discussion boards from the two forums to provide suitable diversity of the topics discussed. In 2015, 233 threads were downloaded from 10 boards on the Breast Cancer Care forum and 135 were downloaded from five boards on the MND forum. The threads spanned a range of years (2006-2014), reflecting the different dates in which the boards originated. The posts were collected from the start date of the board until the sample quota was reached. The quota was decided pragmatically, the calculation based on the volume of threads within each site, and a requirement to ensure that the amount of data downloaded was manageable. The data collection strategy was designed to ensure that the forum threads sampled were collected in a consistent way across the differing boards. From these, a subset of 84 threads from the Breast Cancer Care forum and 52 from the MND forum were purposively selected to reflect varied experiences (eg, illness stages, topics of conversation, and user characteristics; Table 1). Sampling was undertaken with an awareness of the study purpose to explore concepts of empathy, and a wider focus on sharing and trust. The broad sampling strategy sought to ensure the study gained an understanding of both how empathy develops and operates; therefore, we included threads in which empathy was very apparent (eg, emotional threads, moments of need), threads in which empathy could be perceived as lacking (eg, moments of conflict), and threads in which empathy was not apparent or immediately apparent (eg, everyday conversations).

Table 1. Details of forum posts included in the analysis.

Forum type and section	Board	Date of posts	Threads analyzed, n
Breast Cancer Care forum			
Talk to people like me	Men's board	2006	10
Talk to people like me	Younger women and families	2007	10
Welcome to the forum	New members board	2012-2013	12
Going through treatment	Surgery	2007	11
Going through treatment	Chemotherapy (monthly)	2013	1 ^a
I am recently diagnosed	Diagnosed with breast cancer	2007	10
Have I got breast cancer?	Appointments and waiting	2007-2008	10
I have secondary breast cancer	End-of-life board	2009-2010	10
Living with and beyond breast cancer	Coping with fear and anxiety	2012-2014	5
Living with and beyond breast cancer	Sex and relationships	2012-2014	5
Motor neuron disease forum			
Help and advice	Tips and experiences	2011-2013	10
Miscellaneous discussion	Off topic	2011 and 2013	10
General discussion	Life with MND	2010-2013	11
General discussion	Introduce yourself	2010-2013	11
General discussion	For carers	2010-2013	10

^aThe chemotherapy board differed from other boards because it was in one long thread. Stages within one thread were sampled (eg, start, one-quarter way through, halfway through, and so on).

Interviews

The interview participants were recruited by posting a message on the Breast Cancer Care and Motor Neurone Disease Association forums. The message explained the aim of the study and people who were interested in participating were invited to contact the study team. Criteria for interview were as follows: participants were aged 18 years and older, users of the forums with a diagnosis of breast cancer or MND, or a relative or friend of someone with the condition. All interview volunteers recruited were people living with the health conditions, with the exception of the partners of two participants with MND who sat in on and contributed to the interviews. Interviews were conducted by JE either face-to-face, via phone, Skype, or email. The final sample, therefore, was self-selecting (ie, those people who contacted the study team to take part). Although there is potential for a biased sample, this method yielded a range of experiences (Table 2). Some interview participants with MND communicated via a speech synthesizer in face-to-face sessions and one participated via email. Interviews were conducted at a time and in a place convenient for the interviewees, typically at home. Table 2 shows the characteristics of the interviewees. There were 14 breast cancer interviewees and five MND interviewees.

The interviews were semistructured, and a topic guide was used to prompt discussion on topics relating to the person's use of the forum, their experiences of sharing online, relationships with other forum users, and experiences and perceptions of how trust and empathy operated in the forums. A flexible schedule was used to guide participants to issues of relevance to the study, but also to give freedom to explore other aspects of importance to participants. The topic guide included asking participants about their use of the Internet and online health forums; their experiences of online sharing of information, resources, stories, emotions, and experiences; their relationships with other people they met in the forums; and how they experienced trust and empathy in relation to other forum users. The duration of the breast cancer interviews ranged from 49 minutes to 2.5 hours, with an average length of approximately 1 hour. The duration of the MND interviews ranged from 1.3 hours to 2.4 hours, with an average length of 1.7 hours; this longer duration reflected the slower pace of communication for people with MND. All interviews were audio-recorded and fully transcribed. After each interview, field notes were taken by JE to document any immediate contextual and analytical insights.

Data Analysis

The data were analyzed thematically following the methodology outlined by Braun and Clarke [33]. In the first stage (familiarization) two researchers (SD and SH) read through the datasets (forum posts, interview transcripts, and field notes). SH took the lead on analysis for this paper and, in the next stage, discussed initial ideas with the interviewer (JE) and the lead investigator (PB). Initial coding was undertaken manually to

give a greater immediacy to the data sources. Comparisons were made with earlier initial coding developed by JE. A subset of data were coded independently by SH and SD to check consistency of coding and data interpretation. The data were then analyzed thematically by SH using NVivo 10 software, with interpretation of codes discussed with JE, PB, and SD. Codes were grouped into themes and then reviewed and refined by rereading data extracts, thus ensuring a fit between the datasets and interpretation. Data interpretation was discussed in an ongoing dialog between SH and SD to further ensure that analysis reflected participant experience. Findings were also discussed with PB. The analysis was also informed by interactions with members of the public at dissemination events for the study, and who shared experiences of online health forums.

Data analysis from the interviews yielded a rich understanding of the role of empathy in online communication and in these online health forums, with a variety of perspectives (generating a breadth of understanding) and a detailed understanding of themes (depth of understanding). It was unnecessary to seek further interviewees.

Analysis of Both Data Types

The datasets (interview and forum posts) were analyzed separately and then together to gain a better understanding of how empathy operates within the forums. This interplay between datasets enabled the study to explore a greater range and depth of understanding because each data source revealed both overlapping and differing aspects.

We did not seek to understand empathy at a fixed point in time. The data sources were not anchored to one single time point. Interviews and forum samples covered a range of time frames, which did not necessarily overlap. However, this was not considered inconsistent given that memories and experience bridge a wide time frame. This approach was considered appropriate to gain a broad understanding of how empathy operates within the forums. These varied time frames do not preclude triangulation of data sources because it provided a means of exploring consistency of themes across the differing sources and times.

Themes were informed by both sources. The theme "transformative experience," for example, came initially from interview data where interviewees from both forums gave vivid accounts of life-changing experiences of illness onset and diagnosis. These events happened "off-stage" from the forum. The interview data provided a depth of understanding of the devastation caused by diagnosis, and the process by which this both created a need for empathetic understanding and drew interviewees to the forum, where they could share their story with others in the same boat. These events were often summarized in introductory posts on the forums.

Table 2. Participant characteristics (N=19).

Forum and participant (pseudonym)	Age range of person with illness	Gender	Interviewee	Length of time since diagnosis	Period/time of joining the forum	Current level of activity on the forum
Breast Cancer Care forum						
Anne	50-59	Female	User	~15 months	Joined before formal diagnosis	Posting daily and checking posts
Beth	40-49	Female	User	3 years	At biopsy	Less frequent contact with forum; moved on to Facebook group with friends
Christine	50-59	Female	User	~13 months	1 year ago (joined chemo thread)	Checking posts and posting daily
Danielle	40-49	Female	User	~4 months	~4 months ago	Checking post every 3-4 days
Eleanor	50-59	Female	User	~2.5 years	2 years ago	Checking posts 2-3 times a week
Frances	50-59	Female	User	~16 months	Started using the forum at chemo	Checking posts twice weekly
Gayle	50-59	Female	User	~18 months	Started using the forum at chemo	Checks posts once or twice a week
Hazel	60-69	Female	User	~3 years	Joined ~19 months ago	Accessing forum at points of worry
Isobel	40-49	Female	User	~15 months	Joined chemo group ~15 months ago	Regularly checks posts
Janice	60-69	Female	User	~13.5 months	Joined chemo group (~13 months ago)	Regularly checks posts
Kathryn	50-59	Female	User	First diagnosis: 23 years; second diagnosis: 4.5 months	~4 months ago	Checks posts daily
Libby	40-49	Female	User	~10 months	Since chemo (~8 months ago)	Checks posts twice a day
Nancy	40-49	Female	User	~7 months	Since diagnosis (~7 months ago)	Lurker (reading posts twice daily)
Olivia	50-59	Female	User	~5 years	Since diagnosis (~5 years ago)	Checks forum every 6-8 weeks
Motor neuron disease forum						
Pippa & Michael	60-69	Female	User and partner	10 months	First went on forum 3 months before formal diagnosis	Daily user
Robert & Meg	60-69	Male	User and partner	~19 months	Joined forum at unidentified point and become active user 3 months ago when realized the disease was becoming more aggressive	Daily user
Sue	70-79	Female	User	5 years	Joined forum 4.5 years ago	Daily user
Terry	70-79	Male	User	~4 months	Joined forum soon after diagnosis	On the forum most days
Vincent	60-69	Male	User	~3 years	Forum member for 3 years	Very active—checking posts 2-3 times a day

Ethical Considerations

Both forums are openly accessible by anyone choosing to read the message boards online, although users are required to register and log in to actually make posts. At the time of the study, terms and conditions in each forum stated that posts were publicly visible and, with permission of the charitable organization, the posts may be used for research purposes. Recommendations provided by the University of Sheffield's Research Ethics Committee regarding the forum data were followed, with steps taken to preserve user anonymity by removing individual identifiers and changing details that might identify individuals or organizations. When preparing material for publication, we have reworded forum posts carefully to retain their original meaning and nuance while ensuring that phrases cannot be used in Internet search engines to trace quotations back to individual users.

Potential interviewees were provided with study information sheets and a copy of the consent form at least 24 hours prior to the interview to obtain informed consent. Interviewees were informed that they could stop the interview and/or withdraw from the study at any time up until the point of publication. We were mindful that participants may become distressed during the interview, and thus a protocol was devised with both charities to manage this sensitively and ensure, when necessary, participants were appropriately signposted to support. To prepare for interviews with individuals with MND who may experience communication difficulties, JE spent time discussing appropriate approaches with the Motor Neurone Disease Association.

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

Through our analysis, we developed a conceptual framework representing how empathy developed and operated within the forums. We found that empathy was built on shared experiences and connections (Textbox 1). In this paper, we explore these themes, shared experience, and connections, and also discuss their subthemes to highlight how empathy develops in online health forums. An additional theme—knowledge—will be discussed in further work.

Empathy Built on a Shared Experience

Users of the forums had one thing in common: that in some way their lives had been affected by the diagnosis of a life-threatening health condition. The majority of users were individuals who had been diagnosed with the health condition, although some experienced this indirectly as family caregivers. This bond of a shared experience and, most importantly, of knowing what it felt like to receive a diagnosis and live with a serious health condition, formed the shared emotional backdrop and common understanding within both forums.

Across each forum, users shared some similar experiences and emotions relating to diagnosis, but the process from becoming aware that they had a serious illness to diagnosis was quite different for the two conditions. Users of the Breast Cancer Care forum typically described a speedy transition to the world of ill health, going from the assumption that their breast lump would be benign, to receiving a diagnosis that was often both unexpected and shocking, to then often having to adapt swiftly to treatment regimens:

...the shock of the diagnosis, the fact that the treatment was starting so quickly, the operation and all of that, um, and it was kind of just trying to [pause] internalize it all and make sense of it and deal with all the different emotions and kind of my kids and my husband and everything... [Christine, breast cancer interviewee]

The process of diagnosis was typically slower for MND participants: there was an awareness of things going wrong with their body and then the uncertainty of undergoing diagnostic tests before a diagnosis was eventually made. However, receiving the diagnosis was still a devastating blow and, for some participants, the hopelessness of their situation was made worse by the way the diagnosis was communicated:

...and the neurologist said to me...in a very blunt fashion... "I think it's MND." I'll make a second opinion and I asked, "What's MND?" Because I hadn't a clue. Um, he said, "Well, let me say you'd best go home and sort your affairs out and do what you want to do because," he said, "two to five years." [Vincent, MND interviewee]

For both users with breast cancer and MND, the transition to ill health and diagnosis was a traumatic experience as they sought to deal with the emotional impact and to live with uncertainty. At this point of need, most participants found the forums by chance by searching for sources of help on the Internet. Only two interviewees in our study, one with breast cancer and one with MND, were directed to the forums by a health professional.

Introductory posts on both forums summarized people's experiences, offering an important starting point for forum users to reach out and connect:

Thank you everyone for sharing all of that. Everyone's experiences are really important so other people can empathize with... [moderator post, MND forum, in response to someone's post]

It's Saturday 16th June & I was diagnosed (grade 3) last Wednesday. The tumor is 1.5 cm, which sounds not too bad to me. I'm booked in to have a [name of procedure] op next Weds. At the moment, it's difficult to come to terms with the speed of it all, and I feel as though I am totally in limbo. Just waiting for Wednesday really. [breast cancer forum post]

Textbox 1. Thematic map of themes and subthemes for how empathy is built with the breast cancer and MND forums.

Theme 1: Shared Experience
1. Transformative experience
Theme 2: Connections
1. Forum community
<ul style="list-style-type: none"> • A meaningful space • Feeling set apart from others
2. Similarity
<ul style="list-style-type: none"> • Different but similar • Drawn to and seeking out similarity • New spaces for empathy
3. Relationships
<ul style="list-style-type: none"> • Building relationships • Friendship
4. Feelings
<ul style="list-style-type: none"> • Expression of feelings • Emotional impacts

These quotations demonstrate the need for individuals to share their experiences and emotions with other people at an early stage in their diagnosis. It also gave fellow users the opportunity to provide empathic responses. This information sharing was fundamental to the process of building connections between forum users and provided a basis on which to express, offer, and experience empathy.

Empathy Built on Connections

The second theme, “connections,” explores the means by which empathy is built through differing types of connections that are developed within the forums. The first aspect of this was having a sense of being connected to online support communities (eg, within the Breast Cancer Care or Motor Neurone Disease Association forums), as well as the wider community of people living with these conditions:

I do see it as a supportive community... [Terry, MND interviewee]

It is important to remember is that we will all be helped and supported through our distressing experience by the others on this site who have or are “walking in our shoes”—and for which I have found a godsend! [breast cancer forum post]

Within this theme, we explore key attributes of the forum that drew interviewees to the forums and enabled them to form an attachment to the people within these spaces.

Connected by Forum Community

Finding a Meaningful Space

Interviewees with both health conditions perceived the forums as meaningful spaces where they could share their own

experiences and emotions with people in a similar position and receive an empathic response:

I’ve certainly found it useful myself, certainly to know that there’s people there who really understand what you’re going through, your thoughts and worries, et cetera. [Anne, breast cancer interviewee]

You need to be around people who are rowing in the same boat trying to keep afloat in the face of real adversity and on the forum you find this empathy because it is so unique. There is nowhere else like it. [Pippa, MND interviewee]

There was also recognition that the forums could be a unique source of information as well as an emotional support, provided by others undergoing similar journeys and who shared their own experiences and information:

[The forum is] a safety net of, you know, good, solid information and the most enormous amount of kind of warmth and support from other women, yeah. [Olivia, breast cancer interviewee]

We all have a real need for this forum as a platform for our fears and anxieties, for advice, for guidance, for solace, and for friendship. [Pippa, MND interviewee]

All the interviewees perceived this support to be useful and relevant to them, and it was also considered beneficial in helping them through difficult situations. Both forums were perceived as supportive and welcoming spaces, offering warmth, comfort, and human understanding. These qualities made the spaces conducive to empathy, as shown in [Table 3](#).

Table 3. Qualities that made the communities conducive to empathy (perceptions derived from interviews with people with the health conditions).

Elements	Breast Cancer Care forum	Motor Neurone Disease Association Forum
Support	Source of information (a repository), expertise, help, advice, emotional support, saying what I needed to hear, a means of survival	Help, advice, assistance, finding the answers, friendship, usefulness, support (a support group), a lifeline
Community spirit	Comforting, welcoming, warmth, human contact, checking up on people	Warmth, caring, peaceful, sensitivity, give one another a cuddle, hope
A space to go to	A safety net, picking people up when they are down, a place to go in dark moments, sharing problems	A space to share, an outlet
Tone	Noncritical, nonthreatening, supportive, nonjudgmental, respectful, positive	Nice people, core people, jokes and banter, positive, respectful, but at times it could be contentious
Understanding	Unique understanding, understanding that mental well-being is important	A unique understanding, sharing the same boat

These humane qualities were especially appreciated by participants (particularly those with MND) who had felt this quality lacking from some of their interactions with health professionals. Human contact and connection was a key factor in the sense of attachment that participants felt toward the forums. Although they were connecting to an online space where they could not see others within the forums, there was a common perception that they were joining a group of people sharing the same situation:

...it's kind of human contact with people in a similar situation. [Gayle, breast cancer interviewee]

I just thought there was a lot of support on it [the MND forum] from people who were in the same boat. [Pippa, MND interviewee]

Feeling Set Apart From Others

Interviewees from both health conditions reported that ill health distanced them from friends and family, who had not been through the same transformative experiences. This was also mentioned in posts on both forums:

The whole breast cancer DX [diagnosis] turns life upside down and sadly, some people (including family) just don't get the impact it can have on the mind and body. [breast cancer forum post]

...I've got...family support. I've got friends' support. Um, I've got...medical support,...but this [the forum] is another form of support because none of the aforementioned people have got this disease, that's it. That's the difference. [Terry, MND interviewee]

This contrasted with views about members of the forum communities. Interviewees perceived that forum users understood what they were going through because they all had lived through, and therefore shared, these transformative experiences. This enabled participants to share and express concerns, experiences, and feelings with other forum users that could not necessarily be said to family and friends for fear they would not be understood in the same way, and to express things that could only be said to others sharing that same situation:

The amazing thing about talking to fellow breast cancer sufferers is that they totally understand what you're talking about, no matter how much you think you're not making sense! [breast cancer forum post]

...you know, you feel an affinity with the other people on the forum and you know they understand what it is that's happening to me... [Vincent, MND interviewee]

Some interviewees talked about sharing experiences with forum users with an emotional honesty that they were unable to extend to conversations with family and friends. These honest exchanges were also encouraged by being able to write and share their feelings within an anonymous space. Some participants felt constrained in sharing feelings with family members because they were striving to keep their feelings in check to present a positive face to others close to them who they wanted to protect. Participants from both forums valued the opportunity to offload feelings and fears in forum posts—to express what could not be said to family and friends for fear of burdening them:

I think without the forum I would have had huge depression here because there's nobody for me to share anything with and I can't bring my husband down all the time. He doesn't understand everything anyway... [Christine, breast cancer interviewee]

I don't feel comfortable with talking to [my family] about what the future holds for me...I'm more prepared to share my feelings and fears with the forum members than my immediate family, just because it—I know it upsets them more than it does me. [Vincent, MND interviewee]

The interview participants indicated the importance of connecting with a supportive community of others who share similar experiences; they understood very well the emotional impact of living with a serious illness and demonstrated a willingness to share information about their own experiences, which were of value to others. This depth of understanding and reciprocity of emotional exchange opened up opportunities for empathic discourse. Our data suggest that empathic qualities are enhanced further by the atmosphere and ethos of the online space, with users of both forums clearly valuing the comforting and humane qualities they found within the forums. In addition to this, some users experienced a greater depth of empathic connection when they formed connections with individuals or groups of users with whom they shared similar circumstances.

Connected by Similarity

Participants from both forums frequently reported feeling a connection to users with whom they felt a shared similarity. These connections operated on different levels, from a fundamental understanding of sharing the same general diagnosis, to an even greater mirroring of experiences, such as sharing a specific breast cancer diagnosis or experiencing similar symptoms and/or rates of deterioration due to MND. This influenced empathic communication, as users reported that they felt better able to converse with those they perceived to be similar to them. This connection stimulated conversation, users were more prepared to invest time in conversing with similar users, and this increased the potential for conversations to develop, increasing the likelihood of empathetic exchanges.

Different but Similar

Although people within the forums shared the same broad diagnosis, there were many individual differences in terms of experiences of living with a health condition (eg, varied pace of deterioration in MND, different stages, and types of breast cancer), the treatment and services received, family circumstances, and attitudes:

...it just shows, although all of us may have cancer the similarities end there. Our diagnoses are different and how we respond to treatment is different, and our opinions are all different, let's welcome that. [breast cancer forum post]

...everybody is so different, even if they have the same diagnosis... [MND forum post]

Although people were aware that these differences existed, the connections went beyond these, operating at a more profound level:

I don't know. It's just you—when you related to people on such a sort of deep and personal and painful part of your life, you just—it just makes a connection, um, that goes beyond the sort of superficial really. [Janice, breast cancer interviewee]

Really deep connections because there is no hope, no cure. We are all fighting the same fight with the same enemy. [Pippa, MND interviewee]

The quotations indicate that although users were aware of individual differences, these were overridden by the shared experience of living in extremis and often users formed deep connections. However, it was also common for interviewees to talk about strong bonds forged through greater similarities.

Drawn to and Seeking Out Similarity

Within both forums, people sought or made connections based on similarity. These similarities varied, such as living in a close geographical location, shared hobbies, past history, family circumstances (eg, children of the same age), or shared attitudes. Users either identified similarities while reading posts and, from this shared similarity, conversations grew, which could potentially lead to friendships. Other connections were sought, with users asking specifically if others shared their experience or situation, or accessing spaces set aside for users with the sought after characteristic. This was evidenced by interviewee

data in which participants talked about wishing to connect with others who were very similar to them. Within the MND forum, connections were sought based on shared problems or the pace of their deterioration, and these provided the necessary bridge between people often living without face-to-face contact with others in the same situation:

There have been a few who seem on a similar journey so they make a lasting impact and one I regularly contact by PM [Sue, MND interviewee]

It was apparent that some Breast Cancer Care forum users may seek specific connections, posting detailed descriptions of their situation and diagnosis in introductory posts, in order to find users sharing similar circumstances. This was confirmed by interview data, with participants expressing a need to connect with similar others. Moreover, the Breast Cancer Care forum created particular demarcated spaces to connect people sharing specific circumstances or situations. One powerful example of this was the monthly chemotherapy threads that were set up specifically for users (overwhelmingly women) who were commencing chemotherapy within a specific calendar month:

[We were] all going through chemo at the same time. So you were able to talk about symptoms and how you were feeling and that kind of thing... [Isobel, breast cancer interviewee]

This provided a context-specific exchange of empathy that was based on temporality. The fact that these particular forum users were experiencing the same kind of treatment at the same time gave them a synchronized experiential basis on which to develop very strong relationships and form subgroups within the forum environment.

New Spaces for Empathy

Users could create their own spaces within the forums by requesting a new space within the forum (eg, a themed board, starting a new thread, exchanging private messages, or moving conversations out of the forum and setting up their own closed Facebook groups). This desire for new spaces came from both a perception that people within the wider forum would not necessarily share or understand the same concerns, and simultaneously a need to connect with others sharing the same particular situation, concerns, approaches, etc. In the following quotation, Olivia described how, as a lesbian woman going through breast cancer, she felt a need for a separate space:

...well actually, there are things that people don't quite get [within the wider forum] and it would be, you know, really nice to have our own forum. So Evie and I between us requested, bullied, pushed, battered Breast Cancer Care until they did put a little bit up [a space for us]... [Olivia, breast cancer interviewee]

Olivia had felt an initial reluctance to mention her sexuality within the wider forum, and her first post had felt like “coming out.” For Olivia, this new space meant that lesbian and bisexual women coming to the forum would feel both welcomed and accepted by others who could perhaps more readily empathize with their experiences more broadly. The space encouraged conversations of particular relevance to this group so they could be had without the need to consider reactions from the wider

community. This lack of constraint opened up empathetic communications with others sharing a similar experience, creating a space within the online space of the forum.

Users also set up Facebook groups; this appeared to be more common for breast cancer participants. There were many breast cancer Facebook groups. Users set up groups to talk about experiences that were unique to them, such as a specific cancer diagnoses. For Isobel, the opportunity to connect with others sharing the same diagnosis provided a space where she was better able to open up and share with others who she considered could have greater empathy for her situation:

I think it was mainly because they'd all gone through the same cancer I had and they'd all gone through very similar treatments to what I had and they were dealing with very similar fears to what I was...I think for me it was just...the kind of emotional...you know, fear of recurrence or whatever was easier to discuss with the people who were facing very similar situations to me rather than in a wider group... [Isobel, breast cancer interviewee]

Users sought to connect newcomers to these spaces, by signposting to the appropriate group. This signposting role was an important means of building empathy by enabling users with a particular situation/characteristic to connect with similar others.

Invitations posted to the forums about joining Facebook groups emphasized certain advantages offered by these alternative spaces—primarily that Facebook was private and the closed groups offered a secure space to share as the messages posted could not be read by anyone who was not a registered member of the group:

...Josie has set up a FB [Facebook] group and many of us in the December group are on it so no other friends or family can go on it... [breast cancer forum post]

Being away from the forum allowed forum users to be less anonymous with others who joined the Facebook group; the Facebook groups were not moderated by staff from the forum organization and personal details could be exchanged. However, although some breast cancer interviewees felt that Facebook offered a space to share with greater openness, there was also awareness that closed Facebook groups excluded some users and took valuable conversations away from the forum.

Finally, within both forums, personal messages sent between individual users via direct messaging and email offered an important means of building more personal communications. Users sent messages directly to others whom they perceived needed greater support, giving them the option of further contact via direct messaging or email. Frequently within both sites, this offer was made without obligation, giving the user the choice to make contact only if they wanted. Personal messages were used within both forums to talk about specific issues, which could not be shared openly within the forum, and to expand and discuss issues at greater length. Personal messaging was also used to express concern. If users were absent from the forum, or were known to be going through a difficult time, members

would send a personal message checking up on that person and offering support:

...sometimes one of the members will have gone quiet for a week or so or will have posted something that says, you know, we're feeling really low about something then you don't hear from them for a day or two. So I'll just send a little message to say are you alright, you know, do you want to talk? I'm here [Anne, breast cancer interviewee]

I messaged her on FB (Facebook)...Getting worried now...hope gets back soon. I miss her.

Hi Rosie, I have also pmd [private messaged] her xx [MND forum posts]

These more individual messages differed from those shared on the open forum by their personal nature, focused as they were on offering particular support to one person. The technologies provided a private space in which friendships grew, and where intimate and personal conversations took place.

Connected by Relationships

Empathy is a relational experience and it was evident that it was something that developed through forum relationships. Although some users preferred a matter-of-fact relationship with the forum using it as a source of help rather than friendship, the majority of interviewees with both health conditions felt a strong bond to people within the sites. This is not to say that those with a more pragmatic involvement did not feel empathy for other users; however, relationships encouraged a greater depth of empathetic connection.

Building Relationships

Over time, relationships developed in both forums. Users connected with one another as described previously, and friendships grew as they shared information about their situation and lives. Participants were drawn to people with whom they felt a particular connection or affinity. Interviewees described “gelling” or “connecting with” others: one breast cancer interviewee described feeling drawn like a “magnet” to her online friend. In the monthly breast cancer chemotherapy groups, initial sharing focused on treatment experiences and the quickly developed to encompass chatting about a whole range of life experiences. A willingness to share more holistically became “culture” for the groups and increased opportunities of dialog between those participating as well as creating more scope to demonstrate empathy (eg, remembering to ask how a holiday was, whether that glass of wine was enjoyed). During early days of treatment, users typically corresponded daily with the group, and with time and sharing, users came to know one another as friends. These friendships could be intense because they were founded at a time of extremis, and thus there was a depth to the relationships that facilitated sharing and empathy:

I suppose it started off with a—bit all about treatment and cancer and all the rest of it but a lot of the time it isn't now and—and we've also been through a lot. Individuals have shared lots of things on the group, things that have happened to them. I mean...one other woman lost her father, another one whose father was

extremely ill...you get to know things about people that you would have to be really—you'd have to have quite a sort of established friendship with other people to get to that point but because you go in on this deep level, it's sort of much easier to talk about those things and share them. [Janice, breast cancer interviewee]

This same process of building friendships through sharing experiences also occurred on the MND forum. Users built up a sense of the people they were communicating with through sharing, gaining a sense of their personality, and their inner thoughts:

...she gets to know the people, gets to know their personality, and their thoughts and their fears [Michael, husband of Pippa MND interviewee]

Two of the MND participants spent a great deal of time communicating with their forum friends, and this constancy strengthened bonds and deepened relationships:

I am in constant touch with them all. [Pippa, MND interviewee]

These relationships, built on empathy and shared experiences were very important to forum users and could lead to long-term friendships.

Friendship

Friendships operated at different levels, within groups or communities, or in one-to-one relationships. The majority of interviewees described their relationships with other forum users as friendships, sometimes qualifying the description, for example, as forum friends or cancer friends. Some interviewees were keen to emphasize that these were friendships in the truest sense (ie, these friendships were felt, emotional bonds). Some MND interviewees talked about feeling a closeness and affinity to their forum friends, bonded by the unique experience of MND. The majority of interviewees had never met these friends face-to-face and did not have plans to do so, but they supported one another within the online spaces. It was not uncommon for interviewees to remark about the paradox that these forum friends knew more about them than some their offline friends:

...I mean, I always remember one of the people on the forum just sort of saying who would have believed you could get so much support from a computer and a bunch of strangers? Which I think is an absolutely wonderful quote, um, and actually is very true...we are in a sense strangers but we also know each other better than you probably know most—a lot of people in your life really. It's a very strange relationship. [Janice, breast cancer forum interviewee]

Conversations were found to move away from the forum so that private thoughts could be expressed without fear of being seen by others:

There was another lady that, um—she was in a different group than I was at the time and we actually still email each other regularly now, um, and I suppose she's the only person that I have really opened up to and likewise her to me. We've never

met. She lives up north—um, she was a couple of months behind or one month behind and we used to talk and then we'd talk about our darkest moments and our fears and, you know, our families and things. So she's—to me, it's more of an intimate relationship. [Frances, breast cancer interviewee]

Let's take this thread to email and let's keep in touch. [MND forum post]

Friendships provided a space for empathy, where the connections and understanding enabled empathic feelings to grow:

Yeah, understanding somebody, empathy, all those terms you'd associate with a friendship... [Vincent, MND interviewee]

...I feel for them when things are going bad and I'm glad for them when things are going well and I enjoy kind of talking to them. [Isobel, breast cancer interviewee]

Friendship motivated users to act in empathic ways—watching out for one another and coming to the aid of friends in moments of need:

...we're all going through similar experiences and I've made a number of...friends, err, who I would do whatever I can to assist...physically...or emotionally... [Vincent, MND interviewee]

I lost my father in December and within the—put in a message—you know, I put oh, I've just had a phone call I didn't want. You know, I've just—Dad's just died. I had about 13 messages within an hour. So—you know, it's—it's been a huge support. [Christine, breast cancer interviewee]

These were bonds of mutual support, whereby users gave and received comfort and support within the context of an ongoing relationship.

Connected by Feelings

Feelings were a strong theme within both forums, mentioned frequently both within interviews and forum posts. An emotional understanding of ill health formed an important means of connecting users. The expression of feelings and emotional vulnerabilities formed both a language of empathy and an empathetic cue for users to provide support to others in need. These remarkably honest and open expressions of feeling were encouraged by the anonymity of conversing within an online space. However, some interviewees were unaware that posts could be read by anyone searching the Internet and not all users felt comfortable with emotional expression on the forum. This view was expressed by a minority of interviewees reflecting their worries around emotions getting out of hand:

...I don't personally discuss my deepest fears and how I feel on there...I think because it is opening the door again, you know, so once [unclear], you've got to deal with those emotions. [Frances, breast cancer interviewee]

I just think the forum is for sharing experiences and really private matters about how you feel needs

proper counseling otherwise it could get out of hand.
[Robert, MND interviewee]

Both forums had a strong emotional undercurrent because people within these spaces lived with fear and uncertainty, but the depth of distress was perhaps more evident from these comments from one person within the MND forum:

I worry about my family because I won't be there for them.

I have to leave a room to spare people from my uncontrolled emotions.

I cried after the diagnosis, my wife held me like a small child.

I used to laugh and tell jokes, now I fade into the background so others don't have to slow conversation down whilst I tap away on the wretched smartphone. People look with sympathy and I want to scream.

For MND users, although deterioration and ultimately death were inevitable, they also faced uncertainty in their near futures about the nature of their decline and the speed with which it would happen. Users were brutally reminded of the reality of their situation when other users died, and although members of the Breast Cancer Care community also lost their lives to the disease, for MND sufferers the nature of the condition meant that end-of-life issues were more imminently germane for them. The fact that these losses were documented within the forum and experienced by the users affected the empathetic tone of the community. Members were bonded by their grief, the knowledge of their own mortality, and the loss of the person they once were, as the disease stripped away previous normalities. This made for a profound emotional connection with others sharing the same fate:

There is a tremendous empathetic bond between the forumites. We share a life sentence. It cannot be more powerful than that

The feeling between us all on the forum has been strengthened through all these deaths. It is tangible.
[Pippa, MND interviewee]

Within the Breast Cancer Care forum, death tended to be discussed within particular spaces, such as the Living with Secondary Breast Cancer end-of-life board; this effectively shielded the wider forum from the experiences and meant that users who were at a different stage of the illness (eg, awaiting a diagnosis) did not encounter these difficult issues when they were not ready for them. Forum users could choose whether they wished to view these discussions or they could avoid them altogether. This effectively created spaces within spaces, in which users who were facing difficult situations could discuss these openly with others, while the wider forum membership were protected from these conversations.

Although participants felt sadness or grief on hearing about the deaths of other users, the experience also highlighted individual vulnerabilities:

Um, someone on the Younger Breast Cancer Network died last weekend and I have to say I did shed a tear, even though I've never met her, um, just because of

the really sad story of it all. So um, I don't know because it's a bit mawkish sometimes to read—read the sadder things that happen but you have to be aware that it's a possibility... [Libby, breast cancer interviewee]

Users from both forums understood the emotional cost of living with ill health. The emotions and tensions of living with uncertainty provided a language for empathetic expression.

Expression of Feelings

Discussing feelings was a recurring theme within both health forums. Some expressions used to describe feelings were common in both forums, such as the emotional roller coaster many experienced at the time of diagnosis or during other points of particular turmoil. Feeling alone was also a commonly expressed emotion, as people sought to deal with the isolation of living with health conditions that separated them from others, both emotionally—and especially for people with MND—physically:

I imagined saying goodbye to my children, thought about the instructions and letters I would need to leave...Was wondering about at 3 am scared lonely frightened I am sure that most of us feel like this but isn't it horrible!! [breast cancer forum post]

My legs failed first...but I have found it harder now my arms are deteriorating badly. I don't know if anyone here has ever felt like they're alone but especially at my age it is hard to get my head round this day to day. [MND forum post]

Emotional venting provided an important means of “letting go” of emotions, of catharsis, an expression more powerful and meaningful because it was undertaken with an audience of users going through the same thing. Users also asked others to validate feelings, often asking specifically if anyone else within the forum shared the same feeling. This request occurred more frequently within the Breast Cancer Care forum, but was also found within MND posts, as illustrated in the previous quotation. The need to check out emotions seemed to occur at times of great emotional intensity or when users experienced unexpected feelings:

All this is good news [now that I have got to the end of treatment] and yet why am i so fed up? People say how well I've done and now it's nearly over, but I feel like crying all the time like I did in the beginning.
[breast cancer forum post]

These types of post often motivated forum members to reach out and to offer reassurance that the emotions described were normal and shared by others:

...you'll find somebody [who can relate to the emotion] who can kind of pitch in say—validate that yes, you're not going crazy. [laughing]. You're just being like that today or whatever it is, it just kind of—it's part of that roller coaster that you're just going to have to go through... [Danielle, breast cancer interviewee]

...Your [sic] not on your own feeling the way you do, I know I [expletive] well try very hard to hang onto

keeping sane, so don't think everything is down to you, it's what this [expletive] disease does to you...I understand the loneliness you feel even with people around don't help much, and seeing what we had now become what you have now... [MND forum post]

Users shared their own emotional experiences with the intention of comforting others. Libby (a breast cancer interviewee) spoke of sharing in order to calm and to reassure, "...to calm her (another forum user) down and say 'Well yes, totally understand what you're going through, been there...'" People from both forums sent encouraging messages to struggling forum members either sending "strength" or advising that the users should "keep strong."

Emotional Impacts

The emotional impact of serious illness was understood within the forum where all participants could relate to the impact of the illness on relationships and daily life. Participants described giving up jobs they loved and they discussed the pressures placed on families, as partners became carers, relationships were put under pressure, and friendships did not necessarily last. Some participants felt that there was not the same understanding of the emotional consequences of illness from people who had not experienced it first hand, but they were able to gain this from other forum users who had been through similar experiences:

um, you talk to somebody on this forum and he has a—perhaps a better emotional understanding of where you are, not just the physical stuff but, you know, he's perhaps been through the emotional side and with great respect to the medic—the medics, they might have seen it but they haven't done it. [Terry, MND interviewee]

...you see, the thing is the medics are very good here, err, but there's no emotional support. That is completely what's missing so you know... [Christine, breast cancer interviewee]

Members of both forums documented struggles to deal with the emotional impact of living with a changed body. The MND forum users reported physical losses as a result of disease progression, whereas users on the Breast Cancer Care forum posted experiences of coming to terms with a body changed as a result of therapies or surgery. They worked through the implications of this for social and personal identity as well as the ability to carry out everyday tasks. Some feelings were more likely to be expressed within one particular forum or space within the forum, for example, the monthly chemotherapy threads on the Breast Cancer Care forum. Frustrations regarding the physical realities of living with MND were commonly expressed, with users often describing the impact of the disease on their body and everyday activities. Part of this frustration was borne out of the "battles" that some users experienced to get appropriate support from services to manage their condition and also the lack of progress in finding a cure for MND:

...frustration is probably the only emotion I feel constantly; frustration, not anger. [Vincent, MND interviewee]

For those living with breast cancer, fear was commonly expressed (eg, fear of not surviving, fears about test results, and emotional strain of living with such fears day to day). When articulated, these emotional outpourings had particular resonance within these spaces and elicited great empathy from others:

...I'm reading this because I'm in the same boat—anxious and fearful especially at night. I wish I had more peace of mind... [breast cancer forum]

Discussion

Principal Findings

The aim of this paper is to develop a better understanding of how empathy develops and operates in online spaces where people share information, experiences, and emotions relating to living with a serious illness. Although other papers on empathy have tended to quantify and deconstruct processes of empathy documented on forums, our study provides a deeper understanding of the human experiences and human processes that build and foster empathy. Our approach of combining both interview and forum data enabled us to gain a new and deeper understanding of the processes of how empathy is developed and operates within the forums and therefore provides a novel contribution to current literature. The interview data enabled us to situate empathy building within the wider human experience of ill health and to hear directly from participants how they experienced empathy within the space, their thoughts, and actions. The forum data provided a demonstration of how these processes were enacted within the two forums. The analysis provides a unique insight into the development and operation of empathy from the perspective of participants with two very different health conditions. We found that although differences existed, there were points of similarity and key to this was the experience of uncertainty found in both conditions. These experiences both drew participants to the online forums and made for empathetic spaces. Our study found a common means by which empathy is built within both online health forums. Empathy emerges as a process beginning outside the forums in the shared experience of diagnosis, and then develops and operates within the forums through connections sought and made among users.

Participants experienced diagnosis as a life-changing event. This shocking and devastating event marked the transition from previous normalities, into an uncertain world [34-36]. The sense of devastation was particularly marked for participants with MND told that they had a debilitating condition without cure [29,36-38]. Participants with both health conditions described a period when they struggled to make sense of their situation, and felt the need for support and information [38-40]. Participants felt that there was insufficient focus on the emotional impacts of living with breast cancer and MND [41,42]. These intensely emotional experiences both drew participants to the forums and provided an emotional understanding which informed empathetic interactions with other users [15,43]. Other studies have observed that forum participation can be prompted by negative experiences within the offline world [44]. Interestingly, most participants in our study came across the forums while searching on the Internet

[45], and although they may not have been consciously seeking empathetic support, they recognized the potential value when they came across it.

Both forums were experienced as very human spaces, connecting people to a community of shared experiences. The communities were *felt*, and were perceived as warm and comforting communities, where members worked to support one another within a shared supportive ethos [46,47]. This feeling came primarily from the sense that they were connecting to human experience [14,48]. The dimensions of experience discussed within both forums felt more meaningful and true to the lived experience than interactions with health professionals focusing on medical aspects [42] or with their family where conversation may be constrained by fear of causing upset [12]. It is the case that some users can feel distant from their usual sources of support, most frequently family and friends [49]. Often they do not want to burden them with their fears or they are reluctant to share because they feel that they will not adequately understand the experience [12,50,51]. The useful and relevant sharing undertaken within both sites provided a unique source of support, of experiential and practical information, and emotional support [6]. This fit between support provided by the forum and what participants felt they needed echoes the idea of empathetic accuracy [22].

The informational and emotional support provided within both forums fit within the criteria of social support [52,53]. However, the added element of empathy from others with greater homophily profoundly altered the experience of giving and receiving social support within the context of both online health forums. Consider, for example, emotional support. The fact that users from both forums shared first-hand experience of what it is like to live with both conditions took this emotional support to a higher level. Users knew from the inside what it meant to live with the differing conditions and what is at stake, and used this to inform empathetic responses. Emotional support within this context was considered more meaningful (eg, users were able to use short-hand descriptions to describe situations and they knew others within the forum would recognize and understand). The emotional support was highly valued and trusted [17], coming as it did from an audience of others sharing similar situations and who took the time to provide support even though they themselves may be suffering negative impacts of ill health. Emotional support provided by both online health forums differed from that provided by family, friends, and health professionals. Family and friends could imagine/approximate what the person with breast cancer or MND was going through, but the people within the forums knew. Health professionals provide support based on clinical expertise and observations of patient experiences, but again this is based on indirect experience. Thus, within the communities created by both online health forums, users found a unique source of social support with empathy that could not be found elsewhere. This paper provides a deeper understanding of the importance of social support with empathy provided by peers within online health forums.

Most participants in our study experienced both health forums as very human, intimate spaces, within which they felt empowered to share personal information and experiences. Both

forums were understood as a community of people connected by the same traumatic experience, where human understanding and comfort emanated from shared experiences and from sharing their experiences with one another [1]. Sharing experiences, as a way to support others, is foundational in the process of creating empathy; users share personal experiences, information, and emotions in the hope that it can help others [19]. Empathy was encouraged by a willingness to express vulnerability and this was commonly found in posts shared on both forums. These open and honest narratives encouraged fellow users to reciprocate in kind, opening up spaces for empathy. Participants were encouraged to share by connections that they formed to other users, based on shared interests, and within relationships and via expression of shared emotions. Both forums provide a means of connecting with others at a one-to-one level or in subgroups with shared interests, and technologies enabled conversations to move in and out of the forum. Thus, empathy was built within different spaces, groups, and technologies, as users sought to work through the emotional and practical work of living with an uncertain future [54]. What was clear from our analyses was that sharing and making connections are dependent on each other: sharing facilitates the development of connections and making connections encourages sharing.

Within both spaces, users sought out connections based on similarity. Within the literature, similarity is considered a key facilitator of empathy (ie, people who are similar are more likely to empathize) [19] and to detect another person's feelings with accuracy [22]. Interviewees gave accounts of connections made with people in similar circumstances or with shared experiences, which provided opportunities to interact on a profound level. The level of similarity sought varied between the two conditions [55], with participants with breast cancer more likely to seek out others sharing a specific diagnosis or undergoing similar treatment [51,56], reflecting the varied types of breast cancer and differing treatment pathways. In such instances, participants were better able to open up and disclose fears, perhaps because shared risks and illness identity created a stronger tie strength [57]. The fact that users actively created new spaces to share with similar others demonstrates the need for spaces which provide relevant support and an environment which encourages users to unburden their fears.

Empathy was fostered by relationships developed in the forums. Participants described how they sought out or chanced upon an individual or group that they felt either an affinity or a connection to, because of a shared similarity. Relationships grew and deepened through acts of narrative sharing. Initial phases of sharing may have focused on the illness experience, but soon broadened to encompass everyday experiences [1,30]. Over time, users gained an understanding and a sense of the other person, their personality, circumstances, and illness experiences [14,19]. These acts of sharing and connection enabled feelings to grow for fellow users [30], and encouraged empathetic and supportive behaviors [15,19]. Friendships provided an emotional space for users to express their deepest fears in acts of unburdening [25,30]. Online friendships provided a buffer against the isolating impacts of ill health [35]; this was of particular importance for participants with MND, for whom speech and mobility difficulties may combine to inhibit

interactions in the offline world [2]. Conversations between friends were shared in the open forum, including within the spaces dedicated for specific groups, such as the monthly chemotherapy threads and the board for lesbian and bisexual women on the Breast Cancer Care forum, as well as in external spaces such as closed Facebook groups where privacy may foster even greater intimacy [58]. Although there is a risk that conversations within these private spaces might take valuable sharing away from the health forum [59], these alternative spaces were needed and supported practices in everyday life whereby people converse both within groups and more openly within private conversations.

Emotions were an important facilitator of empathy. The experience of ill health was “felt” by interviewees who described in vivid terms the emotional impacts of breast cancer and MND. This understanding of what it *feels* like to live through a journey from diagnosis into the unknown, and the wider impacts on family and previous ways of living, was shared, recognized, and understood within the forums [3,14]. Participants perceived that they gained an emotional understanding from the forums that was not found elsewhere because it came from both *lived* and *felt* experiences [60]. Feelings were anchored in experiences, thus enabling users to recall how they felt at particular times within their illness [30,61]. This motivated users to reach out and support others, and provide the type of support that was needed, particularly during times of great emotional need [19]. This recognition of feelings by others within both forums was a key element bonding users together into an empathetic community [60].

The study demonstrates the importance of time in the development of empathy. Participants from both health conditions described individual journeys, going from diagnosis, to finding the forums, and building connections and friendships within the forums. These processes occurring over time facilitated the development of empathy. Users took time to get to know one another. Empathy developed and grew within relationships. Thus, time both underpinned the development of empathy and intersected with the key enablers of empathy identified by this study. This interconnection between temporality and other key enablers is noted by a study exploring the development of trust in online health forums [17].

Empathy was built within both forums on the same building blocks of shared experiences and connections; however, the breast cancer and MND forums also differed in some respects. The differing disease trajectories altered how empathy was enacted within the two forums. The hoped-for trajectory in breast cancer of treatment and recovery was played out within the forum, with waves of new users joining the monthly treatment groups and working their way through varied treatment pathways. The emotional peaks and troughs of these individual journeys, and the fact that the majority of the users were female [23], all shaped how empathy played out within the space. The MND forum had fewer members and fewer boards within the forums, reflecting the rarity of the condition, the lack of a curative pathway, and the focus on supportive treatments and living with and managing the condition. Users were bonded by shared experiences; the debilitating impact of the conditions stripped away independence and narrowed

interactions with the outside world. This made for a distinctive type of empathetic bond [2]. The gender balance of the forum also differed. The majority of core members, posting most frequently, were male; however, the site itself was mixed gender [62]. There was also a greater presence of family caregivers on the site than that found on the Breast Cancer Care forum. However, despite these differences there were clear points of similarity, with participants from both conditions experiencing uncertainty. For participants with breast cancer, this often meant an ongoing fear of not knowing what would happen to them in the future (eg, treatment outcomes or risk of secondary cancers). Participants with MND knew that they would deteriorate, but did not know when and how fast this would happen, and how it would impact on their lives. People from both conditions dealt with this by coming together on forums to make sense of their situation within the empathetic space provided by the forums.

Strengths and Limitations

Our findings are strengthened by combining interview data and forum posts, the two sources providing differing views and perspectives [16,63] as well as new insights. Use of both datasets together strengthened our analysis and the validity of our claims [17]. This approach has particular value when researching a topic such as empathy, which is abstract and intangible [64]. We combined these differing methodological approaches with the exploration of two contrasting health forums and argue that this approach yielded a richer level of analysis and deeper understanding [63].

This paper adds to knowledge of how empathy is defined by users of health forums. The definition utilized was chosen because it matched interviewee participant definitions. Core elements of this definition—knowing and feeling—could be defined as emotional intelligence, informed by experiences specific to the health conditions. Users knew what others in the forums were feeling, recalling (often vividly) how they themselves had felt during the stage of illness identified or feeling a resonance with emotions described. Users were able to imagine themselves in the circumstances of others, using this understanding to inform compassionate and considered responses.

Interviewees were self-selecting and there is a danger that they represented an overly positive view of health forums, although participants did discuss both positive and negative issues. The interview data were combined with forum data, which provided insights into the perceptions and experiences of a wider body of forum members [65]. However, neither of these approaches can represent the views and experiences of individuals who choose not to share in online environments (eg, people who lurk and read posts but do not post themselves or who are digitally excluded). A further limitation is that there were fewer interviewees with MND, possibly reflecting perhaps the physical and communicative barriers to participation in research for people with this condition.

Implications for Practice

The study highlights areas of unmet need for individuals living with breast cancer and MND, particularly around emotional and informational support [37,38,66]. Forums provide a means of

fulfilling needs within a supportive peer-to-peer environment, and that aspect is particularly valued [65]. Study participants commonly found the forums while searching the Internet, rather than being signposted by health or social care professionals. The potential value of forums to patients and families may not be fully understood by health professionals [44,45], and additional factors, such as level of awareness and time constraints, may reduce the likelihood that professionals will sign-post patients to forums. Raising awareness of the potential benefits of online health forums among health professionals could help them to encourage patients to use these as support, to supplement the care that they, as health professionals, provide.

The study highlights important issues for organizations hosting online health forums. Our research brings a new understanding of how users operate within the forums, seeking support that they perceive as beneficial to fulfill emotional and practical needs. Hosts may consider how best to enhance forums to foster elements that users found most helpful and appealing. The human aspects were of key importance (ie, the supportive and warm atmosphere, the ability to connect with others who shared and understood, and to build relationships). Provision of useful information on practical support was also a key element. Thus, structures or affordances are needed to enable users to both find useful information and satisfy needs to connect with others [47]. This requires empathetic design, which should reflect the needs of users coming to the space, but also provides a means of enabling users to interact in an empathetic way within the space [67]. Design needs to take account of needs changing over time [50], for example, so that there are spaces available for different stages of an illness. Attention should also be focused on protecting the supportive atmosphere of the space [46] because this was perceived to be more conducive to empathetic interactions. Forum moderators might consider how best to moderate spaces to protect the supportive and warm ambiance.

Users interacted with one another in different spaces and conversed in different ways within the spaces, often sharing most private fears within a one-to-one setting. Thus forums should reflect these differing needs, providing access to differing levels of communication (forum, subgroups, one-to-one messaging). Some users may leave the forum space to form special interest groups within other platforms (eg, Facebook). Forum hosts may consider how best to interact with Facebook groups, given the popularity of this social media site [68] and the ease with which social media enables people to form groups independently. Health forums exist within this dynamic space, and thus it should not be considered a failure if users meet within the forum and move off into another setting. The informal, less-anonymous space provided by Facebook may enable users to get to know one another with greater immediacy, and thus encourage empathy. However, there are risks associated with interacting within a private unmoderated space [69] of which users should be aware.

Implications for Future Research

Further research on empathy in online health forums could examine barriers to the development of empathy and how these affect people's sharing of their information and experiences, as well as how they receive information from people with whom they have little or no empathic connection. Although our research has highlighted the importance of sharing in the development of empathy, and how empathy can lead to further sharing, there was little evidence from the discussion boards we sampled, or from the interviews, of how conflicts and disagreements within the forums affect empathy and sharing, or of the extent to which empathy acted as a buffer during disagreements and enabled people to continue to share experiences and emotions. Further research could also examine how empathy operates in relation to other aspects of the lives of people with life-threatening or terminal illnesses in relation to sharing information and experiences; for example, perceptions of the risks of sharing and self-disclosure, and the importance of trust [17] in the development of empathy. As new online platforms and functionalities develop, and people become more aware of issues relating to privacy, developing a better understanding of how empathy and trust operate will enable forum designers and providers to develop spaces in which users feel confident about sharing information, experiences, and emotions.

Conclusion

This study contributes new knowledge to the underresearched and important area of how empathy develops and operates in online environments by an exploration of two online health forums. Our study demonstrated that empathy develops through shared experiences and connections. The process begins outside the forums with the transformative experience of diagnosis. Forum users in our study were motivated to seek online support to meet emotional and informational needs unfulfilled by usual sources of support. They were often at a point of great emotional need, ready to both receive and give empathy. These empathic processes are developed through connections formed on the basis of shared needs, feelings, similarities, and relationships. Empathy was also fostered by a range of structural possibilities within the forums which gave users the opportunity and means to interact within public, restricted, and more private spaces, as well as within groups and one-to-one exchanges. The atmosphere and feeling of sites and perceived audiences were also important, with forum users sharing a perception of a virtual community of caring and supportive people, offering comfort and support. Our findings are of value to organizations hosting health forums that seek to offer support to individuals living with long-term and acute life-threatening conditions. Our findings show the importance of providing and protecting how empathetic interactions are formed and maintained, given the key importance of these interactions in encouraging sharing among forum users, which in turn nurtures well-being and resilience.

Acknowledgments

We thank the people who volunteered to be interviewed and shared their experiences with us. We thank Breast Cancer Care and the Motor Neurone Disease Association for their support in allowing us access to their online forums and enabling us to recruit forum users to take part in the interviews.

We are grateful to our colleague, Professor Nigel Ford, Information School, University of Sheffield, for helpful feedback on an earlier draft of this paper and to the four anonymous reviewers who provided helpful feedback for improving the paper.

The “A Shared Space and A Space for Sharing” project is one of several projects funded through the EMoTICON network (grant no: ES/M00354X/1), which is funded through the following UK cross-council programs: (1) Partnership for Conflict, Crime and Security Research (led by the Economic and Social Research Council), (2) Connected Communities (led by the Arts & Humanities Research Council), and (3) Digital Economy (led by the Engineering and Physical Sciences Research Council), in partnership with Defence Science and Technology Laboratory and Centre for the Protection of National Infrastructure.

Conflicts of Interest

None declared.

References

- 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](https://doi.org/10.2196/jmir.5260)] [Medline: [26965990](https://pubmed.ncbi.nlm.nih.gov/26965990/)]
- Stewart Loane S, D'Alessandro S. Communication that changes lives: social support within an online health community for ALS. *Commun Quart* 2013 Apr;61(2):236-251. [doi: [10.1080/01463373.2012.752397](https://doi.org/10.1080/01463373.2012.752397)]
- Bar-Lev S. “We are here to give you emotional support”: performing emotions in an online HIV/AIDS support group. *Qual Health Res* 2008 Apr;18(4):509-521. [doi: [10.1177/1049732307311680](https://doi.org/10.1177/1049732307311680)] [Medline: [18192435](https://pubmed.ncbi.nlm.nih.gov/18192435/)]
- Malik S, Coulson N. Coping with infertility online: an examination of self-help mechanisms in an online infertility support group. *Patient Educ Couns* 2010 Nov;81(2):315-318. [doi: [10.1016/j.pec.2010.01.007](https://doi.org/10.1016/j.pec.2010.01.007)] [Medline: [20144521](https://pubmed.ncbi.nlm.nih.gov/20144521/)]
- Zieband S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? *Milbank Q* 2012 Jun;90(2):219-249 [FREE Full text] [doi: [10.1111/j.1468-0009.2012.00662.x](https://doi.org/10.1111/j.1468-0009.2012.00662.x)] [Medline: [22709387](https://pubmed.ncbi.nlm.nih.gov/22709387/)]
- Zhao J, Abrahamson K, Anderson JG, Ha S, Widdows R. Trust, empathy, social identity, and contribution of knowledge within patient online communities. *Behav Inform Technol* 2013 Oct;32(10):1041-1048. [doi: [10.1080/0144929X.2013.819529](https://doi.org/10.1080/0144929X.2013.819529)]
- Han J, Shah DV, Kim E, Namkoong K, Lee SY, Moon TJ, et al. Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. *Health Commun* 2011 Mar;26(2):185-197 [FREE Full text] [doi: [10.1080/10410236.2010.544283](https://doi.org/10.1080/10410236.2010.544283)] [Medline: [21318917](https://pubmed.ncbi.nlm.nih.gov/21318917/)]
- Kingod N, Cleal B, Wahlberg A, Husted G. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. *Qual Health Res* 2017 Jan;27(1):89-99. [doi: [10.1177/1049732316680203](https://doi.org/10.1177/1049732316680203)] [Medline: [27956659](https://pubmed.ncbi.nlm.nih.gov/27956659/)]
- Cuff B, Brown S, Taylor L, Howat D. Empathy: a review of the concept. *Emot Rev* 2014 Dec 20;8(2):144-153. [doi: [10.1177/1754073914558466](https://doi.org/10.1177/1754073914558466)]
- Levenson R, Ruef A. Empathy: a physiological substrate. *J Pers Soc Psychol* 1992 Aug;63(2):234-246. [Medline: [1403614](https://pubmed.ncbi.nlm.nih.gov/1403614/)]
- Mazanderani F, Locock L, Powell J. Biographical value: towards a conceptualisation of the commodification of illness narratives in contemporary healthcare. *Sociol Health Illn* 2013 Jul;35(6):891-905. [doi: [10.1111/1467-9566.12001](https://doi.org/10.1111/1467-9566.12001)] [Medline: [23551016](https://pubmed.ncbi.nlm.nih.gov/23551016/)]
- Angouri J, Sanderson T. ‘You'll find lots of help here’ unpacking the function of an online Rheumatoid Arthritis (RA) forum. *Lang Commun* 2016 Jan;46:1-13. [doi: [10.1016/j.langcom.2015.10.001](https://doi.org/10.1016/j.langcom.2015.10.001)]
- Popham S. Developing empathy and expertise in online health forums. 2017 Aug 24 Presented at: Professional Communication Conference (ProComm), 2017 IEEE International; Jul 23-26, 2017; Madison, WI p. 1-9 URL: <https://ieeexplore.ieee.org/abstract/document/8013966/> [doi: [10.1109/IPCC.2017.8013966](https://doi.org/10.1109/IPCC.2017.8013966)]
- Preece J. Empathic communities: balancing emotional and factual communication. *Interact Comput* 1999 Sep;12(1):63-77. [doi: [10.1016/S0953-5438\(98\)00056-3](https://doi.org/10.1016/S0953-5438(98)00056-3)]
- Attard A, Coulson N. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. *Comput Hum Behav* 2012 Mar;28(2):500-506. [doi: [10.1016/j.chb.2011.10.022](https://doi.org/10.1016/j.chb.2011.10.022)]
- Seale C, Charteris-Black J, MacFarlane A, McPherson A. Interviews and internet forums: a comparison of two sources of qualitative data. *Qual Health Res* 2010 May;20(5):595-606. [doi: [10.1177/1049732309354094](https://doi.org/10.1177/1049732309354094)] [Medline: [20008955](https://pubmed.ncbi.nlm.nih.gov/20008955/)]
- Lovatt M, Bath P, Ellis J. Development of trust in an online breast cancer forum: a qualitative study. *J Med Internet Res* 2017 May 23;19(5):e175 [FREE Full text] [doi: [10.2196/jmir.7471](https://doi.org/10.2196/jmir.7471)] [Medline: [28536093](https://pubmed.ncbi.nlm.nih.gov/28536093/)]
- Preece J. Empathy online. *Virtual Real* 1999 Mar;4(1):74-84. [doi: [10.1007/BF01434996](https://doi.org/10.1007/BF01434996)]

19. Pfeil U, Zaphiris P. Patterns of empathy in online communication. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press; 2007 Presented at: CHI '07 SIGCHI Conference on Human Factors in Computing Systems; Apr 28-May 03, 2007; San Jose, CA p. 07 URL: <https://dl.acm.org/citation.cfm?id=1240763>
20. Kaufman S, Whitehead KA. Producing, ratifying, and resisting support in an online support forum. *Health (London)* 2018 May;22(3):223-239. [doi: [10.1177/1363459315628043](https://doi.org/10.1177/1363459315628043)] [Medline: [26851264](https://pubmed.ncbi.nlm.nih.gov/26851264/)]
21. Fan H, Lederman R, Smith SP, Chang S. How trust is formed in online health communities: a process perspective. *Commun Assoc Inform Sys* 2014;34(28):531-560 [FREE Full text]
22. Ickes W. Empathic accuracy. *J Pers* 1993 Dec;61(4):587-610. [doi: [10.1111/j.1467-6494.1993.tb00783.x](https://doi.org/10.1111/j.1467-6494.1993.tb00783.x)]
23. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. *Soc Sci Med* 2006 May;62(10):2577-2590. [doi: [10.1016/j.socscimed.2005.11.016](https://doi.org/10.1016/j.socscimed.2005.11.016)] [Medline: [16361029](https://pubmed.ncbi.nlm.nih.gov/16361029/)]
24. Siriaraya P, Tang C, Ang C, Pfeil U, Zaphiris P. A comparison of empathic communication pattern for teenagers and older people in online support communities. *Behav Inform Technol* 2011 Sep;30(5):617-628. [doi: [10.1080/0144929X.2011.582146](https://doi.org/10.1080/0144929X.2011.582146)]
25. Preece J, Ghazati K. Experiencing empathy online. In: Rice RE, Katz JE, editors. *The Internet and Health Communications: Experiences and Expectations*. Thousand Oaks, CA: SAGE Publications; 2001:237-260.
26. Preece J, Ghazati K. In search of empathy online: a review of 100 online communities. In: *AMCIS 1998 Proceedings 33*. 1998 Presented at: Americas Conference on Information Systems (AMCIS); Aug, 1998; Baltimore, MD p. 92-94 URL: <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1460&context=amcis1998>
27. Bronstein J. An examination of social and informational support behavior codes on the Internet: the case of online health communities. *Libr Inform Sci Res* 2017 Jan;39(1):63-68. [doi: [10.1016/j.lisr.2017.01.006](https://doi.org/10.1016/j.lisr.2017.01.006)]
28. Department of Health, Macmillan Cancer Support, NHS Improvement. National Cancer Survivorship Initiative Vision. London: Department of Health; 2010. URL: <http://webarchive.nationalarchives.gov.uk/20100809113601/http://www.improvement.nhs.uk/cancer/LinkClick.aspx?fileticket=4apVUSvGcow%3D&tabid=214> [accessed 2018-06-06]
29. Mistry K, Simpson J. Exploring the transitional process from receiving a diagnosis to living with motor neurone disease. *Psychol Health* 2013;28(8):939-953. [doi: [10.1080/08870446.2013.770513](https://doi.org/10.1080/08870446.2013.770513)] [Medline: [23464923](https://pubmed.ncbi.nlm.nih.gov/23464923/)]
30. Rubenstein E. "They are always there for me": the convergence of social support and information in an online breast cancer community. *J Assn Inf Sci Tec* 2014 Nov 06;66(7):1418-1430. [doi: [10.1002/asi.23263](https://doi.org/10.1002/asi.23263)]
31. Cancer Research UK. Breast cancer statistics URL: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer> [accessed 2018-06-06] [WebCite Cache ID 6zyCGIPIF]
32. Motor Neurone Disease Association. 2017. Brief guide to MND URL: <http://mnda.stage2.reason.digital/what-is-mnd/brief-guide-to-mnd/> [accessed 2018-06-06] [WebCite Cache ID 6zyCoMIZI]
33. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
34. Bury M. Chronic illness as biographical disruption. *Sociol Health Ill* 1982 Jul;4(2):167-182. [doi: [10.1111/1467-9566.ep11339939](https://doi.org/10.1111/1467-9566.ep11339939)]
35. Carel H. *Illness: The Cry of the Flesh*. Durham: Acumen; 2008.
36. Locock L, Ziebland S, Dumelow C. Biographical disruption, abruption and repair in the context of motor neurone disease. *Sociol Health Ill* 2009 Nov;31(7):1043-1058 [FREE Full text] [doi: [10.1111/j.1467-9566.2009.01176.x](https://doi.org/10.1111/j.1467-9566.2009.01176.x)] [Medline: [19659736](https://pubmed.ncbi.nlm.nih.gov/19659736/)]
37. Locock L, Mazanderani F, Powell J. Metaphoric language and the articulation of emotions by people affected by motor neurone disease. *Chronic Ill* 2012 Sep;8(3):201-213. [doi: [10.1177/1742395312443390](https://doi.org/10.1177/1742395312443390)] [Medline: [22457342](https://pubmed.ncbi.nlm.nih.gov/22457342/)]
38. Pavey A, Allen-Collinson J, Pavey T. The lived experience of diagnosis delivery in motor neurone disease: a sociological-phenomenological study. *Sociol Res Online* 2013 Jun 06;18(2):1-12. [doi: [10.5153/sro.2927](https://doi.org/10.5153/sro.2927)]
39. Fiszer C, Dolbeault S, Sultan S, Brédart A. Prevalence, intensity, and predictors of the supportive care needs of women diagnosed with breast cancer: a systematic review. *Psychooncology* 2014 Apr;23(4):361-374. [doi: [10.1002/pon.3432](https://doi.org/10.1002/pon.3432)] [Medline: [24677334](https://pubmed.ncbi.nlm.nih.gov/24677334/)]
40. Dervin B. Sense - making theory and practice: an overview of user interests in knowledge seeking and use. *J Knowl Manage* 1998 Dec;2(2):36-46. [doi: [10.1108/13673279810249369](https://doi.org/10.1108/13673279810249369)]
41. Campbell-Enns H, Woodgate R. The psychosocial experiences of women with breast cancer across the lifespan: a systematic review. *Psychooncology* 2017 Nov;26(11):1711-1721. [doi: [10.1002/pon.4281](https://doi.org/10.1002/pon.4281)] [Medline: [27648726](https://pubmed.ncbi.nlm.nih.gov/27648726/)]
42. Turner J, Kelly B. Emotional dimensions of chronic disease. *West J Med* 2000 Feb;172(2):124-128 [FREE Full text] [Medline: [10693376](https://pubmed.ncbi.nlm.nih.gov/10693376/)]
43. Paul L. Précis of transformative experience. *Philos Phenomenol Res* 2015 Oct 26;91(3):760-765. [doi: [10.1111/phpr.12249](https://doi.org/10.1111/phpr.12249)]
44. Cipolletta S, Votadoro R, Faccio E. Online support for transgender people: an analysis of forums and social networks. *Health Soc Care Commun* 2017 Sep;25(5):1542-1551. [doi: [10.1111/hsc.12448](https://doi.org/10.1111/hsc.12448)] [Medline: [28329916](https://pubmed.ncbi.nlm.nih.gov/28329916/)]
45. Panesar A. The Guardian. 2016 Jun 20. Online communities can help the NHS so why does it ignore them? URL: <https://www.theguardian.com/healthcare-network/2016/jan/20/online-communities-help-nhs> [accessed 2018-06-06] [WebCite Cache ID 6zyTBGFAC]
46. Lupton D. How does health feel? Towards research on the affective atmospheres of digital health. *Digit Health* 2017 Apr 10;3:1-11. [doi: [10.1177/2055207617701276](https://doi.org/10.1177/2055207617701276)]

47. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. *J Am Med Inform Assoc* 2011 May 01;18(3):298-304 [FREE Full text] [doi: [10.1136/amiainl-2010-000058](https://doi.org/10.1136/amiainl-2010-000058)] [Medline: [21486888](https://pubmed.ncbi.nlm.nih.gov/21486888/)]
48. Betts D, Dahlen H, Smith C. A search for hope and understanding: an analysis of threatened miscarriage internet forums. *Midwifery* 2014 Jun;30(6):650-656. [doi: [10.1016/j.midw.2013.12.011](https://doi.org/10.1016/j.midw.2013.12.011)] [Medline: [24439850](https://pubmed.ncbi.nlm.nih.gov/24439850/)]
49. Høybye MT, Johansen C, Tjørnhøj-Thomsen T. Online interaction. Effects of storytelling in an internet breast cancer support group. *Psychooncology* 2005 Mar;14(3):211-220. [doi: [10.1002/pon.837](https://doi.org/10.1002/pon.837)] [Medline: [15386774](https://pubmed.ncbi.nlm.nih.gov/15386774/)]
50. Eschler J, Pratt W. "I'm so glad I met you": designing dynamic collaborative support for young adult cancer survivors. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. New York: ACM Press; 2017 Presented at: Computer-Supported Cooperative Work and Social Computing; Feb 25-Mar 1, 2017; Portland, OR p. 1763-1774 URL: <https://dl.acm.org/citation.cfm?doid=2998181.2998326> [doi: [10.1145/2998181.2998326](https://doi.org/10.1145/2998181.2998326)]
51. Sandaunet A. The challenge of fitting in: non-participation and withdrawal from an online self-help group for breast cancer patients. *Sociol Health Illn* 2008 Jan;30(1):131-144 [FREE Full text] [doi: [10.1111/j.1467-9566.2007.01041.x](https://doi.org/10.1111/j.1467-9566.2007.01041.x)] [Medline: [18254837](https://pubmed.ncbi.nlm.nih.gov/18254837/)]
52. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. *J Med Internet Res* 2017 Apr 24;19(4):e130 [FREE Full text] [doi: [10.2196/jmir.6834](https://doi.org/10.2196/jmir.6834)] [Medline: [28438725](https://pubmed.ncbi.nlm.nih.gov/28438725/)]
53. Zhang S, O'Carroll BE, Owen J, Bakken S, Elhadad N. Online cancer communities as informatics intervention for social support: conceptualization, characterization, and impact. *J Am Med Inform Assoc* 2017 Dec 01;24(2):451-459 [FREE Full text] [doi: [10.1093/jamia/ocw093](https://doi.org/10.1093/jamia/ocw093)] [Medline: [27402140](https://pubmed.ncbi.nlm.nih.gov/27402140/)]
54. Corbin J, Strauss A. Managing chronic illness at home: three lines of work. *Qual Sociol* 1985;8(3):224-247. [doi: [10.1007/BF00989485](https://doi.org/10.1007/BF00989485)]
55. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. *J Med Internet Res* 2008 May 27;10(3):e15 [FREE Full text] [doi: [10.2196/jmir.1053](https://doi.org/10.2196/jmir.1053)] [Medline: [18504244](https://pubmed.ncbi.nlm.nih.gov/18504244/)]
56. 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](https://doi.org/10.1089/cyber.2013.1512)] [Medline: [23659724](https://pubmed.ncbi.nlm.nih.gov/23659724/)]
57. Gilbert E. Predicting tie strength in a new medium. In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. New York: ACM Press; 2012 Presented at: Computer Supported Cooperative Work; Feb 11-15, 2012; Seattle, WA p. 147-156 URL: <https://dl.acm.org/citation.cfm?id=2145360> [doi: [10.1145/2145204.2145360](https://doi.org/10.1145/2145204.2145360)]
58. Partridge SR, Gallagher P, Freeman B, Gallagher R. Facebook groups for the management of chronic diseases. *J Med Internet Res* 2018 Jan 17;20(1):e21 [FREE Full text] [doi: [10.2196/jmir.7558](https://doi.org/10.2196/jmir.7558)] [Medline: [29343460](https://pubmed.ncbi.nlm.nih.gov/29343460/)]
59. Pendry L, Salvatore J. Individual and social benefits of online discussion forums. *Comput Hum Behav* 2015 Sep;50:211-220. [doi: [10.1016/j.chb.2015.03.067](https://doi.org/10.1016/j.chb.2015.03.067)]
60. Rubenstein EL. "Things my doctor never told me": bridging information gaps in an online community. In: *Proc Am Soc Info Sci Tech*. 2013 Jan 24 Presented at: ASIS&T 75th Annual Meeting; Oct 26-30, 2012; Baltimore, MA p. 1-10. [doi: [10.1002/meet.14504901126](https://doi.org/10.1002/meet.14504901126)]
61. Chen AT. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. *Patient Educ Couns* 2012 May;87(2):250-257. [doi: [10.1016/j.pec.2011.08.017](https://doi.org/10.1016/j.pec.2011.08.017)] [Medline: [21930359](https://pubmed.ncbi.nlm.nih.gov/21930359/)]
62. Gefen D, Ridings C. If you spoke as she does, sir, instead of the way you do. *SIGMIS Database* 2005 Jun 07;36(2):78-92. [doi: [10.1145/1066149.1066156](https://doi.org/10.1145/1066149.1066156)]
63. Stewart Loane S, Webster C, D'Alessandro S. Identifying consumer value co-created through social support within online health communities. *J Macromarketing* 2014 May 29;35(3):353-367. [doi: [10.1177/0276146714538055](https://doi.org/10.1177/0276146714538055)]
64. Chen AT. What's in a virtual hug? A transdisciplinary review of methods in online health discussion forum research. *Library Info Sci Res* 2014 Apr;36(2):120-130. [doi: [10.1016/j.lisr.2013.10.007](https://doi.org/10.1016/j.lisr.2013.10.007)]
65. De Simoni S, Shanks A, Balasooriya-Smeekens C, Mant J. Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative study of a sample of participants. *BMJ Open* 2016 Apr 06;6(4):e010501 [FREE Full text] [doi: [10.1136/bmjopen-2015-010501](https://doi.org/10.1136/bmjopen-2015-010501)] [Medline: [27053271](https://pubmed.ncbi.nlm.nih.gov/27053271/)]
66. McGarry S, Ward C, Garrod R, Marsden J. An exploratory study into the unmet supportive needs of breast cancer patients. *Eur J Cancer Care (Engl)* 2013 Sep;22(5):673-683. [doi: [10.1111/ecc.12076](https://doi.org/10.1111/ecc.12076)] [Medline: [23738768](https://pubmed.ncbi.nlm.nih.gov/23738768/)]
67. Carmel-Gilfilen C, Portillo M. Designing with empathy: humanizing narratives for inspired healthcare experiences. *Herd-Health Env Res* 2016;9(2):130-146. [doi: [10.1177/1937586715592633](https://doi.org/10.1177/1937586715592633)] [Medline: [26169208](https://pubmed.ncbi.nlm.nih.gov/26169208/)]
68. Bender J, Jimenez-Marroquin M, Jadad A. Seeking support on Facebook: a content analysis of breast cancer groups. *J Med Internet Res* 2011 Feb 04;13(1):e16 [FREE Full text] [doi: [10.2196/jmir.1560](https://doi.org/10.2196/jmir.1560)] [Medline: [21371990](https://pubmed.ncbi.nlm.nih.gov/21371990/)]
69. 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 Apr 23;15(4):e85 [FREE Full text] [doi: [10.2196/jmir.1933](https://doi.org/10.2196/jmir.1933)] [Medline: [23615206](https://pubmed.ncbi.nlm.nih.gov/23615206/)]

Abbreviations

MND: motor neuron disease

Edited by G Eysenbach; submitted 21.12.17; peer-reviewed by C Hoving, S Atanasova, G Petrič, X Wang; comments to author 23.01.18; revised version received 17.04.18; accepted 08.05.18; published 14.06.18.

Please cite as:

Hargreaves S, Bath PA, Duffin S, Ellis J

Sharing and Empathy in Digital Spaces: *Qualitative Study of Online Health Forums for Breast Cancer and Motor Neuron Disease (Amyotrophic Lateral Sclerosis)*

J Med Internet Res 2018;20(6):e222

URL: <http://www.jmir.org/2018/6/e222/>

doi: [10.2196/jmir.9709](https://doi.org/10.2196/jmir.9709)

PMID: [29903695](https://pubmed.ncbi.nlm.nih.gov/29903695/)

©Sarah Hargreaves, Peter A Bath, Suzanne Duffin, Julie Ellis. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.06.2018. 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.

Original Paper

Detecting Suicidal Ideation on Forums: Proof-of-Concept Study

Ahmet Emre Aladağ^{1,2}, MSc; Serra Muderrisoglu³, PhD; Naz Berfu Akbas⁴, MD; Oguzhan Zahmacioglu⁵, MD; Haluk O Bingol¹, PhD

¹Department of Computer Engineering, Bogazici University, Istanbul, Turkey

²Amazon Research, Madrid, Spain

³Department of Psychology, Bogazici University, Istanbul, Turkey

⁴Medical School, Department of Psychiatry, Yeditepe University, Istanbul, Turkey

⁵Medical School, Department of Child and Adolescent Psychiatry, Yeditepe University, Istanbul, Turkey

Corresponding Author:

Ahmet Emre Aladağ, MSc

Department of Computer Engineering

Bogazici University

Bebek

Istanbul, 34342

Turkey

Phone: 90 2123594523

Fax: 90 2122872461

Email: emre.aladag@boun.edu.tr

Abstract

Background: In 2016, 44,965 people in the United States died by suicide. It is common to see people with suicidal ideation seek help or leave suicide notes on social media before attempting suicide. Many prefer to express their feelings with longer passages on forums such as Reddit and blogs. Because these expressive posts follow regular language patterns, potential suicide attempts can be prevented by detecting suicidal posts as they are written.

Objective: This study aims to build a classifier that differentiates suicidal and nonsuicidal forum posts via text mining methods applied on post titles and bodies.

Methods: A total of 508,398 Reddit posts longer than 100 characters and posted between 2008 and 2016 on SuicideWatch, Depression, Anxiety, and ShowerThoughts subreddits were downloaded from the publicly available Reddit dataset. Of these, 10,785 posts were randomly selected and 785 were manually annotated as suicidal or nonsuicidal. Features were extracted using term frequency-inverse document frequency, linguistic inquiry and word count, and sentiment analysis on post titles and bodies. Logistic regression, random forest, and support vector machine (SVM) classification algorithms were applied on resulting corpus and prediction performance is evaluated.

Results: The logistic regression and SVM classifiers correctly identified suicidality of posts with 80% to 92% accuracy and F1 score, respectively, depending on different data compositions closely followed by random forest, compared to baseline ZeroR algorithm achieving 50% accuracy and 66% F1 score.

Conclusions: This study demonstrated that it is possible to detect people with suicidal ideation on online forums with high accuracy. The logistic regression classifier in this study can potentially be embedded on blogs and forums to make the decision to offer real-time online counseling in case a suicidal post is being written.

(*J Med Internet Res* 2018;20(6):e215) doi:[10.2196/jmir.9840](https://doi.org/10.2196/jmir.9840)

KEYWORDS

suicide; suicidal ideation; suicidality; detection; prevention; classification model; text mining; machine learning; artificial intelligence; suicidal surveillance

Introduction

Background

Suicide was the tenth leading cause of death for all ages in 2016 with 44,965 completed suicides in the United States, corresponding to 123 deaths by suicide per day [1]. According to estimations, approximately 2.7 million US adults (1.1% of the population) made a suicide plan in 2014. Among these, 1.3 million made an actual suicide attempt [2,3]. It was shown that suicide plans were more commonly (2.5%) observed among adults aged between 18 and 25 years [2]. Considering the high social media penetration rates of this group, many suicide attempts could possibly be prevented via social media surveillance. Suicide does not only affect the victim, but also their family, friends, and even society. The economic cost of death due to suicide in the United States in 2013 was estimated to be US \$58.4 billion [4]. To prevent suicides, monitoring is significantly important. Risk factors of suicide include previous suicide attempt(s), history of depression or other mental illness, alcohol or drug abuse, family history of suicide or violence, physical illness, and feeling alone [2]. There is no agreement on the definition of the suicide stages yet [5]. However, suicidal people were commonly divided into two main classes: ideators (ideators, planners) and attempters (attempters, completers). Considering 80% of the patients attempting suicide were found to score in the depressed range [6], we consider depression a potential precursor to suicidal ideation: depression, suicidal ideation, plan, attempt, and completion.

Suicidal Surveillance and Suicide Prevention

While monitoring for suicide prevention, it is prudent to “cast a wider net,” meaning it is fine to see a few false positive overhead, up to an acceptable level, for the sake of not missing suicidal people. It has been shown that half of all suicides are likely to occur in lower-risk groups [7]. In this sense, sensitivity/recall (the ability to accurately detect suicidal behavior) is more important than precision (the ability to accurately discard nonsuicidal behavior) and accuracy (predicting suicidal status correctly) [8]. However, having 100% recall and very low precision (which is unacceptable) would be a waste of resources and potentially annoying for the patient. Thus, a balance between the two should be sought.

Any suicidal sign should be taken seriously and the patient should be questioned on existence of suicidal thoughts. In common clinical practice, patients with depressive symptoms are asked whether they have any suicidal thoughts for early diagnosis. To perform a better risk assessment, suicide prevention researchers analyze patient history, statements, and suicide notes. It is known that more than 50% of suicide completers and 20% of suicide attempters left suicide notes [9]. Thus, leaving suicide notes should be considered a significant signal [10-13].

Suicidal Ideation on Social Media

In the past, suicidal individuals could only write suicide notes to express their feelings and some studies focused on linguistic and content analysis of these notes [14]. However, with the introduction of social media, such as Facebook, Twitter, Reddit,

and Tumblr, Internet users are now sharing their suicidal feelings and thoughts on these platforms [15,16]. Because postings occur in the earlier phases of ideation, these people can potentially be saved if proper support is given. For each post on such platforms, the following questions can be asked:

1. Does the author of this post have suicidal ideation?
2. Does this person have potential to attempt suicide?
3. Is this post a suicide note?
4. Is this post authentic?
5. Has this person already committed suicide?

Answering each question is a different problem on its own. Because every individual with suicidal or depressive expressions should be provided support, answering the first question (which is our objective) is more beneficial in suicide prevention. For this purpose, detecting suicidal and potentially suicidal people via surveillance is important.

Text Mining Methods for Suicidal Ideation Detection

In efforts of suicidal surveillance, applying text processing and supervised machine learning (classification) techniques for performing suicidal text detection is becoming more popular in suicide research. In this approach, textual features are extracted from posts for discriminating suicidality of a text. Then statistical classification algorithms, such as logistic regression, random forest, and support vector machine (SVM) algorithms are applied to discover patterns (relationships between the features and the suicidality status). Finally, models resulting from training with these classifiers are evaluated with test data and evaluation metrics. Accuracy, precision, recall (sensitivity), and F_1 score are the commonly used metrics for evaluation of the classifier performance. In suicidal post classification context, accuracy represents the fraction of posts classified correctly over all posts. Recall represents the fraction of suicidal posts that are correctly classified as suicidal. Precision represents the fraction of posts that are actually suicidal among the posts classified as suicidal. The F_1 score is the harmonic mean of precision and recall, leading to a more balanced evaluation because precision and recall are complementary metrics.

The prediction performance heavily relies on extracting the best features. Several techniques are used for extracting significant features (feature extraction). Some are the bag-of-words model [17], term frequency-inverse document frequency (tf-idf) [18], linguistic inquiry and word count (LIWC) [19], and sentiment analysis. These techniques provide an analysis of words, themes, or tones commonly used in suicidal posts. See [Multimedia Appendix 1](#) for a detailed explanation of these techniques.

Related Work

There have been efforts on differentiating suicide note content from regular content [20-22]. With the rapid rise of social media, recent studies have begun to utilize text mining on online posts with depression [22], suicidal ideation [23-30], and mental health disorders [30]. These studies have shown the potential of using online posts to assess suicide risk or depression in English (Twitter) and Chinese (Weibo). However, character limitations make the prediction on these microblogging platforms error-prone because thoughts are spanned over multiple posts, making it harder to grasp the context if posts are

evaluated independently during the machine learning process. On the other hand, evaluating all posts of a user results in a dilution of suicidality [25] because suicidality is not expected to be expressed in all posts. Therefore, there is a need to analyze all consecutive posts by the author and define boundaries of a suicidal set of posts. Moreover, recent studies on microblogging platforms used limited number of short posts and few of them provided sufficient performance. After further improvements, classifiers in the study by Guan et al [25] can be used for passive surveillance on Weibo to track users with suicidal mood spanning over a long time period, whereas classifiers in the study by O'Dea et al [24] can be used to detect impulsive suicidal expressions on Twitter. However, it was seen that strongly concerning suicide-related tweets and Weibo posts had higher word count [27,28]. This indicates that seriously suicidal individuals may need longer space to express themselves, such as blogs, forums, or Facebook posts. Thus, classifiers resulting from these studies may fall short in detecting serious and thoughtful suicidal ideators. Furthermore, longer posts have higher chance of being identified correctly due to their longer content. At this point, using Reddit as the dataset and aiming to predict suicidality on longer-form posts may be more effective.

Goal of This Study

This study aims to build a classifier that detects long passages like forum and blog posts containing suicidal ideation via text mining methods to assist authorities in preventing potential suicide attempts.

Methods

Classification models were developed to predict whether a given post with title and body text contains suicidal ideation using a dataset consisting of Reddit posts. The performance objective was to achieve a prediction performance that would mimic a human expert.

Data Collection

In this study, a dataset containing publicly available Reddit posts was used. Reddit was chosen as the data source because it allows longer posts and has a special section on suicidal ideation. However, the generated models can be applied on blogs or any other social media platform, especially ones allowing long posts. No personally identifiable information apart from usernames (in many cases not revealing the real identity) and explicitly stated information were provided with the data. Nevertheless, usernames were not downloaded from the data source during this study and ethics committee approval was not sought. As a consequence, posts from the same authors were handled as separate posts. Because we did not have concerns about differentiating the author-post relationship, this limitation did not pose a problem.

Using Google Cloud BigQuery, posts with a text body of at least 100 characters and that were posted on the subreddits

SuicideWatch, Depression, Anxiety, and ShowerThoughts between September 2008 and October 2016 inclusive (508,398 posts) were downloaded. Each post had an ID, title, body, and subreddit name. SuicideWatch is a subreddit where thousands of people write about their suicidal ideations. The majority of authors on this subreddit are depressed and thinking about suicide. They share their feelings and some ask for help. It is unknown whether any of these people killed themselves unless they left comments stating they changed their mind or published new posts afterwards. Nevertheless, the contents of these posts can definitely be seen as signs of suicidal ideation. Posts on the subreddits Depression and Anxiety contain depressive and anxious thoughts, respectively. A minority of posts on these two subreddits may contain suicidal thoughts as well because suicidal people may have anxiety and depressive feelings, which may lead them to write in these places. Posts on ShowerThoughts, on the other hand, contain authentic personal thoughts that came to mind in the shower on any topic. Therefore, ShowerThoughts is a good candidate for comparison against the aforementioned thought-oriented subreddits and not many suicidal posts are expected on this subreddit.

Data Annotation

Among the posts collected, random posts were selected from all subreddits and manually annotated. In total, 785 posts on SuicideWatch (n=175), Depression (n=200), Anxiety (n=200), and ShowerThoughts (n=210) were manually annotated. Because the dataset would be balanced by binary annotation using oversampling, a balance in the number of annotated posts among subreddits was not sought. A post was labeled as suicidal (1) if the author of the post clearly seemed to have suicidal thoughts; otherwise, it was labeled nonsuicidal (0). One exception to this rule was posts on the SuicideWatch subreddit with strong depression and anticipated suicide risk. These posts were annotated as suicidal even if they did not have a suicidal language because posting on that subreddit is an implicit sign of suicidality. Posts on SuicideWatch were annotated by psychiatrists (NBA and OZ) with an initial agreement rate of 93% and a Cohen kappa [31] coefficient of $\kappa=.74$. The conflicts were then resolved by these authors reaching a consensus. It was seen that 150 of 175 (85.7%) posts were actually posts of people with suicidal ideation; 25 of 175 (14.3%) posts were not. The nonsuicidal segment contained (1) posts asking what to do for a suicidal friend, (2) posts of people who had a suicidal ideation in mind but who were not willing to die anymore, (3) posts asking questions about suicidal people, and (4) a few posts unrelated to the topic. These 25 posts seemed suicidal at first by solely looking at their choice of words; however, they were not suicidal although having been about suicidality. There were posts in a similar situation in other subreddits as well. Although these posts may cause noise when used as a test set, we included them in our dataset for better generalization. Posts in other subreddits were annotated by the AEA under consultancy and guidance of NBA and OZ. See Table 1 for distribution of the suicidality label among annotated posts.

Table 1. Suicidality label distribution of posts in subreddits.

Subreddit	Nonsuicidal, n	Suicidal, n	Total, n
SuicideWatch	25	150	175
Depression	152	48	200
Anxiety	193	7	200
ShowerThoughts	210	0	210
Total	580	205	785

Dataset Formation

Four experiments were carried out with different samplings from four subreddits. For each experiment (E_i), a custom dataset (D_i) and a corresponding label vector (L_i) indicating binary suicidality status was generated. The custom dataset contained post information from selected subreddits (or annotated post set) with rows corresponding to posts, columns corresponding to ID, subreddit, title, and body fields for posts. The label vector $L_i=[l_1, l_2, \dots, l_{mi}]^T$ was a vector where l_j was the binary label for the corresponding post (p_j). The label value was set $l_j=1$ if the corresponding post p_j was annotated as suicidal, $l_j=0$ otherwise (see [Table 2](#)).

Feature Extraction

First, two features were extracted for all posts: LIWC matrices (W_t and W_b) for title and body, and sentiment matrices (S_t and S_b) for title and body. These were the constant features that did not change by composition of posts in datasets. Then, specifically for each dataset D_i to be used in E_i , two more features were extracted: document term matrix for title (T_{it}) and document term matrix for post body (T_{ib}). See [Multimedia Appendix 2](#) for a diagram of feature extraction and the experiment design steps.

Linguistic Inquiry and Word Count Matrix

Initially, LIWC 2015 tool [19] was run on all 508,398 posts (on titles and bodies separately), producing two LIWC matrices (W_t and W_b) where rows corresponded to posts and columns (of size 93) corresponded to LIWC features. Each cell contained a calculated feature score for a post. Feature scaling (standard normalization) was applied on these scores to have all the features in the same range. Then, for each experiment E_i , subsamples of the resulting matrix were extracted for each

dataset to contain only rows that also existed in D_i , resulting in W_{it} and W_{ib} .

Sentiment Matrices

To build sentiment score matrices— S_{it} (for title) and S_{ib} (for body) for D_i dataset—Python TextBlob library [32] (which uses Python Natural Language Toolkit [NLTK] library [33] internally) was incorporated. This process yielded two augmented matrices $S_{it}=[S_{itp}|S_{itj}]$ and $S_{ib}=[S_{ibp}|S_{ibj}]$ each with two columns: polarity (P) and subjectivity (J) in the range $[-1,1]$.

Document Term Matrices

To build T_{it} and T_{ib} matrices, title and body fields in D_i were used. For each row of D_i , text in title/body field was converted to lowercase and applied the Porter stemming algorithm [34] with the NLTK library [33] to obtain the word stems. This allowed words to be evaluated in their canonical forms. Words of stem “suicide” were ignored in all subreddits to avoid classifying solely by existence of the word “suicide.” Then tf-idf document term matrices T_{it} and T_{ib} were built using Python scikit-learn library [35]. Having a large vocabulary (number of columns), one-way analysis of variance (ANOVA) F-test [36] was applied to the matrices to reduce the number of features to 200 for each of the two matrices, leaving the most important columns. This reduced the time required to train models with the classification algorithms.

Combining Features

At the end, these features were concatenated, resulting in corpus $C_i=[W_{it}|W_{ib}|S_{it}|S_{ib}|T_{it}|T_{ib}]$ with 590 columns: 93, 93, 2, 2, 200, and 200, in respective order, for each dataset D_i . These corpora were combined with corresponding label vectors previously tied to D_i , forming an augmented matrix $[C_i|L_i]$ (see [Table 3](#)).

Table 2. Hypothetical dataset (D_i) matrix and corresponding label vector (L_i) for an experiment with two sample posts. A table in this form was generated for each experiment with different posts.

D_i				L_i
Post ID	Subreddit	Title	Body	Label
1	SuicideWatch	I don't wanna live anymore	Since the day I was born,...	1
2	ShowerThoughts	Why are the oceans blue?	I have always wondered...	0

Table 3. Sample table representing concatenated $C_i | L_i$ matrix containing 590 corpus feature columns (C_i) plus one label column (L_i) that were provided to machine learning algorithms for classification. A matrix in this form was generated for each experiment with different posts.

Post ID ^a	$W_{it1} \dots W_{it93}$ ^b	$W_{ib1} \dots W_{ib93}$ ^c	S_{itp} ^d	S_{itj}	S_{ibp} ^e	S_{ibj}	$T_{it1} \dots T_{it200}$ ^f	$T_{ib1} \dots T_{ib200}$ ^g	L_i
1	0.3...0.00	0.15...0.22	-0.75	0.70	0.25	0.35	0.15...0.54	0.14...0.32	1
2	0.11...0.08	0.00...0.00	0.20	0.90	-0.45	0.78	0.07...0.93	0.01...0.63	0
Column #	1...93	94...186	187	188	189	190	191	391...590	1

^aPost IDs are hypothetical.^b W_{it} : Linguistic inquiry and word count (LIWC) matrice for title.^c W_{ib} : LIWC matrix for body.^d S_{it} : sentiment score matrix for title.^e S_{ib} : sentiment score matrix for body.^f T_{it} : document term matrix for title.^g T_{ib} : document term matrix for body.

Experiment Design

Each subreddit contained posts with different levels of suicidality. SuicideWatch mostly contained suicidal posts, Depression contained highly depressive and partly suicidal posts, and Anxiety contained some suicidal but mostly nonsuicidal posts. ShowerThoughts contained mostly nonsuicidal posts. Four experiments were conducted with different compositions of posts to see if discrimination for different levels of suicidality was possible. A new data table was generated for each experiment (see Table 4).

Experiment 1 was designed to see if it is possible to differentiate suicidal posts from posts talking about random daily matters. For this purpose, 175 annotated posts from SuicideWatch and 210 annotated posts from ShowerThoughts subreddits were selected because they are on different sides of the suicidality scale and provide good samples for contrast. To avoid a potential overfit, the experiment was evaluated with 10-fold cross-validation. This experiment was expected to yield good results because the two subreddits were expected to have mostly different vocabulary.

Experiment 2 was designed to see if it is possible to differentiate suicidality when posts with anxious/depressive vocabulary are involved. For this purpose, 200 Anxiety subreddit and 200 Depression subreddit posts, which can be seen as some of the closest psychological moods to suicidality, were included in addition to the composition of experiment 1, forming the second experiment. Because the vocabulary use of depressive, anxious, and suicidal people are expected to have commonalities and posts with these moods are harder to classify, a performance loss was expected in this experiment when compared to experiment 1. However, the diversity of the posts made the models in this experiment a finer-grain predictor in real-life applications.

Experiment 3 was designed to see if it is safe to assume all posts in SuicideWatch are suicidal ($l_j=1$) and all posts in ShowerThoughts are nonsuicidal ($l_j=0$) when training a model. For this purpose, models were trained with randomly selected, nonannotated 5000 SuicideWatch and 5000 ShowerThoughts

posts under this assumption. The trained models were then tested against 175 SuicideWatch and 210 ShowerThought posts, which were already annotated to see if the model trained under the aforementioned assumption could perform well against the gold standard. Because the majority of SuicideWatch posts tend to be suicidal and the majority of ShowerThoughts posts tend to be nonsuicidal, only a slight performance loss was expected when compared to experiment 1.

Experiment 4 was designed to battle-test our model trained with the assumptions in experiment 3 against all 785 annotated posts including the depressive and anxious posts, which are difficult to make judgment on. Because the model was not trained with difficult cases, it was inevitable for it to fail in such cases. However, the models were still expected to perform better than the baseline model trained with ZeroR algorithm.

Model Training and Evaluation

In experiments 1 and 2, rows (posts) from selected subreddits were appended, resulting in datasets D_1 and D_2 . The datasets were then applied feature extraction steps to result in corpus C_1 and C_2 . 10-fold cross-validation was applied on C_1 and C_2 and their corresponding label vectors L_1 and L_2 . In each split, random synthetic minority oversampling technique (SMOTE) [37] was applied before training to obtain an equal number of posts from both classes and avoid imbalanced data bias. Models for each fold were trained with ZeroR (set to always classify posts as suicidal), logistic regression ($\text{delta}=1.0$) and random forest (with 10 trees) and SVM (with radial basis function kernel) classification implementations in Python scikit-learn library. To reduce fluctuations in scores (due to randomization and limited number of samples), experiments were repeated 100 times. Average metric scores were then evaluated. Logistic regression was chosen due to its efficiency, interpretable nature, ability to provide probabilities, and online learning (ability to update model parameters after being exposed to new labeled data) support. Random forest and SVM were chosen due to their high classification performance, especially for datasets with high number of instances and features. ZeroR was chosen as the baseline classifier for comparison.

Table 4. Summary of post distribution used in experiments (E).

Subreddit	Whole data (10-fold) posts, n		Train data posts, n		Test data posts, n	
	E1	E2	E3	E4	E3	E4
SuicideWatch	175 ^a	175 ^a	5000	5000	175 ^a	175 ^a
ShowerThoughts	210 ^a	210 ^a	5000	5000	210 ^a	210 ^a
Depression		200 ^a				200 ^a
Anxiety		200 ^a				200 ^a

^aAnnotated post.

In experiments 3 and 4, rows of train and test data were initially appended and followed the feature extraction steps. After building C_3 and C_4 , rows were split (preserving the train/test formation) to form train and test corpuses $C_{\text{train-}i}$ and $C_{\text{test-}i}$ together with corresponding label vectors $L_{\text{train-}i}$ and $L_{\text{test-}i}$ for $i=3,4$. After training models on $C_{\text{train-}i}$ and $L_{\text{train-}i}$ with the aforementioned algorithms, oversampled test corpus $C_{\text{test-}i}$ and $L_{\text{test-}i}$ were used to test the trained models.

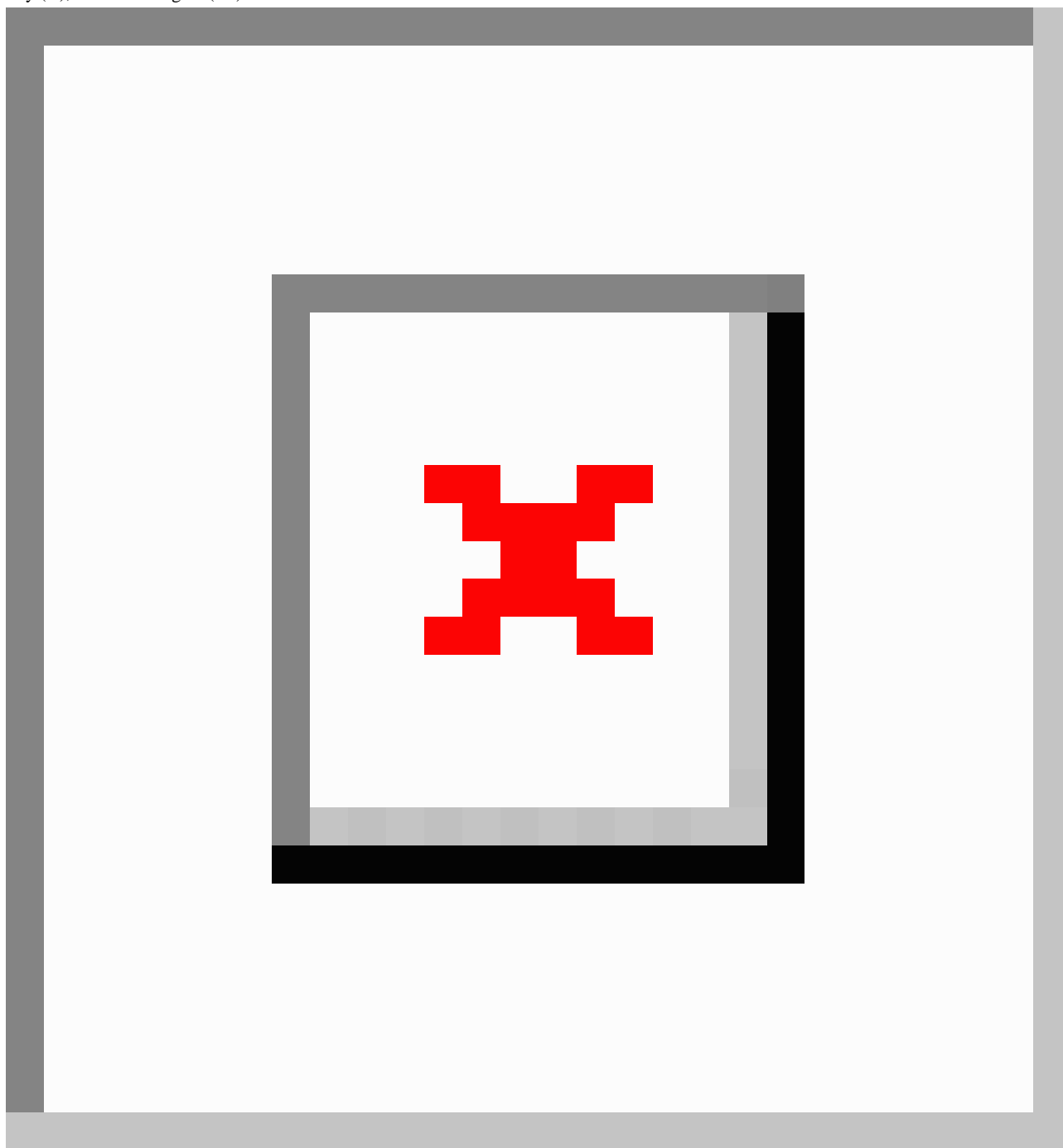
Results

In all the experiments, logistic regression and SVM (except for experiment 2) performed the best, followed by random forest, all much more performant than ZeroR, which provided a baseline (66% for F_1 , 50% for other metrics) for performance evaluation of the classification task (see Figure 1). Although

SVM performance slightly exceeded logistic regression in experiments 3 and 4, logistic regression would be favorable due to its simplicity.

In experiment 1, the logistic regression and SVM classifiers could differentiate suicidal posts from nonsuicidal posts with an F_1 score of 92%, followed by random forest (89%). We can attribute this performance to the suicidality levels of SuicideWatch and ShowerThoughts being on different sides of the scale. The experiment yielded good results as expected. Furthermore, if posts in SuicideWatch mentioning suicidal people other than the author were annotated as suicidal, those scores would be even higher. When LIWC and sentiment features were removed from the feature set, F_1 score went down from 92% to 88%, which is not a significant decrease. This shows that results are still promising even when solely tf-idf matrices were used.

Figure 1. Prediction performance evaluation for the four experiments with different combinations of posts from SuicideWatch (SW), Depression (D), Anxiety (A), ShowerThoughts (ST) subreddits.



When the Depression and Anxiety subreddits were introduced in the experiment 2 dataset, the performance was 11% lower for logistic regression (F_1 /accuracy=81%), followed by random forest (F_1 /accuracy=80%) and SVM (F_1 =73%, accuracy=76%), but still sufficiently high when compared to the baseline classifier (F_1 =66%, accuracy=50%). This also complies with our expectations because the algorithm had to deal with edge cases where depressive and suicidal people have a large intersection in vocabulary. With more labeled data, a better generalization can be achieved.

In experiment 3, where the models were trained with nonannotated posts and tested against annotated posts, F_1 and

accuracy scores were 89% for logistic regression, 3% less than experiment 1. This was an expected reduction because some of the posts in SuicideWatch had suicidal context but were not words of a suicidal person. Nevertheless, lack of annotated posts in the training set was compensated with a higher number of posts from both subreddits, yielding a better generalization. On the other hand, SVM could surpass logistic regression with F_1 =92%.

In experiment 4, lack of posts for training edge cases (annotated posts in Depression and Anxiety) caused a lower but still acceptable prediction performance (an F_1 score of 78% against 66% and an accuracy of 77% against 50%) than in experiment 3. This was an expected result and can be improved by feeding

annotated posts from Depression and Anxiety into the training set.

Discussion

Principal Results

The high (100%) recall performance of the baseline classifier ZeroR is due to its strategy to predict posts as suicidal all the time. This strategy comes with a penalty of false positive overhead, thus 0% precision. This means providing psychological support to everyone, regardless of the content of their posts, which is practically useless, potentially harmful, and costly. The high recall rate—which should be ignored during evaluation for ZeroR—supports a higher than expected illusional F_1 score (66%) in all cases. Although F_1 score is the widely used metric due to its balancing nature, other parameters including accuracy (which is valuable when the dataset is balanced) and precision (to observe the overhead) should be used for comparison with the baseline classifier.

In all the experiments, logistic regression and SVM performed much better than ZeroR baseline algorithm. In the first experiment, the classifiers could predict suicidality with an F_1 score of 92%. There was a reduction in performance when annotated Depression and Anxiety subreddit posts were introduced in experiment 2. This is due to posts in these two subreddits having gradient levels of suicidality, introducing new edge cases the algorithm should handle, making it harder to differentiate nonsuicidal depressives from suicidal people. Solely looking at the vocabulary and psychological meanings of the words seems to have confused the classifier, suggesting a more contextual approach, more data, and maybe a deeper classifier might be required to obtain performance levels in the first experiment.

The high performance in the third experiment shows the assumptions of (1) posts in SuicideWatch subreddit being suicidal and (2) posts in ShowerThoughts being nonsuicidal was valid. It can be said that these assumptions hold when discriminating suicidal posts from nonsuicidal posts in the absence of edge cases such as in Depression and Anxiety posts. When the aforementioned edge cases are introduced in the test set but not in the training set of experiment 4, the performance was lower (which was expected), although still significantly higher than ZeroR algorithm.

Practical Use

Our findings show that text mining methods can be used to detect posts with suicidal ideation online. Being one of the simplest, efficient, and most interpretable models, logistic regression performs very well on the problem. The model trained in this study can be used in spotting people with suicidal ideation while they are writing their forum (or blog) posts right away using a Javascript or mobile app library. Popup dialogs can be shown to authors of posts classified as suicidal. Authors can be asked how they feel and whether they need help or have suicidal thoughts without irritating or leading them to nonexistent suicidal thoughts. Thanks to interpretability and simplicity of logistic regression, the code to embed on mobile apps, blogs, forums, or even Web browsers would add very little overhead.

On admission of having suicidal thoughts on the popup, the author can be offered support immediately via live chat, phone call, or face-to-face counseling. Accepting the offer would be a verification of our prediction and false positives would come mainly from depressed people who have premature suicidal thoughts (who should be supported as well); therefore, lowering precision levels would not pose an overhead on the support staff. This means the model can be tuned further in favor of recall, instead of precision, by changing threshold values of logistic regression. Experiment 3 has validated the assumption that posts on SuicideWatch subreddit can be assumed as suicidal. Thus, a logistic regression model trained as in experiment 3 with the whole SuicideWatch and ShowerThoughts data would be sufficient for a real-world application since the verification system would eliminate the false positives. With the introduction of annotated edge cases (from Depression and Anxiety subreddits) to the training set, the performance can be further improved. Another interesting strategy would be using the responses received from the authors (subject to this prevention system) as annotations to further train the model to make better predictions, leading to an ever-learning online classifier.

Limitations

The prediction system in study is limited to text posts in the English language. Similar models can be trained on other languages given sufficient dataset. Without knowledge of whether the authors of the posts committed suicide, our system can only claim to predict suicidal ideation, not a potential suicide attempt. Although our dataset is limited to Reddit, which is a forum itself, we expect our system to work well on other forums and blogs due to similarity in format and context. However, further research is needed to verify this claim. This study is a proof of concept for online suicidal ideation surveillance, yet further development is needed for a real-time online suicide prevention system after designing appropriate questionnaires to be asked to authors with suicidal markers. In all, 175 SuicideWatch posts were annotated by psychiatrists. Due to the time-consuming nature of annotating hundreds of posts, the rest of the subreddits were annotated by AEA under guidance of the psychiatrists NBA and OZ. To avoid introducing a bias, AEA performed annotation on SuicideWatch posts as well and the similarity between annotations of authors were analyzed. It was seen that AEA and the psychiatrists agreed on annotations 87% and 89% of the time, whereas the psychiatrists agreed 93% of the time among themselves before resolving conflicts. So this indicates annotations of the computer scientist (AEA) are not expected to introduce bias.

Conclusions

To the best of our knowledge, this is the first study that uses forum posts in thousands scale (of which 785 were manually annotated) with an objective of detecting posts with suicidal ideation with high performance in all metrics. Results indicate it is possible to detect suicidal people online to provide them proper immediate support as they are writing. Authors of this text acknowledge that detecting suicidal ideation with high accuracy is a difficult problem even for humans and design of nonintrusive conversation for potential suicidal candidates should be carried out carefully. Application of such a detection

system in real time may save thousands of lives every year if carried out properly.

Acknowledgments

This work was partially supported by TUBITAK BIDEB 2211 and the Turkish Ministry of Development under the TAM Project number DPT2007K120610. This work was done prior to AEA joining Amazon. HOB would like to thank University of Waterloo for his sabbatical year.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Method Descriptions.

[PDF File (Adobe PDF File), 386KB - [jmir_v20i6e215_app1.pdf](#)]

Multimedia Appendix 2

Feature Extraction and Experiment Design Diagram.

[ZIP File (Zip Archive), 632KB - [jmir_v20i6e215_app2.zip](#)]

References

- Centers for Disease Control and Prevention: WISQARS. Fatal injury reports, national, regional and state, 1981 – 2016 URL: <https://webappa.cdc.gov/sasweb/ncipc/mortrate.html> [accessed 2018-05-31] [WebCite Cache ID 6zp3KWDCq]
- Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Suicide: Facts at a Glance. 2015. URL: <https://www.cdc.gov/violenceprevention/pdf/suicide-datasheet-a.pdf> [accessed 2017-09-24] [WebCite Cache ID 6tiRNsbWX]
- Lipari R, Piscopo K, Kroutil LA, Miller GK. NSDUH Data Review. Rockville, MD: SAMHSA; 2015 Sep. Suicidal thoughts and behavior among adults: results from the 2015 National Survey on Drug Use and Health URL: <https://www.samhsa.gov/data/sites/default/files/NSDUH-FRR2-2014/NSDUH-FRR2-2014.pdf> [accessed 2018-05-31]
- Shepard DS, Gurewicz D, Lwin AK, Reed GA, Silverman MM. Suicide and suicidal attempts in the United States: costs and policy implications. *Suicide Life Threat Behav* 2016 Dec;46(3):352-362 [FREE Full text] [doi: [10.1111/sltb.12225](https://doi.org/10.1111/sltb.12225)] [Medline: [26511788](https://pubmed.ncbi.nlm.nih.gov/26511788/)]
- Silverman MM, De Leo D. Why there is a need for an international nomenclature and classification system for suicide. *Crisis* 2016 Mar;37(2):83-87. [doi: [10.1027/0227-5910/a000419](https://doi.org/10.1027/0227-5910/a000419)] [Medline: [27232426](https://pubmed.ncbi.nlm.nih.gov/27232426/)]
- Silver MA, Bohnert M, Beck AT, Marcus D. Relation of depression of attempted suicide and seriousness of intent. *Arch Gen Psychiatry* 1971 Dec;25(6):573-576. [Medline: [5141377](https://pubmed.ncbi.nlm.nih.gov/5141377/)]
- Large M, Kaneson M, Myles N, Myles H, Gunaratne P, Ryan C. Meta-analysis of longitudinal cohort studies of suicide risk assessment among psychiatric patients: heterogeneity in results and lack of improvement over time. *PLoS One* 2016;11(6):e0156322 [FREE Full text] [doi: [10.1371/journal.pone.0156322](https://doi.org/10.1371/journal.pone.0156322)] [Medline: [27285387](https://pubmed.ncbi.nlm.nih.gov/27285387/)]
- Hawgood J, De Leo D. Suicide prediction - a shift in paradigm is needed. *Crisis* 2016 Jul;37(4):251-255. [doi: [10.1027/0227-5910/a000440](https://doi.org/10.1027/0227-5910/a000440)] [Medline: [27809591](https://pubmed.ncbi.nlm.nih.gov/27809591/)]
- DeJong TM, Overholser JC, Stockmeier CA. Apples to oranges? A direct comparison between suicide attempters and suicide completers. *J Affect Disord* 2010 Jul;124(1-2):90-97 [FREE Full text] [doi: [10.1016/j.jad.2009.10.020](https://doi.org/10.1016/j.jad.2009.10.020)] [Medline: [19903573](https://pubmed.ncbi.nlm.nih.gov/19903573/)]
- Barr W, Leitner M, Thomas J. Self-harm or attempted suicide? Do suicide notes help us decide the level of intent in those who survive? *Accid Emerg Nurs* 2007 Jul;15(3):122-127. [doi: [10.1016/j.aen.2007.04.005](https://doi.org/10.1016/j.aen.2007.04.005)] [Medline: [17604173](https://pubmed.ncbi.nlm.nih.gov/17604173/)]
- Callanan VJ, Davis MS. A comparison of suicide note writers with suicides who did not leave notes. *Suicide Life Threat Behav* 2009 Oct;39(5):558-568. [doi: [10.1521/suli.2009.39.5.558](https://doi.org/10.1521/suli.2009.39.5.558)] [Medline: [19929155](https://pubmed.ncbi.nlm.nih.gov/19929155/)]
- Eisenwort B, Berzlanovich A, Willinger U, Eisenwort G, Lindorfer S, Sonneck G. [Suicide notes and their importance to suicide research. The representativeness of suicide note writers]. *Nervenarzt* 2006 Nov;77(11):1355-1356, 1358. [doi: [10.1007/s00115-005-1965-y](https://doi.org/10.1007/s00115-005-1965-y)] [Medline: [16028079](https://pubmed.ncbi.nlm.nih.gov/16028079/)]
- Kuwabara H, Shioiri T, Nishimura A, Abe R, Nushida H, Ueno Y, et al. Differences in characteristics between suicide victims who left notes or not. *J Affect Disord* 2006 Aug;94(1-3):145-149. [doi: [10.1016/j.jad.2006.03.023](https://doi.org/10.1016/j.jad.2006.03.023)] [Medline: [16712947](https://pubmed.ncbi.nlm.nih.gov/16712947/)]
- Foster T. Suicide note themes and suicide prevention. *Int J Psychiatry Med* 2003;33(4):323-331. [doi: [10.2190/T210-E2V5-A5M0-QLJU](https://doi.org/10.2190/T210-E2V5-A5M0-QLJU)] [Medline: [15152783](https://pubmed.ncbi.nlm.nih.gov/15152783/)]

15. Ruder TD, Hatch GM, Ampanozi G, Thali MJ, Fischer N. Suicide announcement on Facebook. *Crisis* 2011;32(5):280-282. [doi: [10.1027/0227-5910/a000086](https://doi.org/10.1027/0227-5910/a000086)] [Medline: [21940257](https://pubmed.ncbi.nlm.nih.gov/21940257/)]
16. Karbeyaz K, Akkaya H, Balci Y, Urazel B. Analysis of suicide notes: an experience in Eskişehir City. *Noro Psikiyatr Ars* 2014 Sep;51(3):275-279 [FREE Full text] [doi: [10.4274/npa.y7000](https://doi.org/10.4274/npa.y7000)] [Medline: [28360638](https://pubmed.ncbi.nlm.nih.gov/28360638/)]
17. Harris ZS. Distributional structure. *WORD* 2015 Dec 04;10(2-3):146-162. [doi: [10.1080/00437956.1954.11659520](https://doi.org/10.1080/00437956.1954.11659520)]
18. Salton G, McGill M. Introduction to Modern Information Retrieval. New York: McGraw-Hill College; Sep 01, 1983.
19. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. Austin, TX: University of Texas at Austin; 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2018-05-31] [WebCite Cache ID 6zp55ObHJ]
20. Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A. Suicide note classification using natural language processing: a content analysis. *Biomed Inform Insights* 2010 Aug 04;2010(3):19-28 [FREE Full text] [Medline: [21643548](https://pubmed.ncbi.nlm.nih.gov/21643548/)]
21. Matykiewicz P, Duch W, Pestian J. Clustering semantic spaces of suicide notes and newsgroups articles. In: Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing. 2009 Presented at: BioNLP '09 Workshop on Current Trends in Biomedical Natural Language Processing; Jun 4-5, 2009; Boulder, CO p. 179-184. [doi: [10.3115/1572364.1572389](https://doi.org/10.3115/1572364.1572389)]
22. White E, Mazlack LJ. Creating a suicide note analysis model using fuzzy cognitive maps. 2011 Presented at: International Conference on Artificial Intelligence (ICAI); Jul 18-21, 2011; Las Vegas, NV URL: <https://world-academy-of-science.org/worldcomp11/ws/conferences/icai11/General%20Information.html>
23. Abboute A, Boudjeriou Y, Entringer G, Bringay S, Poncelet P. Mining Twitter for suicide prevention. In: Natural Language Processing and Information Systems. NLDB 2014. Lecture Notes in Computer Science. 2014 Jun 18 Presented at: NLDB 2014 19th Natural Language Processing and Information Systems; Jun 18-20, 2014; Montpellier, France p. 250-253. [doi: [10.1007/978-3-319-07983-7_36](https://doi.org/10.1007/978-3-319-07983-7_36)]
24. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. *Internet Interv* 2015 May;2(2):183-188. [doi: [10.1016/j.invent.2015.03.005](https://doi.org/10.1016/j.invent.2015.03.005)]
25. Guan L, Hao B, Cheng Q, Yip PS, Zhu T. Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. *JMIR Ment Health* 2015;2(2):e17 [FREE Full text] [doi: [10.2196/mental.4227](https://doi.org/10.2196/mental.4227)] [Medline: [26543921](https://pubmed.ncbi.nlm.nih.gov/26543921/)]
26. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for Twitter data against established measures of suicidality. *JMIR Ment Health* 2016 May 16;3(2):e21 [FREE Full text] [doi: [10.2196/mental.4822](https://doi.org/10.2196/mental.4822)] [Medline: [27185366](https://pubmed.ncbi.nlm.nih.gov/27185366/)]
27. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related Twitter posts. *Crisis* 2017 Sep;38(5):319-329. [doi: [10.1027/0227-5910/a000443](https://doi.org/10.1027/0227-5910/a000443)] [Medline: [28228065](https://pubmed.ncbi.nlm.nih.gov/28228065/)]
28. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing suicide risk and emotional distress in chinese social media: a text mining and machine learning study. *J Med Internet Res* 2017 Jul 10;19(7):e243 [FREE Full text] [doi: [10.2196/jmir.7276](https://doi.org/10.2196/jmir.7276)] [Medline: [28694239](https://pubmed.ncbi.nlm.nih.gov/28694239/)]
29. Leo K. BBC News. 2017. Facebook artificial intelligence spots suicidal users URL: <http://www.bbc.com/news/technology-39126027> [accessed 2017-11-05] [WebCite Cache ID 6ukw9DUW5]
30. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. *Sci Rep* 2017 Mar 22;7:45141 [FREE Full text] [doi: [10.1038/srep45141](https://doi.org/10.1038/srep45141)] [Medline: [28327593](https://pubmed.ncbi.nlm.nih.gov/28327593/)]
31. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol* 2012;8(1):23-34 [FREE Full text] [Medline: [22833776](https://pubmed.ncbi.nlm.nih.gov/22833776/)]
32. Loria S. TextBlob. 2017. URL: <https://github.com/sloria/textblob> [accessed 2017-12-21] [WebCite Cache ID 6vrTwufWI]
33. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Sebastopol, CA: O'Reilly Media; Jul 10, 2009.
34. Porter MF. Snowball: A Language for Stemming Algorithms. 2001 Oct. URL: <http://snowball.tartarus.org/texts/introduction.html> [accessed 2017-09-24] [WebCite Cache ID 6tikyq8wb]
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 2011;12:2825-2830.
36. Elssied N, Ibrahim O, Osman A. A novel feature selection based on one-way anova f-test for e-mail spam classification. *Res J Appl Sci Eng Technol* 2014 Jan 20;7(3):625-638. [doi: [10.19026/rjaset.7.299.s](https://doi.org/10.19026/rjaset.7.299.s)]
37. Chawla NV, Bowyer K, Hall L, Kegelmeyer W. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res* 2002 Jun 01;16:321-357. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]

Abbreviations

LIWC: linguistic inquiry and word count
NLTK: Natural Language Toolkit
SVM: support vector machine

tf-idf: term frequency-inverse term frequency

Edited by G Eysenbach; submitted 27.01.18; peer-reviewed by A Abbe, M Larsen; comments to author 18.02.18; revised version received 22.04.18; accepted 08.05.18; published 21.06.18.

Please cite as:

Aladağ AE, Muderrisoglu S, Akbas NB, Zahmacioglu O, Bingol HO

Detecting Suicidal Ideation on Forums: Proof-of-Concept Study

J Med Internet Res 2018;20(6):e215

URL: <http://www.jmir.org/2018/6/e215/>

doi: [10.2196/jmir.9840](https://doi.org/10.2196/jmir.9840)

PMID: [29929945](https://pubmed.ncbi.nlm.nih.gov/29929945/)

©Ahmet Emre Aladağ, Serra Muderrisoglu, Naz Berfu Akbas, Oguzhan Zahmacioglu, Haluk O. Bingol. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 21.06.2018. 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.

Original Paper

Assessing Statewide All-Cause Future One-Year Mortality: Prospective Study With Implications for Quality of Life, Resource Utilization, and Medical Futility

Yanting Guo^{1,2*}, BM; Gang Zheng^{1*}, PhD; Tianyun Fu^{3*}, BS; Shiyong Hao^{4,5}, PhD; Chengyin Ye^{2,6}, PhD; Le Zheng^{4,5}, PhD; Modi Liu³, BS; Minjie Xia³, BS; Bo Jin³, MS; Chunqing Zhu³, MS; Oliver Wang³, BA; Qian Wu^{2,7}, BS; Devore S Culver⁸, MM; Shaun T Alfreds⁸, MBA; Frank Stearns³, MHA; Laura Kanov³, RRT, BS, MBA; Ajay Bhatia⁹, MD, PhD; Karl G Sylvester², MD; Eric Widen³, MHA; Doff B McElhinney^{4,5}, MD; Xuefeng Bruce Ling^{2,5,10*}, PhD

¹School of Management, Zhejiang University, Hangzhou, China

²Department of Surgery, Stanford University, Stanford, CA, United States

³HBI Solutions Inc, Palo Alto, CA, United States

⁴Department of Cardiothoracic Surgery, Stanford University, Stanford, CA, United States

⁵Clinical and Translational Research Program, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Palo Alto, CA, United States

⁶Department of Health Management, Hangzhou Normal University, Hangzhou, China

⁷China Electric Power Research Institute, Beijing, China

⁸HealthInfoNet, Portland, ME, United States

⁹Department of Pediatrics, Stanford University, Stanford, CA, United States

¹⁰Department of Epidemiology and Health Statistics, School of Public Health, School of Medicine, Zhejiang University, Hangzhou, China

*these authors contributed equally

Corresponding Author:

Xuefeng Bruce Ling, PhD

Department of Surgery

Stanford University

S370 Grant Building

Stanford, CA,

United States

Phone: 1 650 427 9198

Email: bxling@stanford.edu

Abstract

Background: For many elderly patients, a disproportionate amount of health care resources and expenditures is spent during the last year of life, despite the discomfort and reduced quality of life associated with many aggressive medical approaches. However, few prognostic tools have focused on predicting all-cause 1-year mortality among elderly patients at a statewide level, an issue that has implications for improving quality of life while distributing scarce resources fairly.

Objective: Using data from a statewide elderly population (aged ≥ 65 years), we sought to prospectively validate an algorithm to identify patients at risk for dying in the next year for the purpose of minimizing decision uncertainty, improving quality of life, and reducing futile treatment.

Methods: Analysis was performed using electronic medical records from the Health Information Exchange in the state of Maine, which covered records of nearly 95% of the statewide population. The model was developed from 125,896 patients aged at least 65 years who were discharged from any care facility in the Health Information Exchange network from September 5, 2013, to September 4, 2015. Validation was conducted using 153,199 patients with same inclusion and exclusion criteria from September 5, 2014, to September 4, 2016. Patients were stratified into risk groups. The association between all-cause 1-year mortality and risk factors was screened by chi-squared test and manually reviewed by 2 clinicians. We calculated risk scores for individual patients using a gradient tree-based boost algorithm, which measured the probability of mortality within the next year based on the preceding 1-year clinical profile.

Results: The development sample included 125,896 patients (72,572 women, 57.64%; mean 74.2 [SD 7.7] years). The final validation cohort included 153,199 patients (88,177 women, 57.56%; mean 74.3 [SD 7.8] years). The c-statistic for discrimination was 0.96 (95% CI 0.93-0.98) in the development group and 0.91 (95% CI 0.90-0.94) in the validation cohort. The mortality was 0.99% in the low-risk group, 16.75% in the intermediate-risk group, and 72.12% in the high-risk group. A total of 99 independent risk factors (n=99) for mortality were identified (reported as odds ratios; 95% CI). Age was on the top of list (1.41; 1.06-1.48); congestive heart failure (20.90; 15.41-28.08) and different tumor sites were also recognized as driving risk factors, such as cancer of the ovaries (14.42; 2.24-53.04), colon (14.07; 10.08-19.08), and stomach (13.64; 3.26-86.57). Disparities were also found in patients' social determinants like respiratory hazard index (1.24; 0.92-1.40) and unemployment rate (1.18; 0.98-1.24). Among high-risk patients who expired in our dataset, cerebrovascular accident, amputation, and type 1 diabetes were the top 3 diseases in terms of average cost in the last year of life.

Conclusions: Our study prospectively validated an accurate 1-year risk prediction model and stratification for the elderly population (≥ 65 years) at risk of mortality with statewide electronic medical record datasets. It should be a valuable adjunct for helping patients to make better quality-of-life choices and alerting care givers to target high-risk elderly for appropriate care and discussions, thus cutting back on futile treatment.

(*J Med Internet Res* 2018;20(6):e10311) doi:[10.2196/10311](https://doi.org/10.2196/10311)

KEYWORDS

One-year mortality risk prediction; electronic medical records; quality of life; healthcare resource utilization; social determinants

Introduction

Many patients with advanced cancer would prefer to be cared for and die at home. However, among the 50% to 70% of patients with a terminal illness who prefer to be cared for and die at home, only about 25% have a home death, and more than 50% die in the hospital [1]. Nearly a third of Americans who die after age 65 years will have spent time in an intensive care unit in their final 3 months of life, and almost a fifth undergo surgery in their last month [2]. Even more, a disproportionate amount of health care resources and expenditures are spent on patients who are terminally ill [3]. Health care experts estimate that one-quarter of all Medicare costs—US \$150 billion annually—goes to treating patients in their last year of life [4].

Despite aggressive interventions and escalating health costs, delaying unavoidable death may not influence patient outcome and often leads to reduced quality of life [5]. Cancer patients who die in a hospital typically experience more pain, stress, and depression than similar patients who die in hospice or at home [6,7]. Put differently, significant numbers of terminally ill patients may be suitable for and better served by palliative care but are nevertheless readmitted to acute hospitals multiple times [8].

Contributors to this disparity are multifactorial. On one hand, given the complex causal pathways to mortality, it can be difficult for doctors to decide the time and duration of the ultimate episode of decompensation, increasing the uncertainty to making appropriate treatment plans. On the other hand, quality-of-life discussion is associated with less aggressive medical care near death and earlier palliative care, which needs to be balanced with the will of patients to die with comfort, the expectations of families about satisfactory end-of-life care, and saving health resources if possible [5]. Also, mental illness or neurocognitive limitations are common in patients near the end of life, which further complicates assessment and decision making around care in this population [9].

To address this issue, prognostic tools have been developed to identify patients who are approaching a terminal state [10-16]. To date, however, there is no widely acceptable model for timely assessment and risk stratification of all-cause 1-year mortality that can be applied in the general population. Barriers to a widely applicable and accurate model include insufficient risk factors [16], incomplete data available in administrative datasets [17], and lack of generalizability of study patients. Knowledge gaps also exist with regard to the new challenges of social determinants of health (SDH) in terminally ill patients, in terms of the accessibility of health care resources, exposure to hazards, and knowledge of healthy behaviors [18]. Considering SDH in health care decision making could help care teams better target context-informed care, which fills a huge gap between hospital and hospice.

The objective of this analysis was to prospectively validate a machine-learning-based model to estimate a person's risk of all-cause mortality in the next 12 months and assist care providers and families in decision making about appropriate care plans in the last few months of life [5]. The widespread use of electronic medical records (EMRs) affords a unique opportunity to understand health care status and improve care management at the population level. First, our study collected evidence from a rather comprehensive clinical profile, including demographics, medications, diagnoses and procedures, and radiology and laboratory test results for every patient. The breadth and richness of data allowed signals predicting mortality to be detected from the networked clinical patterns. Second, the methodology implemented XGBoost machine-learning techniques to extract valuable information from EMR datasets that could assign a predictive risk score to each individual [19]. Third, a large number of patients from the whole state ensured a certain degree of generalizability. Thus, this study was able to identify patients with diverse demographics and was readily translated to populations of different geographic origins and multiple social disparities.

We hypothesized that the past 12-month clinical histories of patients can be used to predict risk of all-cause mortality within

the next 1 year. This prognostic model aims to provide an objective assessment to aid clinicians in decision making and counseling patients and their families about alternative treatments that incorporate their personal preferences and values. Of equal importance, identifying at-risk elderly patients and providing earlier palliative care may improve their quality of life and thus reduce futile utilization.

Methods

Reporting Method and Ethics Statement

The study was reported according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines for a derivation and validation predictive model. Protected personal health information was removed for the purpose of this research. Analyzing deidentified data, this study was exempted from ethics review by the Stanford University Institutional Review Board (October 16, 2014).

Health Information Exchange Dataset of Maine

Patients for this study were extracted from the Health Information Exchange (HIE) dataset, which covered records of nearly 95% of the population of the state of Maine and was managed by HealthInfoNet, an independent nonprofit organization. The data sources were EMRs collected from 35 hospitals, 34 federally qualified health centers, and more than 400 ambulatory practices in the state of Maine covering about 1 million patients [20,21].

We developed and applied the models using EMR data that included personal demographics, social determinants from the US Census Bureau, laboratory and radiographic tests coded according to Logical Observation Identifier Names and Codes, medication prescriptions coded according to the National Drug Code, and primary and secondary diagnoses and procedures that were coded using the *International Classification of Diseases, Ninth Revision, Clinical Modification*.

Study Sample and Selection Criteria

The study included patients aged 65 years and older who visited any care facility in the Maine HIE network any time from September 5, 2013, to September 6, 2016.

Patients who died before September 5, 2014, were excluded. Those who did not have any active encounters during the 3 years before September 5, 2014, or whose zip codes were not

located in Massachusetts, Maine, or New Hampshire were excluded from the study.

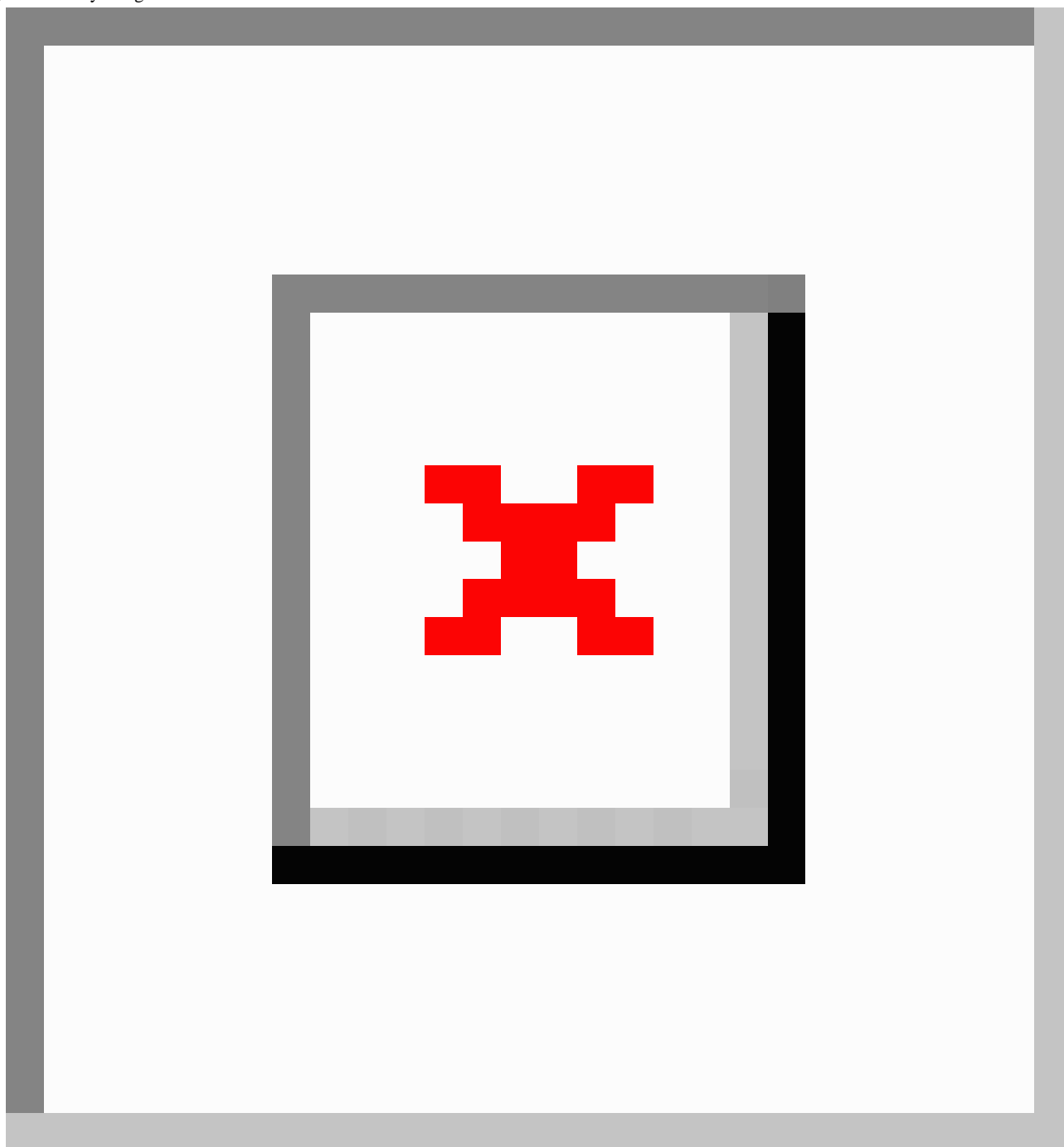
Outcome Definition

A mortality case was defined as a coded date of death in the EMR database in the period from September 5, 2014, to September 4, 2015, in the derivation cohort and from September 5, 2015, to September 4, 2016, in the validation cohort.

Predictive Factors of Mortality

A workflow chart is shown in [Figure 1](#). The selection process was divided into 3 stages: univariate analysis, literature review, and XGBoost selection ([Multimedia Appendix 1](#)). Selection based on P values ($P < .05$, chi-square test or t test) was the initial screening process to trim down the high dimension of the dataset. Literature review was performed in parallel to identify risk factors of mortality that were identified by other studies. Those risk factors included demographics like age, chronic diseases (ie, cerebrovascular disease, cardiovascular disease, rheumatic disease), abnormal laboratory test results (ie, C-reactive protein, potassium), and medication prescriptions (ie, lactulose). Features identified by univariate analysis and literature review went into an XGBoost selection process, where the features were ranked based on their importance of predicting mortality in a model. To improve computational efficiency, we used machine-learning feature selection to determine the features that would go into the model prior to the derivation phase. Chronic disease history variables were modeled as dichotomous using primary and secondary diagnoses. Medication prescriptions were analyzed as the number of prescriptions for a particular medicine during the past 1 year. The thresholds defining laboratory tests as abnormal were set by facilities in the HIE network and treated as continuous variables.

We also assigned 8 SDH variables to each patient: percentage of the population residing in the zip code who were white, percentage of the population residing in the zip code who lived in a rural area, percentage of the population residing in the zip code who attained education at a bachelor's degree level or higher, median household income in the zip code, unemployment rate in the zip code, Gini index of income inequality in the zip code, Social Vulnerability Index in the county (this is a measure of a community's social conditions including socioeconomic status, household composition, minority status, and transportation), and Respiratory Hazard Index in the county (an indication of the adverse effects of pollutants).

Figure 1. Study design.

These social determinants were mapped to the EMR database through a patient's zip code and were categorized according to population quintiles (very low 0%-20%, low 20%-40%, medium 40%-60%, high 60%-80%, very high 80%-100%).

Model Derivation

The derivation cohort was divided into 2 subsets for training and calibration. The initial model was derived based on the training subset: 99 features were input to describe the preceding 1-year clinical profile from September 5, 2013, to September 4, 2014, and the output was set to either 1 or 0 to indicate whether or not a patient was coded with mortality during the period from September 5, 2014, to September 4, 2015.

We adopted XGBoost and tuned the hyperparameters using grid search and cross validation. As a supervised machine-learning technique, it is able to discover statistical patterns in high-dimensional and multivariate data sets and handle nonlinear correlations and random errors both in input features and the output variable.

During the process of model construction, the algorithm generated an ensemble of classification trees and ranked variable importance on the selection frequency of the variable as a decision node [22]. It then summed the scores in the corresponding leaves of each tree to calculate a final predictive estimate \hat{y}_i for the i -th ($i = 1, \dots, n$) instance, as demonstrated in Figure 2, where each f_k corresponded to an independent classification tree and K was the maximum number of trees in

the algorithm. For our study, the depth of each tree was set to be 5 and K equaled 500. We protected against overfitting by penalizing the complexity of the algorithm. Parameters were adjusted to minimize the sum of loss function and the overfitting control term. The sum term at the t iteration was as seen in Figure 3, where l was a differentiable convex loss function that not only measured the difference between the target y_i and the prediction $\hat{y}_i^{(t-1)}$ of the i instance at the $t-1$ iteration but also took the f_t to improve the model most into account. The term Ω was set to penalize the complexity of the regression tree functions in avoid of overfitting. As a splitting method to grow trees, we used an approximate greedy algorithm, and features on each node were sorted to propose a couple of candidates at percentiles. Splitting points were chosen to optimize purity at the next level. The final predictive estimate was summed by individual trees.

In addition, a calibration subset was constructed to convert predictive estimates from the training set to positive predictive values (PPVs), a generalized risk measure with values that described the probability of mortality during the next 1 year. The PPV was calculated as the proportion of mortality in a subset of samples having predictive estimates higher than \hat{y} . In this way, all the predictive estimates were mapped to the calculated PPVs. Patients were then grouped into 3 categories: low risk, intermediate risk, and high risk based on calibrated scores. The relative risk of each patient was calculated as individual score divided by the mean score of all patients in the cohort (baseline). The relative risk indicated the probability of mortality during the next 1 year relative to the baseline.

Model Validation

To test model performance, a validation cohort with clinical history from September 5, 2014, to September 4, 2015, was assembled to predict the risk of mortality from September 5, 2015, to September 4, 2016. The predicted score and relative risk for each patient were calculated. The predictive accuracy of the model was evaluated by calculating the area under the receiver operating characteristic (ROC) curves (discrimination) in both the derivation and validation cohorts, which reflected the ability to distinguish between patients at high and low risk of death. Clinical patterns and social determinants of patients in different risk groups were compared.

Prospective analysis of average and total cost in the year of death and the number of deaths by the top 22 mortality rate commodities in high-risk mortality patients were explored. This was because delaying unavoidable death often contributed to unsustainable and escalating health care costs due to aggressive and expensive interventions. In order to better allocate health care resources spent on treating high-risk patients at the end of life, we evaluated high-risk patients who died and analyzed associations between the cost of care in the last year of life and different chronic diseases. The 22 diagnoses were selected due to their associations with higher mortality among the high-risk patients in our database.

Additionally, studies have documented the escalating treatment cost and poor quality of life associated with significant burden of symptoms. In order to profile the seriously ill elderly population based on the debilitating diseases which may lead to death, we plotted the association between average cost and disease burden grouped by the top 20 chronic diseases of high-risk patients. All analyses were performed using R software (The R Foundation).

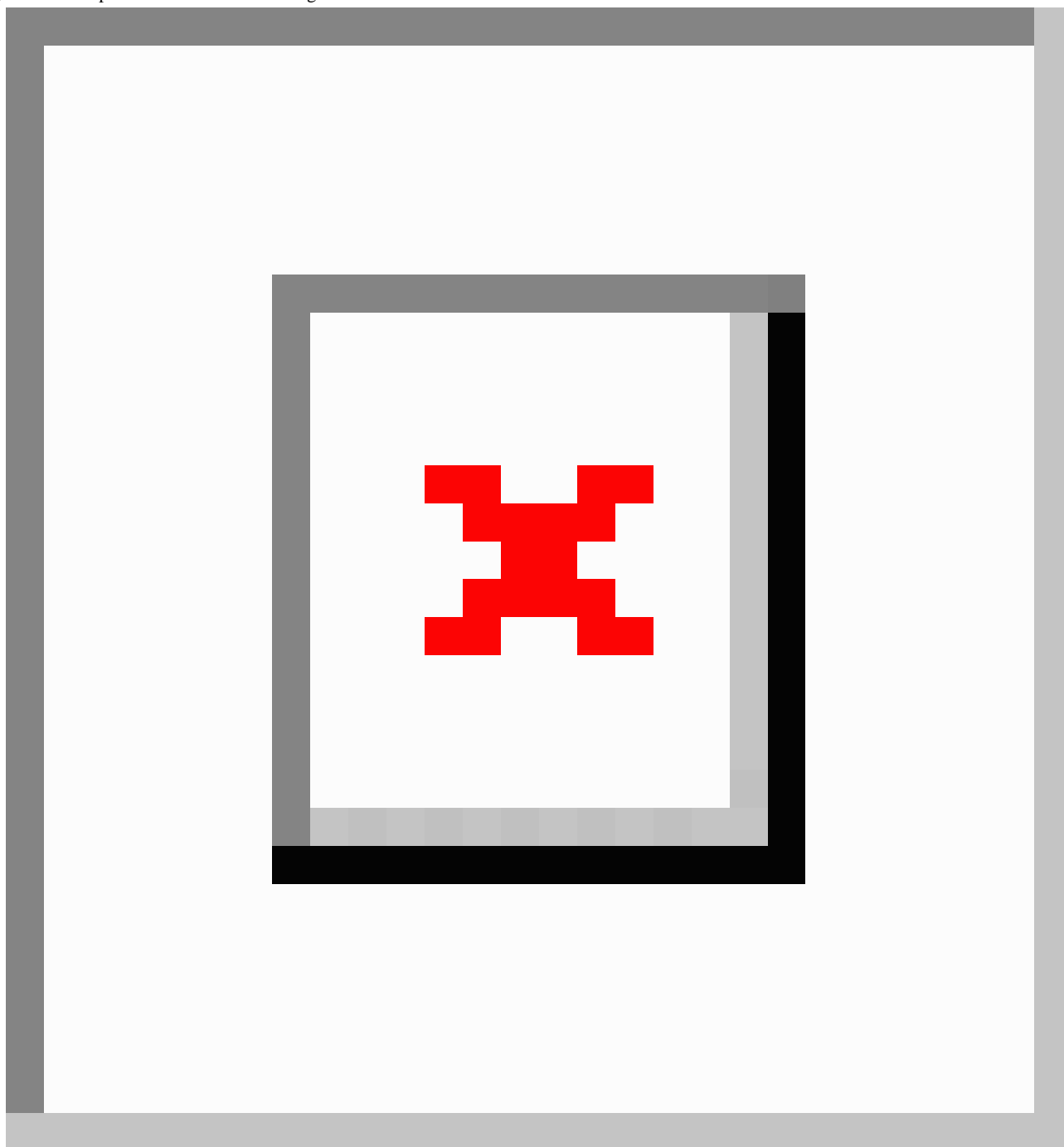
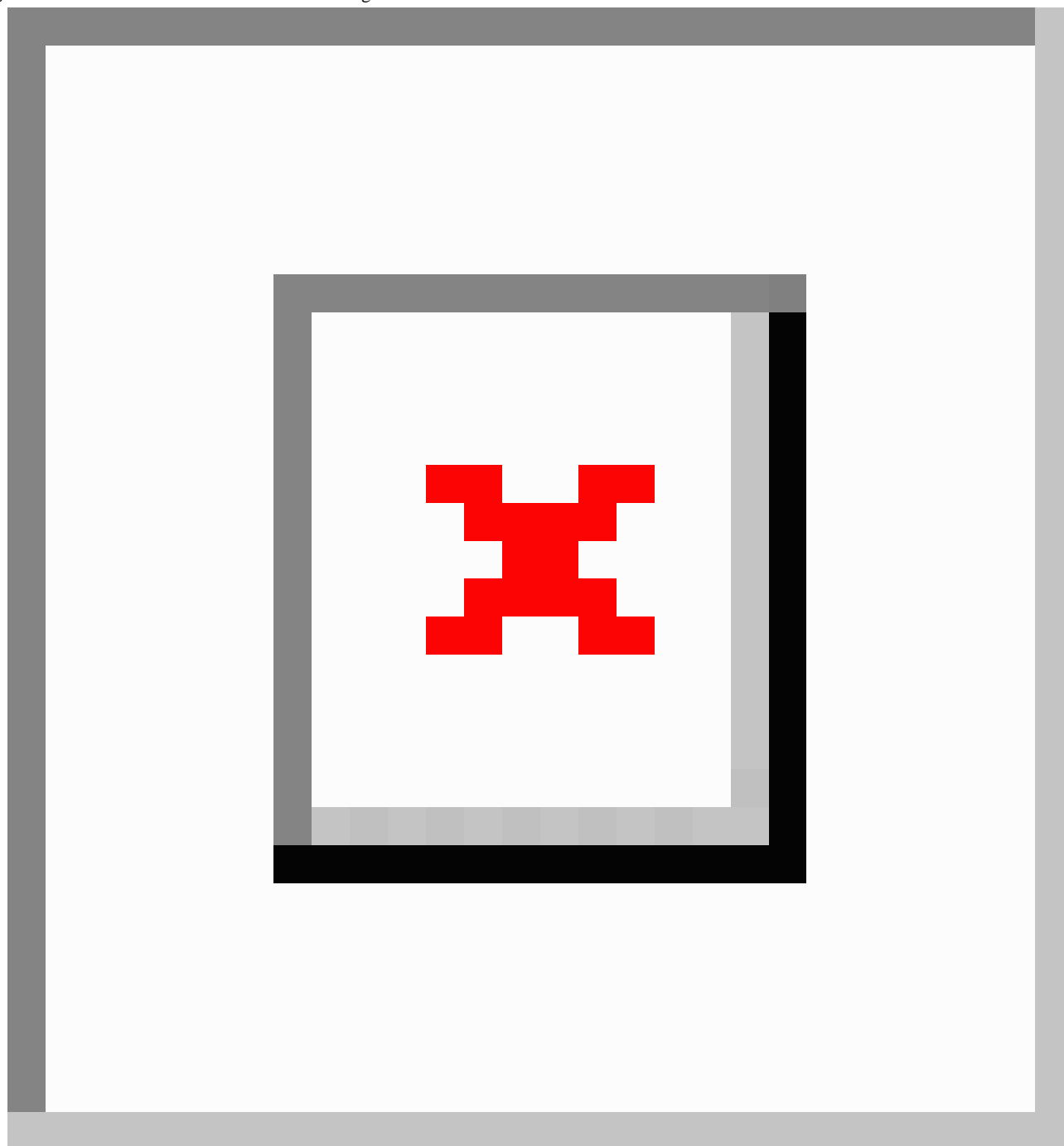
Figure 2. Final predictive estimate of the algorithm.

Figure 3. A sum of loss function and the overfitting control term.

Results

Cohorts and Baseline Characteristics

The final cohort included 125,896 patients for model derivation, 4842 of whom were recorded to have died in the next 1 year (from September 5, 2014, to September 4, 2015), and 153,199 patients for model validation, 5390 of whom died in the next 1 year (from September 5, 2015, to September 4, 2016). A cohort construction diagram is shown in [Figure 4](#).

[Table 1](#) shows the baseline characteristics for patients in derivation and validation cohorts. The 2 cohorts were evenly matched across demographics, payers, and clinical conditions ([Table 1](#)). Specifically, the study involved patients of balanced

age (74.2 years in the derivation vs 74.3 years in the validation) and gender (57.64% [72,572/125,896] in the derivation and 57.56% [88,177/153,199] in the validation). With regard to clinical history, the occurrence of cancer and congestive heart failure, 2 well-established risk features of mortality, were present in 0.65% (989/153,199) and 1.09% (1667/153,199) of the validation cohort, respectively. Type 2 diabetes was present in 4.79% (7337/153,199) in the validation cohort.

Significant Risk Features

Altogether, there were 14,680 features to profile each patient's clinical history in the HIE dataset and socioeconomic status from the public data source. We identified 86 established clinical features of mortality from the literature review. In addition, 653

features survived after the univariate analysis and literature review. XGBoost used the approximate greedy algorithm to split trees by sorting and picking features on each node in order to optimize purity at each splitting level. Finally, a total of 99 features were selected as model predictors. The top 45 univariate features of mortality for elderly patients are shown in Table 2.

In accordance with previous studies, age (≥ 85 years) was recognized as the most impactful demographic feature in mortality risk. In our prospective analysis, the percentage of patients aged 65 to 74 years accounted for 59.25% (90,770/153,199) of the total population, and 25.60% (1380/5390) of all deaths. Comparatively, older adults (≥ 85

years) composed 12.62% (19,311/153,199) of the population and 37.35% (2013/5390) of all deaths. Furthermore, we analyzed the death rate of 4 priority noncommunicable diseases and 2 high-prevalence chronic diseases among different age groups (65-69 years, 70-74 years, 75-79 years, 80-84 years, and ≥ 85 years) (Multimedia Appendix 2). There was a significant rise in the percentage of death cases for cardiovascular disease and hypertension that presented in all patients aged ≥ 85 years. The percent of mortality cases of patients with chronic kidney disease diagnosis also increased nearly 10 times when comparing the 65- to 69-year age group with the 85-year age group. Cardiovascular disease showed the highest percentages of mortality cases in all age groups.

Figure 4. Construction of derivation and validation cohorts.

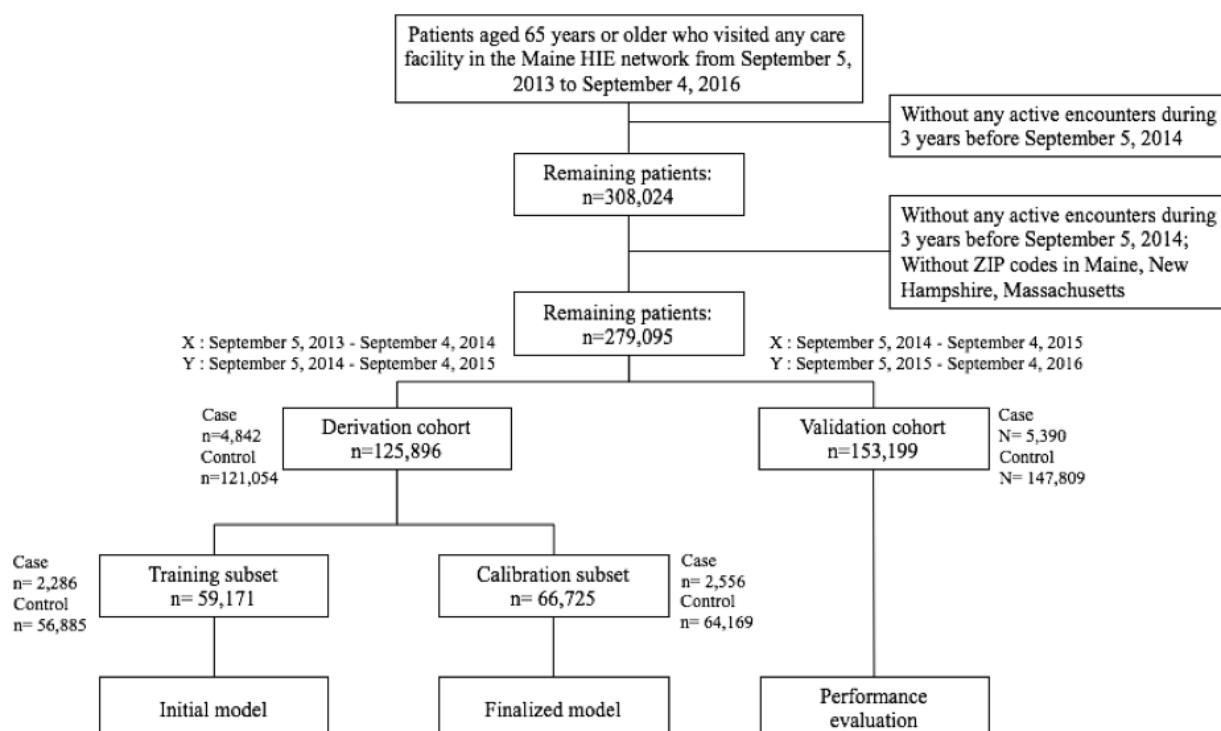


Table 1. Baseline characteristics.

Characteristic	Derivation cohort (n=125,896), n (%)	Validation cohort (n=153,199), n (%)	P value
Age (years)			.009
65-74	73,989 (58.77)	90,770 (59.25)	—
75-84	36,076 (28.66)	43,098 (28.13)	—
≥85	15,831 (12.57)	19,331 (12.62)	—
Female	72,572 (57.64)	88,177 (57.56)	0.6
Race			<.001
White	99,206 (78.80)	123,632 (80.70)	—
Black	126 (0.10)	306 (0.20)	—
Asian	19,010 (15.10)	20,682 (13.50)	—
Other/unknown	7554 (6.00)	8579 (5.60)	—
Medicare	16,841 (13.38)	20,008 (13.06)	.01
Medicaid	263 (0.21)	341 (0.22)	.50
Comorbid conditions			.05
Cancer	841 (0.67)	989 (0.65)	—
Type 2 diabetes	6019 (4.78)	7337 (4.79)	—
Renal disease	1911 (1.52)	2489 (1.62)	—
Anemia	2879 (2.29)	3759 (2.45)	—
Congestive heart failure	1386 (1.10)	1667 (1.09)	—
Cerebrovascular accident/stroke	1747 (1.39)	2280 (1.49)	—
Obesity	1465 (1.16)	1777 (1.16)	—

Model Performance

Model outcomes of the derivation and validation phases are showed in [Table 3](#). For derivation, the model had a c-statistic of 0.960 (see [Multimedia Appendix 3](#) and [Multimedia Appendix](#)). Patients who died in the next 1 year (n=4842) had a mean relative risk of 30.91 (probability of mortality 30.91 times more than baseline). Among the 4842 patients, 595 were stratified as low risk, 1591 as intermediate risk, and 2656 as high risk. The mortality incidence and relative risk increased monotonically from low-risk (0.5%, 0.05) to high-risk (100%, 30.99) groups.

The model performance was slightly lower in the validation cohort (c-statistic 0.912), with 5390 patients who died in the next 1 year and a mean relative risk of 6.15. The total numbers of low-, intermediate-, and high-risk patients were 1384, 1593, and 2413, respectively, with high-risk patients accounting for 44.77% (2413/5390) of all patients who died within the next 1 year. The mortality incidence and relative risk climbed monotonically from low-risk (1.0%, 0.05) to high-risk (72.1%, 36.64) groups.

Risk Stratification and Clinical Patterns

As demonstrated in [Table 4](#), clinical patterns were compared among the low-, intermediate-, and high-risk groups in the

validation cohort. There was an obvious difference in age distribution between the low-risk and other 2 groups. The average ages of low- and high-risk groups were 72 and 84 years, respectively. Survival analysis among the 3 risk groups showed the model to have good risk stratification in general ([Multimedia Appendix 5](#)).

In addition, patients in the high-risk group suffered from more severe comorbidities and used more health care resources. The proportion of patients with cancer of the bronchus (lung) in the high-risk group (6.54%) was much higher than in the low-risk group (0.11%) ($P<.001$). Chronic diseases such as type 2 diabetes, congestive heart failure, and chronic obstructive pulmonary disease (COPD) were present in 16.19% (542/3347), 19.18% (642/3347), and 16.61% (556/3347) of the high-risk group, compared with 3.04% (4273/140,344), 0.06% (82/140,344), and 1.46% (2055/140,344) of the low-risk group ($P<.001$), respectively. With regard to laboratory tests, abnormal complete blood count, metabolic panel, urinalysis, and coagulation tests were present 48.07% (1609/3347), 51.33% (1718/3347), 24.59% (823/3347), and 11.95% (400/3347), respectively, of high-risk patients, while in the low-risk group, the percentages were much lower, at 0.18% (216/140,344), 0.27% (384/140,344), 0.06% (84/140,344), and 0.04% (58/140,344), respectively.

Table 2. Top 45 risk features in the final model with odds ratio and 95% confidence interval.

Category and differentiating features	Odds ratio	95% CI
Demographics		
Age ≥85 years	1.41	1.06-1.48
Social determinant		
Respiratory Hazard Index	1.24	0.92-1.40
Unemployment rate	1.18	0.98-1.24
Percent of population who lived in rural area	1.10	1.00-1.10
Diagnosis		
Congestive heart failure	20.90	15.41-28.08
Cancer of ovary	14.42	2.24-53.04
Cancer of colon	14.07	10.08-19.08
Cancer of stomach	13.64	3.26-86.57
Cancer of bronchus, lung	12.38	2.91-36.04
Chronic kidney disease	11.96	8.49-16.29
Cancer of liver and intrahepatic bile duct	11.59	1.81-41.01
Renal failure	11.22	8.88-14.06
Cerebrovascular accident/stroke	9.31	5.59-14.68
Cancer of brain and nervous system	8.65	2.07-24.4
Rheumatic disease	6.15	3.85-9.12
Myocardial infarction	6.13	5.21-7.29
Leukemia	5.01	1.23-13.89
Malnutrition	4.66	1.07-22.32
Peripheral arterial disease	4.58	1.77-9.49
Somnolence	2.99	1.85-4.43
Cancer of breast	2.70	1.59-4.26
Dementia	2.57	1.76-8.67
Diabetes mellitus	1.43	0.36-2.22
Laboratory test		
Hematocrit	4.13	2.00-6.31
Potassium	3.55	2.50-4.76
B-type natriuretic peptide	2.76	2.08-3.57
Glucose	1.54	1.42-1.57
C-reactive protein test	1.41	1.30-1.62
Platelets	1.32	1.02-1.42
Medication		
Pazopanib hydrochloride	3.66	1.92-10.65
Lactulose	1.89	1.04-2.13
Abiraterone acetate	1.85	1.34-2.45
Metolazone	1.67	1.37-1.93
Omeprazole	1.67	1.04-1.89
Phenytoin sodium extended	1.61	0.96-1.78
Furosemide	1.58	1.13-1.71
Venlafaxine hydrochloride	1.54	0.98-1.63

Category and differentiating features	Odds ratio	95% CI
Clotrimazole	1.38	1.05-1.54
Cephalexin	1.30	1.17-1.46
Fluticasone/salmeterol	1.26	1.07-1.25
Rifaximin	1.22	0.95-1.23
Glipizide	1.19	1.07-1.36
Olanzapine	1.13	1.00-1.69
Carvedilol	1.10	1.07-1.13
Utilization		
Inpatient days in the past 12 months	1.33	1.13-1.72

Table 3. Comparison of the model outcome in derivation and validation cohorts.

Outcome	Derivation cohort (n=125,896)	Validation cohort (n=153,199)
Died in the next 1 year, n (%)	4842 (3.84)	5390 (3.52)
Risk score model		
Baseline score, mean (SD)	0.032 (0.035)	0.011 (0.072)
Baseline score for mortality patients in the next 1 year, median (1st, 3rd quartile)	0.99 (0.11, 0.99)	0.067 (0.01, 0.34)
Relative risk ^a for mortality patient in the next 1 year, median (1st, 3rd quartile)	30.91 (3.48, 31.06)	6.15 (0.86, 31.42)
Mortality risk category: low/intermediate/high	595/1591/2656	1384/1593/2413
Percent incidence of mortality (95% CI)		
Low	0.50 (0.40, 0.60)	1.00 (0.80, 1.20)
Intermediate	11.5 (11.0, 12.4)	16.80 (16.20, 17.52)
High	100 (100, 100)	72.10 (71.50, 73.10)
Relative risk for the population baseline (95% CI)		
Low	0.05 (0.04, 0.05)	0.052 (0.048, 0.055)
Intermediate	2.76 (2.67, 2.88)	2.45 (2.41, 2.48)
High	30.99 (30.9, 31.0)	36.64 (36.12, 37.07)

^aRelative risk of each patient was defined as the ratio of the risk score of the patient to the baseline score (ie, the mean risk score of total population).

For high-risk elderly patients, we performed survival analysis for 4 leading causes of death defined by the World Health Organization: cardiovascular diseases, cancers, COPD, and type 2 diabetes. We found that all 4 chronic disease categories had a steep decrease in survival opportunity over time, indicating that our prognosis model aligned with current findings regarding major health burdens and high mortality among high-risk aged patients ([Multimedia Appendix 6](#)).

With respect to SDH, more high-risk patients lived in a community with high unemployment rate (29.46% [986/3347] vs 21.84% [30,646/140,344] in the low-risk cohort). Differently, 10.16% (340/3347) of high-risk patients had low median household income in their community, slightly lower than

low-risk counterparts (17,560/140,344, 12.51%). Unemployment rate and education attainment (percentage of population who attained education at bachelor's degree level or higher) contributed toward mortality risk in high-risk elderly patients, which made them more vulnerable to end-of-life care (see [Multimedia Appendix 7](#)).

For health care resource utilization, the mean cost during the last 12 months per patient in the high-risk group (US \$10,575) was substantially higher than in the low-risk cohort (US \$680). Seriously ill patients also used more health care resources as indicated by the greater number of outpatient visits at the end of life, 12 per patient per year, compared with 3 in low-risk patients.

Table 4. Clinical patterns of patients by risk categories in the validation cohort.

Characteristic	Low risk (n=140,344)	Intermediate risk (n=9508)	High risk (n=3347)
Age, years, median (1st, 3rd quartile)	72 (68, 78)	86 (80, 91)	84 (77, 90)
Female, n (%)	81,041 (57.74)	5356 (56.33)	1780 (53.18)
Race (white) , n (%)	113,678 (81.00)	7530 (79.20)	2510 (74.99)
Diagnosis, n (%)			
Cancer of bronchus (lung)	163 (0.11)	60 (0.63)	219 (6.54)
Cancer of prostate	1306 (0.93)	97 (1.02)	92 (2.74)
Cancer of bladder	218 (0.15)	43 (0.45)	50 (1.49)
Cancer of breast	1052 (0.75)	65 (0.68)	63 (1.88)
Cancer of head and neck	138 (0.09)	18 (0.19)	14 (0.42)
Cancer of colon	68 (0.05)	23 (0.24)	49 (1.46)
Anemia	660 (0.47)	251 (2.64)	492 (14.70)
Pure hypercholesterolemia	5733 (4.08)	399 (4.19)	414 (12.37)
Type 2 diabetes	4273 (3.04)	468 (4.92)	542 (16.19)
Chronic kidney disease	266 (0.19)	114 (1.19)	285 (8.51)
Chronic liver disease and cirrhosis	434 (0.31)	52 (0.54)	104 (3.10)
Congestive heart failure	82 (0.06)	140 (1.47)	642 (19.18)
Chronic obstructive pulmonary disease	2055 (1.46)	483 (5.08)	556 (16.61)
Leukemia	57 (0.04)	4 (0.04)	24 (0.72)
Dementia	250 (0.17)	171 (1.79)	75 (2.24)
Community-level social determinant, n (%)			
Zip code with high median household income	17,560 (12.51)	1129 (11.87)	340 (10.16)
Zip code with high percentage of population who lived in rural area	86,577 (61.69)	5275 (55.48)	1962 (58.62)
Zip code with high unemployment rate	30,646 (21.84)	2140 (22.51)	986 (29.46)
Zip code with high percentage of population who attained education at bachelor level or higher	24,802 (17.67)	1634 (17.19)	720 (21.51)
Medication, n (%)			
Hypertension	27,962 (19.92)	5845 (61.47)	2681 (80.10)
Seizures	3775 (2.69)	952 (10.01)	584 (17.45)
Chronic obstructive pulmonary disease	4420 (3.15)	1410 (14.83)	902 (26.95)
Heart	11,897 (8.47)	3054 (32.12)	1622 (48.46)
Mental illness ^a	9144 (6.51)	2602 (27.36)	1352 (40.39)
Lab test, n (%)			
Abnormal complete blood count	216 (0.18)	523 (5.50)	1609 (48.07)
Abnormal metabolic panel	384 (0.27)	646 (6.79)	1718 (51.33)
Abnormal urinalysis	84 (0.06)	229 (2.41)	823 (24.59)
Coagulation test	58 (0.04)	77 (0.81)	400 (11.95)
Utilization, mean (1st, 3rd quartile)			
Cost past 12 months, US \$	680 (340, 1360)	1700 (680, 4420)	10,575 (3230, 23,796)
Mean outpatient visit per 12 months	3 (1, 5)	5 (2, 10)	12 (6, 24)

^aDonepezil hydrochloride, lorazepam, prochlorperazine maleate, memantine hydrochloride, risperidone, haloperidol, paroxetine hydrochloride, rivastigmine, zolpidem tartrate, venlafaxine hydrochloride, temazepam, amitriptyline hydrochloride, olanzapine, and nortriptyline hydrochloride.

When we focused the analysis on high-risk patients with dementia, given the increasing attention to mental illness of terminally ill patients [23,24], we found a higher prevalence of dementia among high-risk elderly patients (2.24% [75/3347] vs 0.17% [250/140,344] in the low-risk cohort, $P<.001$). About 40.39% (1352/3347) of high-risk patients took medications for mental illness health conditions, substantially higher than in the low- (9144/140,344, 6.51%) and intermediate-risk (2602/9508, 27.36%) groups. We also compared mortality and health care use between high-risk patients with and without dementia (see [Multimedia Appendix 8](#)). Although the average ages in 2 groups were similar (82.8 vs 83.3 years), the mortality rate of patients who had dementia was slightly lower (41/75, 54.67%) than those at-risk patients without dementia (2372/3272, 72.49%) ($P=.003$). Dementia patients incurred less health care spending in the past 12 months (US \$2795 vs \$10,805) ($P<.001$) than patients without dementia and had a lower chronic disease burden (7.5 vs 10.9) ($P<.001$) and fewer inpatient days (1.3 vs 8.9), inpatient admissions (0.2 vs 1.3), and emergency department visits (1.1 vs 1.8) ($P<.001$). The community in which dementia patients lived was characterized by lower household income (US \$40,407 vs \$44,588) ($P<.001$).

Average and Total Cost in the Final Year of Life of High-Risk Patients

[Figures 5,6](#), and [Multimedia Appendix 9](#) depict the average and total costs of medical care in the previous 1 year before death,

which characterized the pertinent clinical profile and expenses of patients who died any time in the predictive year.

In the prospective cohort, the average cost in the last year of life overall was US \$2346, and for the high-risk group it was US \$21,799 ([Figure 5](#)). Among high-risk patients who expired, cerebrovascular accident, amputation, type 1 diabetes, obesity, and rheumatic diseases were the top 5 diseases in terms of 1-year average cost: US \$64,756, \$61,692, \$40,329, \$37,548, and \$35,167, respectively. The percentages of high-risk patient deaths in those who had these conditions were 4.10% (99/2413), 2.15% (52/2413), 2.45% (59/2413), 10.73% (259/2413), and 5.64% (136/2413), respectively. This highlighted that elderly patients burdened with these diseases were likely to die with high attendant costs.

For 1-year total cost of mortality among high-risk patients ([Figure 6](#)), myocardial infarction, hyperlipidemia, congestive heart failure, edema, and shortness of breath were the top 5 diseases, given the absolute numbers of deaths in patients with these conditions in our study cohort. These diseases amounted to US \$41,175,717, \$34,227,212, \$17,664,371, \$17,495,202, and \$16,063,029, respectively, and the percentages of high-risk patient deaths in those with these conditions were 62.45% (1507/2413), 51.72% (1248/2413), 26.52% (640/2413), 21.47% (518/2413), and 22.96% (554/2413), respectively.

Figure 5. Prospective analysis of average cost in the year of death and the number of deaths by the top 22 mortality rate commodities in high-risk mortality patients.

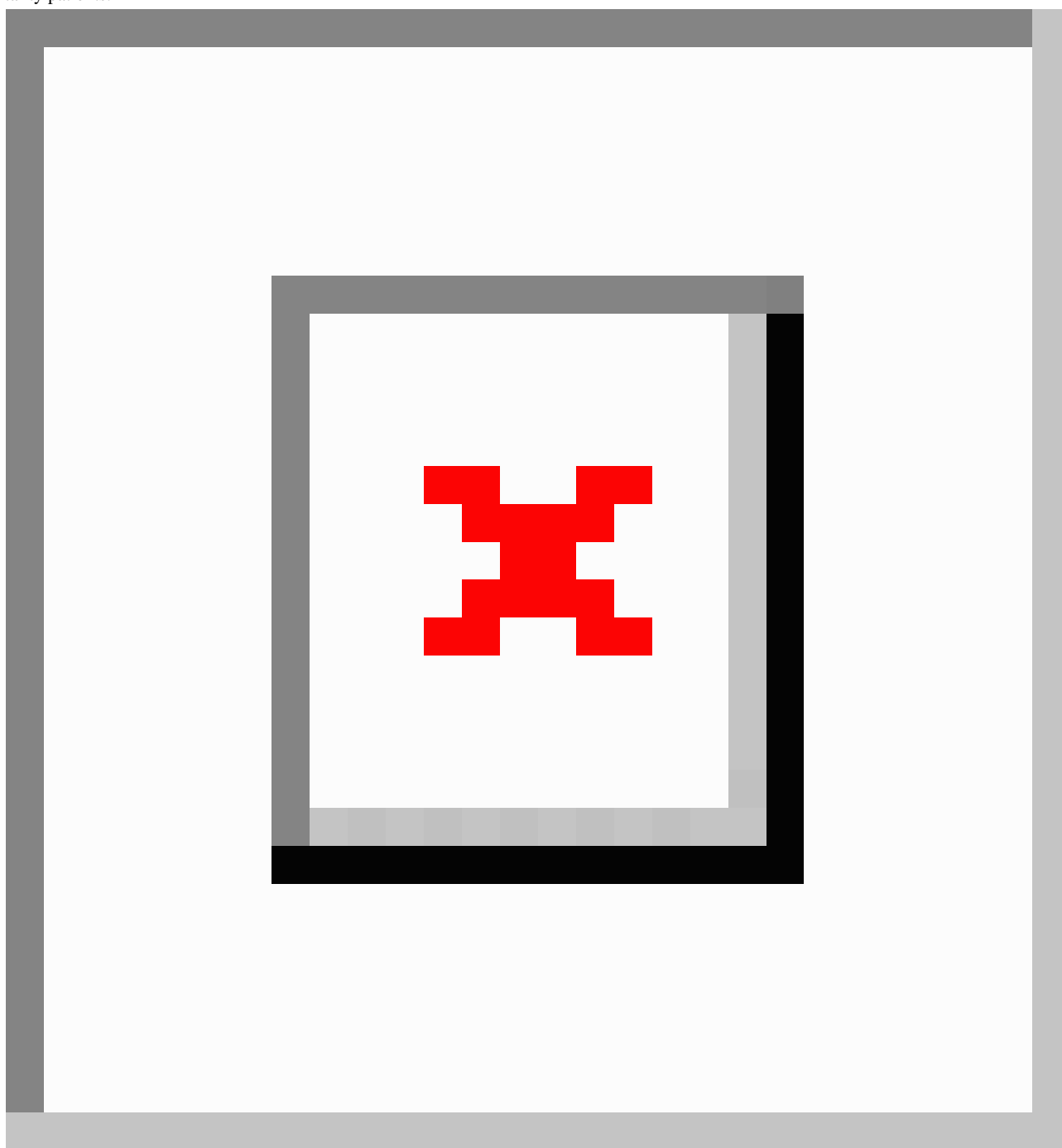
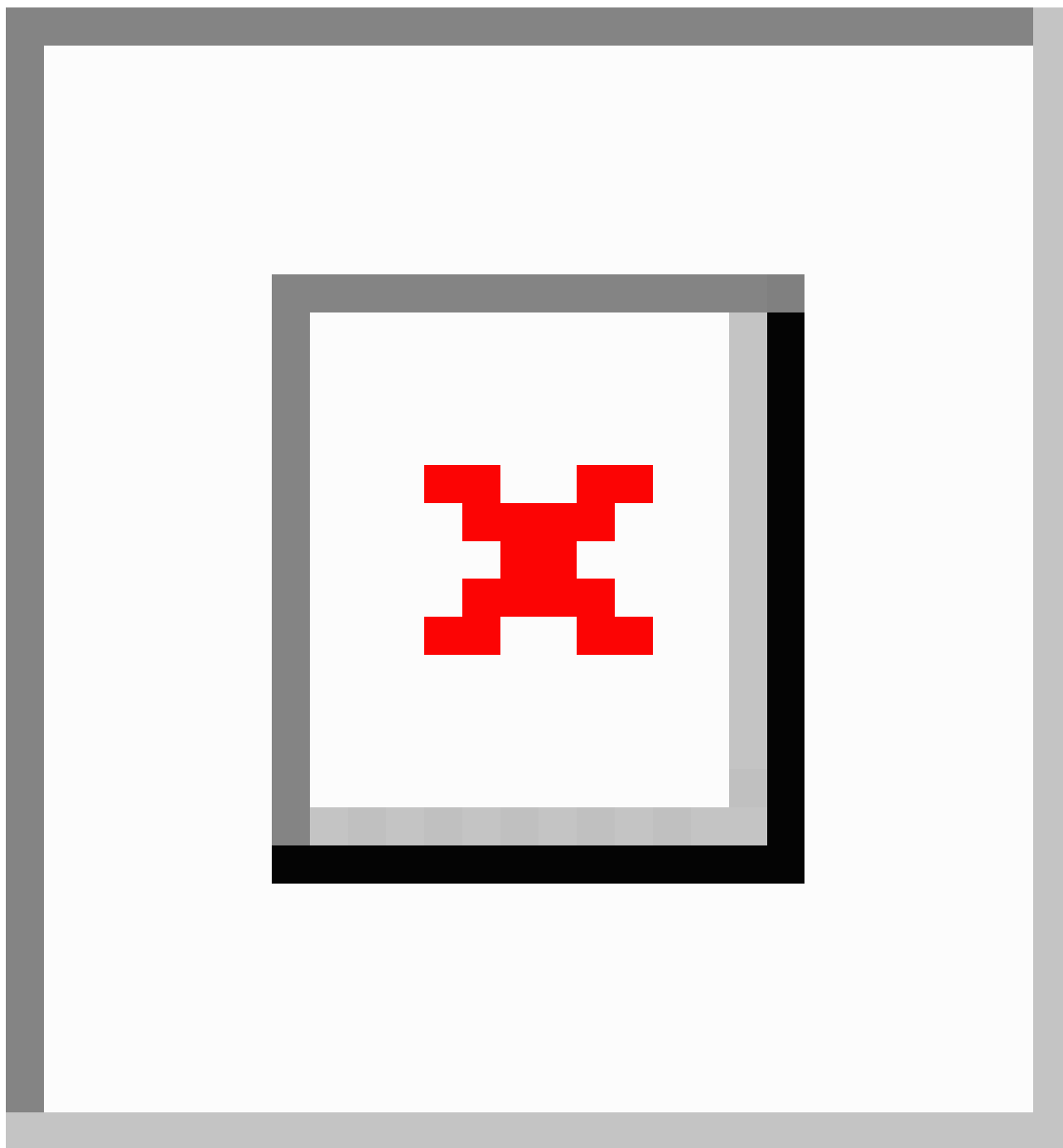


Figure 6. Prospective analysis of total cost in the year of death and the number of deaths by the top 22 mortality rate commodities in high-risk mortality patients.



Associations Between End-of-Life Cost, Resource Utilization, and Disease Burden of High-Risk Patients

We found congestive heart failure (642/3347), COPD (556/3347), and type 2 diabetes (542/3347) were the 3 chronic diseases with the largest populations among these 20 diseases (see [Multimedia Appendix 10](#)). Patients with these 3 diagnoses had an average of 16.9, 14.0, and 16.1 chronic diseases. Patients with somnolence had the largest number of chronic diseases (20.1 per member) and the fourth highest average cost (US \$34,378) of high-risk patients. Patients with leukemia were comorbid with nearly 18.5 chronic diseases at their end of life, with a high average cost of US \$32,514 (compared to the overall

average of US \$21,799). The pattern differed from that of low-risk patients (see [Multimedia Appendix 11](#)).

Meanwhile, health care use also increased with end-of-life diseases (see [Multimedia Appendix 12](#) and [Multimedia Appendix 13](#)). Among terminally ill patients, patients with somnolence and renal failure had the most inpatient admissions (2.6 times per member per year [PMPY]) and emergency department visits (3.4 times PMPY) at end of life. As mentioned previously, elderly patients with cerebrovascular accidents had relatively high mean cost in the final year of life as well as a high number of inpatient admissions (2.3 times PMPY) and a heavy disease burden (15.5 per member). For high-prevalence comorbid conditions, terminally ill patients with COPD, chronic

kidney disease, and type 2 diabetes had a mean number of emergency department visits more than twice PMPY. Comparatively, high-risk elderly patients with dementia had fewer inpatient admissions (0.24 times PMPY) and emergency department visits (1.3 times PMPY) in their last year of life.

Discussion

Principal Findings

In an attempt to address the enormous treatment expenditures and unmet needs of patients approaching the end of life, we prospectively validated a tool to predict the all-cause 1-year mortality of statewide patients aged 65 years and older. To our knowledge, this study was the first to examine mortality of an elderly population with respect to end-of-life care, cost, and resource use.

Both the model outcomes and survival analysis supported the effectiveness and accuracy of our model in risk stratification of the high-risk group. The model performances in derivation and validation phases were excellent, with c-statistics of 0.960 and 0.912, respectively. It outperformed other models that were derived from limited numbers of risk factors or administrative data sources (see [Multimedia Appendix 14](#)). Further, the use of a statewide population in this study supported the generalizability of our findings to the nation as a whole.

The overall mortality in our study was 3.52% for elderly patients (65 years and older) in the state of Maine. The overall mortality of the elderly population in 2015 was about 4.3% in the state of Maine [25], which was not far from the rate of our data pool.

Given that our cohort was highly imbalanced with 5390 cases and 147,809 controls derived from the 35 hospitals in the state of Maine with 1 million patients, the negative predictive values were expected to be high. However, our study goal was set to deliver robust PPVs (describing the probability of mortality within the next 1 year), and that can be challenging. Our model has prospectively identified the high-risk patients across the Maine state for population health, balancing the PPV and sensitivity requirements (see [Multimedia Appendix 15](#), where PPV is 0.99% in the low-risk group versus 72.12% in the high-risk group).

Although some argued that conclusions in EMR studies may not be easily drawn [26], we have tried to improve the robustness of our study. First, our study is a prospective analysis originated from a dataset that covered records of nearly 95% of the statewide population in Maine. Unlike a randomized clinical trial (RCT) that has a well-designed sample cohort with targeted outcomes, our study was set to develop a predictive model using a statewide population with comprehensive clinical history including patient demographics, encounter history, vital signs, laboratory and radiology results, medication history, diagnoses, and procedures. Rather than providing a direct solution (ie, optimum treatment option) to each individual to extend life spans, our model intended to identify high-risk patients at early stages to provide early warning signals for improved health care quality and improved health care resource use.

Second, the active case-finding model and associated online real-time application were designed to track the evolving nature of total population risk of mortality in a longitudinal manner across all payers and diseases of elderly patients. Results were visualized on a real-time 24/7 online dashboard. This empowers the accountable care organization field staff and population health managers to visualize the risks derived from each resident's historical medical records in the state of Maine. This tool is able to identify patient at high risk of 1-year mortality among the fragmented nature of population health information and improve quality of life and reduce futile treatment.

Interpretation of Features

A list of risk factors survived after a feature selection process that integrated machine learning with clinical knowledge from the literature and practice ([Table 2](#) and [Multimedia Appendix 16](#)). The importance of age to mortality risk was consistent with prior clinical studies and reflects a clinical scenario in which the association between older age and mortality may be explained by greater disease burden, associated complications, and functional impairment [27]. The age cutoff that best predicts mortality in elderly patients in our study and others is usually in the range of 80 to 85 years, an age group with a high prevalence of frailty, dependence, and geriatric syndromes according to related findings [28].

The model also highlighted the detrimental impacts of several chronic conditions, including congestive heart failure, kidney disease, cerebrovascular accident, rheumatic diseases, and myocardial infarction. Consistent with prior studies, we found congestive heart failure contributed significantly to the risk of 1-year mortality [11,29]. The illness trajectory of most heart and other organ system failure was distinct from that of cancers and comprised gradual decline, punctuated by episodes of acute deterioration and some recovery, with more sudden and seemingly unexpected death [30]. The hematocrit was the top laboratory test feature in this study and can indicate anemia or leukemia with lower values or lung and heart disease with higher values. Serum potassium levels can be associated with adverse outcome in patients with cardiovascular disease, with a U-shaped relationship between serum potassium levels and mortality in patients with acute myocardial infarction [31].

In our large cohort of elderly patients with various tumor sites, cancers of the ovary, colon, stomach, bronchus (lung), liver and intrahepatic bile duct, and brain and nervous system were the most impactful factors, independently associated with overall 1-year mortality. Overall, 70% of cancer deaths occur in patients older than 65 years, and for most cancers, there was a trajectory of steady progression and a clear terminal phase over a period of weeks, months, or years before death [30]. Factors such as age older than 80 years, functional impairment, mobility impairment, higher number of severe comorbidities, and malnutrition were common pathways that may increase the risk of cancer-related mortality [28]. Additionally, the adverse effect of metastatic status was greatest for breast and prostate cancers. Pazopanib hydrochloride, the most impactful medication in our model, is used to treat patients with advanced soft tissue sarcoma or gastrointestinal stromal tumors who have received prior chemotherapy.

In this study, dementia was found to increase mortality risk for elderly patients. As mentioned in prior studies, many features were reported as predictive of death in patients with dementia, including age, functional impairment, and disease severity [32]. However, it was likely that these patients had less accessibility to health care resources due to their cognitive impairment or that professional caregivers already viewed palliative care as appropriate for patients with end-stage dementia. Given the concentration of aging patients in the high-risk group, dementia as a test of the comprehensive geriatric assessment, especially for older surgical patients, can predict postoperative outcomes [33] and thus guide personalized approaches to medical care. Further, prescribing of antipsychotic medications for dementia patients was associated with higher mortality rates [34]. We also found that medications used to treat mental or mood disorders contributed to higher mortality probability, in accordance with prior literature. The side effects of psychotropic medications, particularly weight gain and impaired glucose tolerance, may increase the risk of excess mortality in people with mental illness [35]. It was also reported that conventional antipsychotics were associated with higher mortality than atypical antipsychotics [36]. These findings should give rise to more attention to mental illness in terminally ill patients, not only because of higher fatality rates from cancer in psychiatric patients [23] but also because of great opportunities to improve end-of-life care for these vulnerable patients, given their decreased ability to communicate need and the severe physical consequences [37].

Last, this study also featured SDH at the community level (zip code level), which was recently recognized as increasingly influential on morbidity and mortality [38–40]. Among several social determinant inputs, the Social Vulnerability Index and Respiratory Hazard Index were highly weighted. They reflected the degree to which a community exhibited certain social conditions (eg, high poverty or crowded households) and was exposed to pollution, respectively. Notably, patients with different racial or income backgrounds have been found to vary in their treatment preferences, advanced care planning, and access to health care resources [24]. Based on these findings, care providers may consider SDH information in their assessment of end-of-life medical care and prognosis.

Planning for a “Good Death” in Terminal Phase

Death is inevitable, but there are a variety of ways to care for dying patients. A good death—“one that is free from avoidable death and suffering for patients, families, and caregivers in general accordance with the patients’ and families’ wishes” [41]—often optimizes the quality of life of terminally ill patients before a timely, dignified, and peaceful death.

Based on prospective validation of the statewide elderly population, the meaningful use of our model may be to stratify the population and identify patients at high risk of mortality, for whom timely targeted curative treatments may be indicated or palliative care plans may yield better quality of life and lower medical cost.

Planned treatments can be curative or palliative, depending on the diseases. Currently, the delivery of palliative care mainly targets malignant diseases. The trajectory of most cancers may

be punctuated by the positive or negative effects of palliative oncological treatment. Most weight loss, impaired ability, and reduction in performance status for self-care occurs in patients’ last few months [30]. With earlier identification and open discussion about prognosis, there is time to anticipate palliative needs and plan for end-of-life care.

Early intervention (survival prognosis 6 to 24 months) by a palliative care team can help improve symptom control and satisfaction with psychosocial support and decision assistance [42]. Consequently, specialist palliative care is a recommended element of care for patients with cancer, especially cancers with poor survival rates. Positive benefits of specialist palliative care services in hospital teams, home care teams, and inpatient services have been documented [43]. Given the significant advances that have been made in the treatment of certain cancers, local health care systems need to ensure that these treatment advances are accessible in areas of high incidence. Care providers can also act on lifestyle choices to improve prevention, such as smoking, obesity, and diet, which have been identified as leading causes of cancer mortality [44].

Long-term limitations with intermittent serious episodes are typical among seriously ill patients with heart failure, chronic respiratory diseases, or other organ failure. Deteriorations are generally associated with admission to the hospital and intensive treatment. In this sense, advanced identification by the prognostic tool will contribute more to informing the timing of death and planning for terminal care in a preferred setting.

For many life-threatening diseases like congestive heart failure, actions can be taken and conditions can be managed to help avoid escalating pain. The treatment aim of symptom relief held greater importance to physicians for elderly patients, while delay of death was thought to be more important for relatively younger patients, as suggested by an international survey [45]. For example, follow-up monitoring by specially trained staff, access to specialized heart failure clinics, and other multidisciplinary strategies appeared to be efficacious to improve outcomes for heart failure patients [46].

People who escape cancer and organ system failure may die at an older age of either brain failure (such as Alzheimer or other dementia) or generalized frailty of multiple body systems [47]. The disease course that dementia usually follows is one of prolonged and progressive disability, which makes identification of the terminal phase very difficult. Such patients may lose weight and show a variety of symptoms like depression and neurologic signs occurring in combination with declining reserve that can prove fatal. Despite the wishes of the majority of dementia patients and their families to die at home [48], many frail elderly patients with dementia are currently admitted to the hospital to die when terminally ill. The use of end-of-life care pathways in nursing homes is proving increasingly effective in preventing such admissions [7]. The prognostic tool helps identify patients with dementia who are approaching the end of life in order to plan care and make provisions for adequate terminal care. Educational and self-study programs for care assistants in nursing homes appear to improve knowledge and attitudes regarding end-of-life care in dementia, and this knowledge appears to be maintained [49].

Some disease features of short-term death such as dementia were also significant determinants of quality of life [9]. As current research shows, there is a 30% higher mortality rate from cancer in patients with mental illness even though their incidence of cancer is no greater than in the general population [23]. In addition, some elderly patients with dementia have limited literacy and experience large disparities in health care access [50], while many primary care physicians lack competence in dementia care and access to valid assessment tools. Our predictive tool can assist care providers to address knowledge deficits and stratify at-risk patients with dementia for timely referral to specialist palliative care [51].

Cutting Back on Medical Futility

More than 15.5 million Americans with a history of cancer were alive on January 1, 2016. The Agency for Healthcare Research and Quality estimated that the direct medical cost (total of all health care expenditures) for cancer in the United States in 2014 was US \$87.8 billion [52]. The economic burden of cancer in the United States was substantial and expected to increase significantly in the future because of expected growth and aging of the population [53]. Consistent with the intensity of treatment for initial care, recurrence, and end-of-life care, costs of cancer were highest in the initial period following diagnosis and at the end-of-life stage [54]. In this study, seriously ill patients with cancer of the colon, blood (leukemia), stomach, and breast had relatively higher average cost in the final year of life, presenting as US \$34,485, \$32,514, \$22,388, and \$20,780, respectively, as well as high emergency department and inpatient resource use.

Medicalized deaths did not seem to be what cancer patients wanted, however. In a randomized controlled study, when patients with advanced cancer were given palliative care alongside standard treatment such as chemotherapy, the group receiving palliative care had lower rates of depression and were less likely to report pain [42]. In fact, increases in mortality incidence seen in the older patients (≥ 85 years) in the past may have been related to more aggressive diagnostic testing (eg, computed tomography imaging and stereotactic biopsy procedure) for this population [55], indicating potential overtreatment coupled with poor outcomes.

Although it was often assumed that dialysis will restore health, this was not always the case for some prevalent causes of death such as chronic renal disease, cardiovascular disease, and chronic respiratory disease. For example, despite improvements in survival among patients receiving maintenance dialysis over the past 2 decades, mortality rates in the end-stage renal disease population remained disturbingly high [56]. Older dialysis patients spend twice as many days in the hospital during the last month of life compared with Medicare beneficiaries with cancer. This indicates that when patients met dismal probability of survival and poor quality of life in the future, there was an opportunity to cut down the annual direct medical costs for end-stage renal disease, which are nearly US \$28.6 billion [56].

The futility and discomfort of aggressive treatments combined with the underrecognition and undertreatment of pain in patients with severe dementia support the use of palliative care for advanced dementia [57]. Further, limited use of antibiotics has not been associated with increased mortality, and aggressive treatment of infections has not been shown to alter underlying disease processes [58].

Our prognostic model helps address the problem of futility by identifying patients who are receiving aggressive intervention but may benefit from being referred to palliative care at an earlier time. Should therapy fail and the patient choose not to continue with treatment, early referral to palliative care may be a benefit as well. Information from the prognostic model may stimulate an open conversation and provide evidence of why treatment is judged to be medically inappropriate, promoting the synchronization between medical teams, patients, and families [59].

Limitations

This study has several limitations. First, compared to RCTs and other observational studies, the EMR-based study had real-life data challenges including missing or inaccurate values and sparse data. It is possible that some longitudinal clinical data were missing for certain patients in our EMR data warehouse, where the uncoded mortality cases could be outliers of the model and affect accuracy.

Second, the population aged 65 years and older in the United States has different distributions of race (white: 83.1% vs 80.7%; black: 9.1% vs 0.2%; Asian: 4.27% vs 13.5%) and comorbidities (cancer: 0.2% vs 0.6%; diabetes: 25.9% vs 4.7%) than the population of our study, which focused on older patients in Maine [60]. Recalibration and other necessary adjustments would be needed before leveraging the model validated with the population in the state of Maine to other regions of the United States. Using a cohort with race and comorbidity distribution similar to the United States, we shall build a more transferrable model.

Third, cost calculations were based on estimates from the literature rather than medical claims data. Using state cost averages that included items such as type of chemotherapy or laboratory costs, these estimates provided a justifiable approximation of the overall impact on 1-year cost.

Conclusions

Our prognostic model, prospectively validated for identifying elderly patients at risk for mortality, had a good predictive ability and generalized well among the elderly population (≥ 65 years) in the state of Maine. We identified statistically significant and clinically meaningful risk factors to help predict mortality and support clinical decision making by grouping high-risk patients based on clinical history. This tool should be a valuable adjunct for helping patients make better quality-of-life choices and alerting caregivers to target better interventions and counseling to individuals at high risk for mortality.

Acknowledgments

The authors acknowledge the hospitals, medical practices, physicians, and nurses participating in the HIE in the state of Maine and our biostatistics colleagues at the Department of Health Research and Policy for critical discussions.

The work was performed under a business arrangement between HealthInfoNet (HIN), operators of the Maine Health Information Exchange (HIE), and HBI Solutions Inc, located in California. The data are owned by the members of HIN. HIN is responsible for security and access to its members' data and has established data service agreements restricting unnecessary exposure of information. HIN and its board (composed of a cross section of its members) authorized the use of the deidentified data for this research. Access to the data used in the study requires a secure connection to HIN servers and should be requested directly from HIN. For any data request, researchers must contact Phil Prefenno (pprefenno@hinfonet.org) at HIN and follow the HIN compliance requirements to submit an application. HIN agrees to provide access to the deidentified data on a per-request basis to interested researchers.

HBI Solutions Inc is a private commercial company and several authors are employed by HBI. HBI provided funding in the form of salaries to the authors employed by HBI: TF, ML, MX, BJ, CZ, OW, LK, FS, and EW. HBI did not provide any funding or support outside of the salary support for the study and HBI did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of manuscript.

Authors' Contributions

KGS, EW, and XBL contributed to the conceptualization of this study. TF, MX, ML, BJ, CZ, OW, QW, and STA contributed to data curation, methodology, and software. EW, XBL, and FS administrated the project and supervised. YG, GZ, SH, and MX wrote the manuscript and carried out the data analysis with guidance from CY, LK, DBM and XBL. DBM, AB, KGS, and XBL contributed to the review and editing of the manuscript.

Conflicts of Interest

KGS, EW, and XBL are cofounders and equity holders of HBI Solutions Inc, which is currently developing predictive analytics solutions for health care organizations. TF, ML, MX, BJ, CZ, OW, FS, LK, and EW are employed by HBI. From the Stanford University School of Medicine, Stanford, California, KGS and XBL conducted this research as part of a personal outside consulting arrangement with HBI. The research and research results are not associated with Stanford University in any way. There are no patents, further products in development, or marketed products to declare.

Multimedia Appendix 1

Workflow of feature selection.

[[PDF File \(Adobe PDF File\), 131KB - jmir_v20i6e10311_app1.pdf](#)]

Multimedia Appendix 2

Percentage of deaths by diagnosis and age group in the state of Maine.

[[PNG File, 95KB - jmir_v20i6e10311_app2.png](#)]

Multimedia Appendix 3

Receiver operating characteristic curve and c-statistics of derivation cohort.

[[PNG File, 25KB - jmir_v20i6e10311_app3.png](#)]

Multimedia Appendix 4

Receiver operating characteristic curve and c-statistics of validation cohort.

[[PNG File, 26KB - jmir_v20i6e10311_app4.png](#)]

Multimedia Appendix 5

Survival curves for elderly patients in low-, intermediate-, and high-risk groups.

[[PNG File, 49KB - jmir_v20i6e10311_app5.png](#)]

Multimedia Appendix 6

Survival curves of mortality for high-risk patients with 4 main types of chronic diseases defined by the World Health Organization.

[[PNG File, 83KB](#) - [jmir_v20i6e10311_app6.png](#)]

Multimedia Appendix 7

Odds ratio plot of social determinants for death in high-risk and low-risk groups.

[[PNG File, 86KB](#) - [jmir_v20i6e10311_app7.png](#)]

Multimedia Appendix 8

Mortality and health care utilization among high-risk patients with and without dementia.

[[PDF File \(Adobe PDF File\), 13KB](#) - [jmir_v20i6e10311_app8.pdf](#)]

Multimedia Appendix 9

Total cost, average cost, and number of deaths among patients at high-risk for mortality (data in [Figures 3](#) and [4](#)).

[[PDF File \(Adobe PDF File\), 12KB](#) - [jmir_v20i6e10311_app9.pdf](#)]

Multimedia Appendix 10

Association between average cost and disease burden grouped by the top 20 chronic diseases among high-risk patients.

[[PNG File, 89KB](#) - [jmir_v20i6e10311_app10.png](#)]

Multimedia Appendix 11

Association between average cost and disease burden grouped by the top 20 chronic diseases among low-risk patients.

[[PNG File, 84KB](#) - [jmir_v20i6e10311_app11.png](#)]

Multimedia Appendix 12

Association between average number of inpatient admissions and disease burden grouped by the top 20 chronic diseases among high-risk patients.

[[PNG File, 81KB](#) - [jmir_v20i6e10311_app12.png](#)]

Multimedia Appendix 13

Association between average number of emergency department visits and disease burden grouped by the top 20 chronic diseases among high-risk patients.

[[PNG File, 89KB](#) - [jmir_v20i6e10311_app13.png](#)]

Multimedia Appendix 14

Details and performance of other clinical mortality prediction scores.

[[PDF File \(Adobe PDF File\), 14KB](#) - [jmir_v20i6e10311_app14.pdf](#)]

Multimedia Appendix 15

Risk stratification of derivation and validation cohorts.

[[PDF File \(Adobe PDF File\), 12KB](#) - [jmir_v20i6e10311_app15.pdf](#)]

Multimedia Appendix 16

Top risk feature importance (weights).

[[PDF File \(Adobe PDF File\), 13KB](#) - [jmir_v20i6e10311_app16.pdf](#)]

References

1. Wheatley VJ, Baker JJ. "Please, I want to go home": ethical issues raised when considering choice of place of care in palliative care. *Postgrad Med J* 2007 Oct;83(984):643-648 [[FREE Full text](#)] [doi: [10.1136/pgmj.2007.058487](https://doi.org/10.1136/pgmj.2007.058487)] [Medline: [17916873](https://pubmed.ncbi.nlm.nih.gov/17916873/)]

2. The Economist. 2017 Apr 29. How to have a better death URL: <https://www.economist.com/news/leaders/21721371-death-inevitable-bad-death-not-how-have-better-death> [WebCite Cache ID 6yuBu4xeB]
3. Scitovsky A. "The high cost of dying": what do the data show? *Milbank Q* 2005;83(4):825-841 [FREE Full text] [doi: [10.1111/j.1468-0009.2005.00402.x](https://doi.org/10.1111/j.1468-0009.2005.00402.x)] [Medline: [16279969](https://pubmed.ncbi.nlm.nih.gov/16279969/)]
4. Hogan C, Lunney J, Gabel J, Lynn J. Medicare beneficiaries' costs of care in the last year of life. *Health Aff (Millwood)* 2001;20(4):188-195. [doi: [10.1377/hlthaff.20.4.188](https://doi.org/10.1377/hlthaff.20.4.188)] [Medline: [11463076](https://pubmed.ncbi.nlm.nih.gov/11463076/)]
5. Cardona-Morrell M, Hillman K. Development of a tool for defining and identifying the dying patient in hospital: Criteria for Screening and Triaging to Appropriate Alternative Care (CriSTAL). *BMJ Support Palliat Care* 2015 Mar;5(1):78-90 [FREE Full text] [doi: [10.1136/bmjspcare-2014-000770](https://doi.org/10.1136/bmjspcare-2014-000770)] [Medline: [25613983](https://pubmed.ncbi.nlm.nih.gov/25613983/)]
6. Hillman K. Dying safely. *Int J Qual Health Care* 2010;22:339-340.
7. Higginson I, Sen-Gupta GJ. Place of care in advanced cancer: a qualitative systematic literature review of patient preferences. *J Palliat Med* 2000;3(3):287-300. [doi: [10.1089/jpm.2000.3.287](https://doi.org/10.1089/jpm.2000.3.287)] [Medline: [15859670](https://pubmed.ncbi.nlm.nih.gov/15859670/)]
8. Rosenwax L, McNamara BA, Murray K, McCabe RJ, Aoun SM, Currow DC. Hospital and emergency department use in the last year of life: a baseline for future modifications to end-of-life care. *Med J Aust* 2011 Jun 06;194(11):570-573. [Medline: [21644868](https://pubmed.ncbi.nlm.nih.gov/21644868/)]
9. Mitchell A, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. *Lancet Oncol* 2011 Feb;12(2):160-174. [doi: [10.1016/S1470-2045\(11\)70002-X](https://doi.org/10.1016/S1470-2045(11)70002-X)] [Medline: [21251875](https://pubmed.ncbi.nlm.nih.gov/21251875/)]
10. Bates B, Xie D, Kwong PL, Kurichi JE, Ripley DC, Stineman MG. One-year all-cause mortality after stroke: a prediction model. *PM R* 2014 Jun;6(6):473-483. [doi: [10.1016/j.pmrj.2013.11.006](https://doi.org/10.1016/j.pmrj.2013.11.006)] [Medline: [24211696](https://pubmed.ncbi.nlm.nih.gov/24211696/)]
11. Walter L, Brand RJ, Counsell SR, Palmer RM, Landefeld CS, Fortinsky RH, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. *JAMA* 2001 Jun 20;285(23):2987-2994. [Medline: [11410097](https://pubmed.ncbi.nlm.nih.gov/11410097/)]
12. Edwards FH, Cohen DJ, O'Brien SM, Peterson ED, Mack MJ, Shahian DM, Steering Committee of the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. Development and validation of a risk prediction model for in-hospital mortality after transcatheter aortic valve replacement. *JAMA Cardiol* 2016 Apr 01;1(1):46-52. [doi: [10.1001/jamacardio.2015.0326](https://doi.org/10.1001/jamacardio.2015.0326)] [Medline: [27437653](https://pubmed.ncbi.nlm.nih.gov/27437653/)]
13. Miao F, Cai Y, Zhang Y, Li Y, Zhang Y. Risk prediction of one-year mortality in patients with cardiac arrhythmias using random survival forest. *Comput Math Methods Med* 2015;2015:303250 [FREE Full text] [doi: [10.1155/2015/303250](https://doi.org/10.1155/2015/303250)] [Medline: [26379761](https://pubmed.ncbi.nlm.nih.gov/26379761/)]
14. Uematsu H, Kunisawa S, Sasaki N, Ikai H, Imanaka Y. Development of a risk-adjusted in-hospital mortality prediction model for community-acquired pneumonia: a retrospective analysis using a Japanese administrative database. *BMC Pulm Med* 2014 Dec 16;14:203 [FREE Full text] [doi: [10.1186/1471-2466-14-203](https://doi.org/10.1186/1471-2466-14-203)] [Medline: [25514976](https://pubmed.ncbi.nlm.nih.gov/25514976/)]
15. Slenter R, Sprooten RTM, Kotz D, Wesseling G, Wouters EFM, Rohde GGU. Predictors of 1-year mortality at hospital admission for acute exacerbations of chronic obstructive pulmonary disease. *Respiration* 2013;85(1):15-26 [FREE Full text] [doi: [10.1159/000342036](https://doi.org/10.1159/000342036)] [Medline: [23037178](https://pubmed.ncbi.nlm.nih.gov/23037178/)]
16. de Munter L, Polinder S, Lansink K, Cnossen M, Steyerberg E, de Jongh M. Mortality prediction models in the general trauma population: a systematic review. *Injury* 2017 Feb;48(2):221-229. [doi: [10.1016/j.injury.2016.12.009](https://doi.org/10.1016/j.injury.2016.12.009)] [Medline: [28011072](https://pubmed.ncbi.nlm.nih.gov/28011072/)]
17. Fonarow G, Adams K, Abraham W, Yancy C, Boscardin W, ADHERE Scientific Advisory Committee, Study Group, Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA* 2005 Feb 02;293(5):572-580. [doi: [10.1001/jama.293.5.572](https://doi.org/10.1001/jama.293.5.572)] [Medline: [15687312](https://pubmed.ncbi.nlm.nih.gov/15687312/)]
18. Bazemore A, Cottrell E, Gold R, Hughes L, Phillips R, Angier H, et al. "Community vital signs": incorporating geocoded social determinants into electronic records to promote patient and population health. *J Am Med Inform Assoc* 2016 Mar;23(2):407-412. [doi: [10.1093/jamia/ocv088](https://doi.org/10.1093/jamia/ocv088)] [Medline: [26174867](https://pubmed.ncbi.nlm.nih.gov/26174867/)]
19. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. *J Med Internet Res* 2018 Jan 30;20(1):e22 [FREE Full text] [doi: [10.2196/jmir.9268](https://doi.org/10.2196/jmir.9268)] [Medline: [29382633](https://pubmed.ncbi.nlm.nih.gov/29382633/)]
20. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based real-time case finding for the population health management of patients with diabetes mellitus: a prospective validation of the natural language processing-based algorithm with statewide electronic medical records. *JMIR Med Inform* 2016 Nov 11;4(4):e37 [FREE Full text] [doi: [10.2196/medinform.6328](https://doi.org/10.2196/medinform.6328)] [Medline: [27836816](https://pubmed.ncbi.nlm.nih.gov/27836816/)]
21. Hao S, Fu T, Wu Q, Jin B, Zhu C, Hu Z, et al. Estimating one-year risk of incident chronic kidney disease: retrospective development and validation study using electronic medical record data from the state of Maine. *JMIR Med Inform* 2017 Jul 26;5(3):e21 [FREE Full text] [doi: [10.2196/medinform.7954](https://doi.org/10.2196/medinform.7954)] [Medline: [28747298](https://pubmed.ncbi.nlm.nih.gov/28747298/)]
22. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 2012 Nov 01;28(21):2747-2754 [FREE Full text] [doi: [10.1093/bioinformatics/bts526](https://doi.org/10.1093/bioinformatics/bts526)] [Medline: [22942019](https://pubmed.ncbi.nlm.nih.gov/22942019/)]

23. Kisely S, Crowe E, Lawrence D. Cancer-related mortality in people with mental illness. *JAMA Psychiatry* 2013 Feb;70(2):209-217. [doi: [10.1001/jamapsychiatry.2013.278](https://doi.org/10.1001/jamapsychiatry.2013.278)] [Medline: [23247556](#)]
24. Wright AA, Zhang B, Ray A, Mack JW, Trice E, Balboni T, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. *JAMA* 2008 Oct 08;300(14):1665-1673 [FREE Full text] [doi: [10.1001/jama.300.14.1665](https://doi.org/10.1001/jama.300.14.1665)] [Medline: [18840840](#)]
25. Maine Center for Disease Control & Prevention. Deaths URL: <http://www.maine.gov/dhhs/mecdc/public-health-systems/data-research/vital-records/deaths.shtml> [WebCite Cache ID 6yrmHErrl]
26. Harrels, Frank. EHRs and RCTs: outcome prediction vs. optimal treatment selection. URL: <http://www.fharrell.com/post/ehrs-rcts/> [accessed 2018-04-22] [WebCite Cache ID 6yrm3XA0r]
27. Levine S, Sachs G, Jin L, Meltzer D. A prognostic model for 1-year mortality in older adults after hospital discharge. *Am J Med* 2007 May;120(5):455-460. [doi: [10.1016/j.amjmed.2006.09.021](https://doi.org/10.1016/j.amjmed.2006.09.021)] [Medline: [17466658](#)]
28. Ferrat E, Paillaud E, Laurent M, Caillet P, Tournigand C, Lagrange J, ELPACA Study Group. Predictors of 1-year mortality in a prospective cohort of elderly patients with cancer. *J Gerontol A Biol Sci Med Sci* 2015 Sep;70(9):1148-1155. [doi: [10.1093/gerona/glv025](https://doi.org/10.1093/gerona/glv025)] [Medline: [25834194](#)]
29. Wu C, Hannan E, Walford G, Ambrose J, Holmes D, King S, et al. A risk score to predict in-hospital mortality for percutaneous coronary interventions. *J Am Coll Cardiol* 2006 Feb 07;47(3):654-660 [FREE Full text] [doi: [10.1016/j.jacc.2005.09.071](https://doi.org/10.1016/j.jacc.2005.09.071)] [Medline: [16458151](#)]
30. Murray S, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. *BMJ* 2005 Apr 30;330(7498):1007-1011 [FREE Full text] [doi: [10.1136/bmj.330.7498.1007](https://doi.org/10.1136/bmj.330.7498.1007)] [Medline: [15860828](#)]
31. Goyal A, Spertus JA, Gosch K, Venkitachalam L, Jones PG, Kosiborod M. Serum potassium levels and mortality in acute myocardial infarction. *JAMA* 2012 Jan 11;307(2):157-164. [doi: [10.1001/jama.2011.1967](https://doi.org/10.1001/jama.2011.1967)] [Medline: [22235086](#)]
32. Todd S, Barr S, Roberts M, Passmore A. Survival in dementia and predictors of mortality: a review. *Int J Geriatr Psychiatry* 2013 Nov;28(11):1109-1124. [doi: [10.1002/gps.3946](https://doi.org/10.1002/gps.3946)] [Medline: [23526458](#)]
33. Kim S, Han H, Jung H, Kim K, Hwang DW, Kang S, et al. Multidimensional frailty score for the prediction of postoperative mortality risk. *JAMA Surg* 2014 Jul;149(7):633-640. [doi: [10.1001/jamasurg.2014.241](https://doi.org/10.1001/jamasurg.2014.241)] [Medline: [24804971](#)]
34. Kales H, Valenstein M, Kim H, McCarthy J, Ganoczy D, Cunningham F, et al. Mortality risk in patients with dementia treated with antipsychotics versus other psychiatric medications. *Am J Psychiatry* 2007 Oct;164(10):1568-1576. [doi: [10.1176/appi.ajp.2007.06101710](https://doi.org/10.1176/appi.ajp.2007.06101710)] [Medline: [17898349](#)]
35. De Hert M, Detraux J, van Winkel R, Yu W, Correll C. Metabolic and cardiovascular adverse effects associated with antipsychotic drugs. *Nat Rev Endocrinol* 2011 Oct 18;8(2):114-126. [doi: [10.1038/nrendo.2011.156](https://doi.org/10.1038/nrendo.2011.156)] [Medline: [22009159](#)]
36. Park Y, Franklin J, Schneeweiss S, Levin R, Crystal S, Gerhard T, et al. Antipsychotics and mortality: adjusting for mortality risk scores to address confounding by terminal illness. *J Am Geriatr Soc* 2015 Mar;63(3):516-523 [FREE Full text] [doi: [10.1111/jgs.13326](https://doi.org/10.1111/jgs.13326)] [Medline: [25752911](#)]
37. Weinstein L, Stefancic A, Cunningham A, Hurley K, Cabassa L, Wender R. Cancer screening, prevention, and treatment in people with mental illness. *CA Cancer J Clin* 2016;66(2):134-151 [FREE Full text] [doi: [10.3322/caac.21334](https://doi.org/10.3322/caac.21334)] [Medline: [26663383](#)]
38. Marmot M, Shipley MJ. Do socioeconomic differences in mortality persist after retirement? 25-year follow-up of civil servants from the first Whitehall study. *BMJ* 1996 Nov 09;313(7066):1177-1180 [FREE Full text] [Medline: [8916748](#)]
39. Van Lenthe F. Neighbourhood unemployment and all cause mortality: a comparison of six countries. *J Epidemiol Comm Health* 2005;59:231-237.
40. Galobardes B, Davey S, Jeffreys M, McCarron P. Childhood socioeconomic circumstances predict specific causes of death in adulthood: the Glasgow student cohort study. *J Epidemiol Community Health* 2006 Jun;60(6):527-529 [FREE Full text] [doi: [10.1136/jech.2005.044727](https://doi.org/10.1136/jech.2005.044727)] [Medline: [16698985](#)]
41. Cassel C, Field MJ. *Approaching Death: Improving Care at the End of Life*. Washington: National Academies Press; 1997.
42. Zimmermann C, Swami N, Krzyzanowska M, Hannon B, Leighl N, Oza A, et al. Early palliative care for patients with advanced cancer: a cluster-randomised controlled trial. *Lancet* 2014 May 17;383(9930):1721-1730. [doi: [10.1016/S0140-6736\(13\)62416-2](https://doi.org/10.1016/S0140-6736(13)62416-2)] [Medline: [24559581](#)]
43. Higginson I, Evans CJ. What is the evidence that palliative care teams improve outcomes for cancer patients and their families? *Cancer J* 2010;16(5):423-435. [doi: [10.1097/PPO.0b013e3181f684e5](https://doi.org/10.1097/PPO.0b013e3181f684e5)] [Medline: [20890138](#)]
44. Mokdad AH, Dwyer-Lindgren L, Fitzmaurice C, Stubbs RW, Bertozzi-Villa A, Morozoff C, et al. Trends and patterns of disparities in cancer mortality among US counties, 1980-2014. *JAMA* 2017 Dec 24;317(4):388-406 [FREE Full text] [doi: [10.1001/jama.2016.20324](https://doi.org/10.1001/jama.2016.20324)] [Medline: [28118455](#)]
45. Cleland J, Cohen-Solal A, Aguilar J, Dietz R, Eastaugh J, Follath F, IMPROVEMENT of Heart Failure Programme Committees, Investigators, Study Group on Diagnosis of the Working Group on Heart Failure of The European Society of Cardiology. Management of heart failure in primary care (the IMPROVEMENT of Heart Failure Programme): an international survey. *Lancet* 2002 Nov 23;360(9346):1631-1639. [Medline: [12457785](#)]
46. McAlister F, Stewart S, Ferrua S, McMurray J. Multidisciplinary strategies for the management of heart failure patients at high risk for admission: a systematic review of randomized trials. *J Am Coll Cardiol* 2004 Aug 18;44(4):810-819 [FREE Full text] [doi: [10.1016/j.jacc.2004.05.055](https://doi.org/10.1016/j.jacc.2004.05.055)] [Medline: [15312864](#)]

47. Lynn J, Adamson D. Living Well at the End of life. Adapting Health Care to Serious Chronic Illness in Old Age. Washington: RAND Health; 2003.
48. Davies E, Higginson I. Better palliative care for older people. 2004. URL: http://www.euro.who.int/_data/assets/pdf_file/0009/98235/E82933.pdf [accessed 2018-04-29] [WebCite Cache ID 6z3DxLF55]
49. Harris D. Forget me not: palliative care for people with dementia. *Postgrad Med J* 2007 Jun;83(980):362-366 [FREE Full text] [doi: [10.1136/pgmj.2006.052936](https://doi.org/10.1136/pgmj.2006.052936)] [Medline: [17551065](https://pubmed.ncbi.nlm.nih.gov/17551065/)]
50. Sudore R, Mehta K, Simonsick E, Harris T, Newman A, Satterfield S, et al. Limited literacy in older people and disparities in health and healthcare access. *J Am Geriatr Soc* 2006 May;54(5):770-776. [doi: [10.1111/j.1532-5415.2006.00691.x](https://doi.org/10.1111/j.1532-5415.2006.00691.x)] [Medline: [16696742](https://pubmed.ncbi.nlm.nih.gov/16696742/)]
51. Coventry P, Grande G, Richards D, Todd C. Prediction of appropriate timing of palliative care for older adults with non-malignant life-threatening disease: a systematic review. *Age Ageing* 2005 May;34(3):218-227. [doi: [10.1093/ageing/afi054](https://doi.org/10.1093/ageing/afi054)] [Medline: [15863407](https://pubmed.ncbi.nlm.nih.gov/15863407/)]
52. Cancer facts & figures 2017.: American Cancer Society URL: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf> [accessed 2018-04-29] [WebCite Cache ID 6z3EGdnUB]
53. Yabroff K, Lund J, Kepka D, Mariotto A. Economic burden of cancer in the United States: estimates, projections, and future research. *Cancer Epidemiol Biomarkers Prev* 2011 Oct;20(10):2006-2014 [FREE Full text] [doi: [10.1158/1055-9965.EPI-11-0650](https://doi.org/10.1158/1055-9965.EPI-11-0650)] [Medline: [21980008](https://pubmed.ncbi.nlm.nih.gov/21980008/)]
54. Yabroff K, Lamont E, Mariotto A, Warren J, Topor M, Meekins A, et al. Cost of care for elderly cancer patients in the United States. *J Natl Cancer Inst* 2008 May 07;100(9):630-641. [doi: [10.1093/jnci/djn103](https://doi.org/10.1093/jnci/djn103)] [Medline: [18445825](https://pubmed.ncbi.nlm.nih.gov/18445825/)]
55. Legler J. Brain and other central nervous system cancers: recent trends in incidence and mortality. *J Nat Cancer Inst* 1999;91:1382-1390.
56. 2014 Annual Data Report: Volume Two: End-Stage Renal Disease in the United States.: National Institute of Diabetes and Digestive and Kidney Diseases; 2014. URL: https://www.usrds.org/2014/download/v2_esrd_full_14.zip [accessed 2018-04-29]
57. Volicer L, Rheume Y, Brown J, Fabiszewski K, Brady R. Hospice approach to the treatment of patients with advanced dementia of the Alzheimer type. *JAMA* 1986;256(16):2210-2213. [Medline: [3761520](https://pubmed.ncbi.nlm.nih.gov/3761520/)]
58. Fabiszewski K, Volicer B, Volicer L. Effect of antibiotic treatment on outcome of fevers in institutionalized Alzheimer patients. *JAMA* 1990 Jun 20;263(23):3168-3172. [Medline: [1693407](https://pubmed.ncbi.nlm.nih.gov/1693407/)]
59. Aghabarary M, Dehghan N. Medical futility and its challenges: a review study. *J Med Ethics Hist Med* 2016;9:11 [FREE Full text] [Medline: [28050241](https://pubmed.ncbi.nlm.nih.gov/28050241/)]
60. Population bulletin: aging in the United States.: Population Research Bureau URL: <https://www.prb.org/wp-content/uploads/2016/01/aging-us-population-bulletin-1.pdf> [accessed 2018-04-24] [WebCite Cache ID 6yuBC0EVJ]

Abbreviations

COPD: chronic obstructive pulmonary disease
EMR: electronic medical record
HIE: Health Information Exchange
PMPY: per member per year
PPV: positive predictive value
RCT: randomized controlled trial
ROC: receiver operating characteristic
SDH: social determinants of health

Edited by G Eysenbach; submitted 08.03.18; peer-reviewed by C Kurz, K Spanier, K Goniewicz; comments to author 29.03.18; revised version received 24.04.18; accepted 26.04.18; published 04.06.18.

Please cite as:

Guo Y, Zheng G, Fu T, Hao S, Ye C, Zheng L, Liu M, Xia M, Jin B, Zhu C, Wang O, Wu Q, Culver DS, Alfreds ST, Stearns F, Kanov L, Bhatia A, Sylvester KG, Widen E, McElhinney DB, Ling XB
 Assessing Statewide All-Cause Future One-Year Mortality: Prospective Study With Implications for Quality of Life, Resource Utilization, and Medical Futility
J Med Internet Res 2018;20(6):e10311
 URL: <http://www.jmir.org/2018/6/e10311/>
 doi:[10.2196/10311](https://doi.org/10.2196/10311)
 PMID:[29866643](https://pubmed.ncbi.nlm.nih.gov/29866643/)

©Yanting Guo, Gang Zheng, Tianyun Fu, Shiyong Hao, Chengyin Ye, Le Zheng, Modi Liu, Minjie Xia, Bo Jin, Chunqing Zhu, Oliver Wang, Qian Wu, Devore S Culver, Shaun T Alfreds, Frank Stearns, Laura Kanov, Ajay Bhatia, Karl G Sylvester, Eric Widen, Doff B McElhinney, Xuefeng Bruce Ling. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 04.06.2018. 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.

Original Paper

Interactive Visual Displays for Interpreting the Results of Clinical Trials: Formative Evaluation With Case Vignettes

Jiantao Bian¹, MS; Charlene Weir^{1,2}, PhD, RN; Prasad Unni^{1,3}, MD, MS; Damian Borbolla¹, MD, MS; Thomas Reese¹, PharmD; Yik-Ki Jacob Wan¹, BS; Guilherme Del Fiol¹, MD, PhD

¹Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States

²George E. Whalen Veterans Affairs Medical Center, Salt Lake City, UT, United States

³Intelligent Medical Objects, Chicago, IL, United States

Corresponding Author:

Guilherme Del Fiol, MD, PhD

Department of Biomedical Informatics

University of Utah

421 Wakara Way, Suite 140

Salt Lake City, UT, 84108

United States

Phone: 1 801 213 4129

Fax: 1 801 581 4297

Email: guilherme.delfiol@utah.edu

Abstract

Background: At the point of care, evidence from randomized controlled trials (RCTs) is underutilized in helping clinicians meet their information needs.

Objective: To design interactive visual displays to help clinicians interpret and compare the results of relevant RCTs for the management of a specific patient, and to conduct a formative evaluation with physicians comparing interactive visual versus narrative displays.

Methods: We followed a user-centered and iterative design process succeeded by development of information display prototypes as a Web-based application. We then used a within-subjects design with 20 participants (8 attendings and 12 residents) to evaluate the *usability* and *problem-solving* impact of the information displays. We compared subjects' perceptions of the interactive visual displays versus narrative abstracts.

Results: The resulting interactive visual displays present RCT results side-by-side according to the Population, Intervention, Comparison, and Outcome (PICO) framework. Study participants completed 19 *usability* tasks in 3 to 11 seconds with a success rate of 78% to 100%. Participants favored the interactive visual displays over narrative abstracts according to *perceived efficiency*, *effectiveness*, *effort*, *user experience* and *preference* (all *P* values <.001).

Conclusions: When interpreting and applying RCT findings to case vignettes, physicians preferred interactive graphical and PICO-framework-based information displays that enable direct comparison of the results from multiple RCTs compared to the traditional narrative and study-centered format. Future studies should investigate the use of interactive visual displays to support clinical decision making in care settings and their effect on clinician and patient outcomes.

(*J Med Internet Res* 2018;20(6):e10507) doi:[10.2196/10507](https://doi.org/10.2196/10507)

KEYWORDS

clinical decision-making; clinician information needs; information display; information foraging theory; information seeking behavior

Introduction

Background

At the point of care, clinicians have many clinical questions that they are unable to answer with the best available evidence [1]. Unanswered questions are missed opportunities to improve patient care decisions and for just-in-time learning [2]. Primary literature resources (eg, PubMed) contain answers to most of these questions [3], but their use at the point of care is still limited due to barriers such as lack of time and significant cognitive effort imposed by the evidence search and interpretation process [4,5].

Abstracts in scientific manuscripts are typically presented according to the well-established “background, introduction, methods, results, discussion, conclusion” structure [6]. However, this study-centered structure may not be optimal to support clinicians’ patient-centered conceptual models. The average time for clinicians to look up clinical questions on PubMed ranges from 5 to 60 minutes [7]. In addition, clinicians report high levels of dissatisfaction with their information seeking experience [8,9]. Ultimately, clinicians’ challenges in consuming evidence from the primary literature may contribute to slowing the translation of scientific evidence [10].

Few studies have examined optimal methods for displaying the results of clinical research reports. Prior work regarding primary literature has focused on displaying systematic reviews and investigating different methods of displaying results across studies, such as short summaries [11–13], tables [14–19], and harvest plots [20]. One recent study examined a novel presentation of clinical trial reports that restructured the visualization into several panels (ie, study purpose, process model and data grid for viewing results, statistical methods, and result interpretations). Using this visualization, translational researchers spent less time understanding and interpreting the clinical trials but maintained the same accuracy [21]. In general, very few of those studies have used any theory to drive their work.

The purpose of this study was to investigate alternative display approaches to present relevant information from randomized controlled trials (RCTs) to support clinical decision-making. Overall, we hypothesized that interactive visual displays would reduce clinicians’ cognitive workload in interpreting RCTs compared with narrative RCT abstracts. Building on Slager et al’s exploratory study on static tabular displays [22], we employed *information foraging* theory [23] and information visualization techniques to design a high-fidelity prototype with interactive visual displays of RCT results. The information displays were designed to help clinicians rapidly review, synthesize, and compare the results of relevant RCTs for the treatment of a specific patient. In this study, we described the RCT information displays and addressed the following three research questions: (1) Is the interface usable? (2) Is there a difference in *perceived efficiency*, *effort*, *effectiveness*, *user experience*, and *preference* between interactive visual displays and narrative abstracts? and (3) Do clinicians’ *perceived user experience*, *efficacy*, *effort*, and *effectiveness* predict their *intention to use* interactive visual displays?

Theoretical Framework

The design of the information displays was based upon *information foraging* theory [23], Shneiderman’s information visualization principles [24], and the Population, Intervention, Comparison, and Outcome (PICO) framework [25]. Information foraging theory was initially proposed for Web designers [26]. Based on an analogy with animals’ foraging, *information foraging* theory indicates that information seekers use *information scent* (ie, cues indicating the existence of easily accessible and relevant information) to select *information patches* to explore maximizing the value (ie, *perceived utility of the information*) to cost (ie, time and effort required to explore the patch) ratio [23]. Within a certain patch, the concentration of relevant information can be increased through a process called *information patch enrichment* (eg, use of filters). Information foraging theory is grounded on the Holling Disc Equation, which equals the ratio of the total net amount of valuable information gained to the sum of the total amount of time spent between-patches and within-patches [27]. Shneiderman proposed an information visualization principle according to which information displays should first provide an *information overview*, with the ability to *zoom and filter*, and then retrieve *details on demand* [24].

RCTs are the highest-level evidence in evidence-based medicine [28]. The PICO elements are key components of the Consolidated Standards of Reporting Trials (CONSORT) statement for reporting the quality of RCTs [29]. In addition, the PICO elements have been identified in multiple studies as the critical elements of an RCT [30–32] and are almost ubiquitous within medical journal abstracts [31]. PICO has been reported to be a more effective search input format than the standard PubMed search interface for answering clinical questions [25,33]. More recently, Slager et al found that clinicians favored a PICO-based tabular display over the typical narrative abstracts reported in scientific journals [22].

Methods

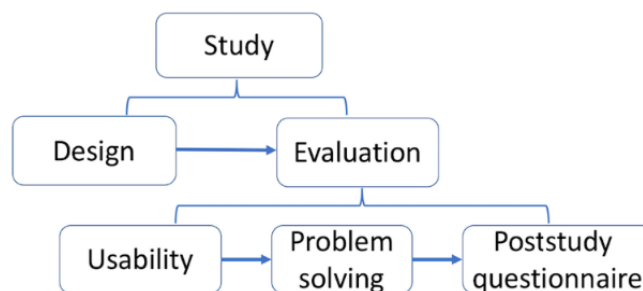
Overview

Our study had two phases. The first phase was the process of designing and implementing the information displays. The second was an experimental formative evaluation assessing *usability* and *problem-solving* impact. The second phase included three stages: (1) *usability* test of the interactive visual displays; (2) *problem solving* for two case vignettes comparing narrative abstracts versus interactive visual displays; and (3) a poststudy questionnaire (Figure 1).

Phase One: Design of the Interactive Visual Displays

Design

We followed a user-centered and iterative design process with feedback from informatics students, clinicians, and human factor experts. The study authors (three informatics and human factors specialists and four informatics students; five with clinical backgrounds) designed the first several iterations, followed by feedback from three independent informatics researchers with clinical backgrounds.

Figure 1. Visualization of study structure.

We started with *low-fidelity* prototypes that were designed with user interface mockup software (*NinjaMock*) [34] and a free website builder (Wix) [35]. After approximately 50 design-feedback iterations, the information displays evolved into a *high-fidelity* interactive prototype: a Web-based application implemented in HTML, CSS, JavaScript, and an open-source third party graphic library called *Highcharts* [36].

Data Structure

First, we searched a set of RCTs in PubMed related to three clinical case vignettes that were used in the evaluation phase. Next, we manually extracted data, including PICO elements, from each of the selected RCTs. The following data were extracted from each RCT on a spreadsheet: *PMID*, *journal*, *publication date*, *study title*, *study acronym*, *population inclusion criteria*, *population age*, *study sample size*, *study country*, *aim*, and *main conclusion*. For *study arms*, we extracted the *study arm intervention* and *number of participants*. For each *study arm*, we extracted *name* and *results* of all *major outcome measures*, *overall adverse event rate*, and *most common adverse events*. Last, we designed an Extensible Markup Language (XML) schema to represent the RCT data, created XML instances for each RCT, and transformed the XML instances into Java Script Object Notation format for consumption by the application. The RCTs used in the prototype were manually searched and selected, as the purpose of the study was to compare the information displays as opposed to search engines. RCT data were manually extracted since the purpose of the study was not to investigate automated methods to extract PICO elements.

Phase Two: Formative Evaluation Comparing Narrative Abstracts Versus Interactive Visual Displays

For formative evaluation, we used a within-subjects experimental design. The within-subjects design has advantages over a randomized, between-subjects design: it has higher statistical power, requiring a smaller sample size; and it enables participants to directly compare two designs. We tested the *usability* of the interactive visual displays and compared subjects' perceptions about narrative abstract displays versus interactive visual displays for clinical *problem solving* using case vignettes.

Study Setting

Formative evaluation sessions were conducted via online meetings using a Web meeting software (WebEx). Participants accessed an instance of the interactive visual displays hosted at the University of Utah Center for High Performance Computing

[37]. The Uniform Resource Locator (URL) was shared with participants at the beginning of the evaluation sessions.

Participant Recruitment

We recruited 20 participants (8 attending physicians and 12 residents) who had not previously been exposed to the interactive visual displays (see [Multimedia Appendix 1](#), Table A1). Participants were recruited from the Departments of Family Medicine at the University of Utah and Partners Healthcare via announcements that were sent to departmental email lists. We also employed the snowball sampling technique which asked study participants to promote the study among their colleagues. All participants received a US \$100 incentive to participate in the study sessions. A previous study with a similar design [22] demonstrated that 20 participants were enough to detect a moderate difference between interactive visual displays and narrative displays with a power of 0.80.

Information Displays Evaluation Procedure

We randomized participants to the order of presentation of the two tools and to the vignette-tool assignment. Each session began with a brief introduction about the study and a short one-page PDF tutorial explaining how to use the interactive visual displays. To ensure consistency, verbal instructions were read from a predefined script for each session. Each formative evaluation session included three stages: (1) *usability* test of the interactive visual displays; (2) *problem solving* for two case vignettes comparing narrative abstracts versus interactive visual displays; and (3) a poststudy questionnaire assessing the participant's perception of the *efficiency*, *effort*, *effectiveness*, *user experience*, and *preference* of interactive visual displays versus narrative abstracts. In the first two stages, the participants were asked to share their screens via WebEx, and their screen interactions were recorded for data analysis. In the third stage, participants were asked to stop screen-sharing while answering the poststudy questionnaire to ensure anonymity and minimize the Hawthorne effect. A waiver of written consent was approved by the Institutional Review Board of the University of Utah. Participants provided verbal consent before the study session.

Case Vignettes and RCTs

We prepared three case vignettes ([Table 1](#)), which presented challenges related to patient treatment. The vignettes were obtained from the literature and adapted by clinicians in our team. For each case vignette, we searched for potentially relevant RCTs using PubMed's Clinical Query treatment filter and a Medical Subject Heading (MeSH) term that matched the main disease of the case.

Table 1. Case vignettes used in the formative evaluation.

Case vignette	Stage Used	Complexity ^a	Designer	Number of studies
Acute coronary syndrome	<i>Usability</i>	Easy	Article author (PU ^b)	2
Rheumatoid arthritis	<i>Problem solving</i>	Complex	Adapted from Medscape [38]	10
Diabetes mellitus	<i>Problem solving</i>	Complex	Adapted from Hirsch et al [39]	10

^aThe complexity level of each vignette was determined by the number of factors involved in each treatment case.

^bPU refers to co-author Prasad Unni.

We manually screened the retrieved articles for RCTs on the diseases of interest and presented the same RCTs in the same order, both within PubMed and the interactive visual displays. The case vignettes and selected RCTs are available in [Multimedia Appendix 1](#) (case vignettes).

Stage One: Usability of Interactive Visual Displays

We developed 19 tasks to test the *usability* of the key information and features provided by interactive visual displays. Most of the tasks required participants to perform an action (eg, *highlight*, *access*, *switch*). We read the tasks aloud one-by-one to each participant and let them complete the tasks independently without assistance. We measured the *time spent* and *success* on each task. We also tape recorded the session and transcribed participants' comments.

Stage Two: Problem Solving

Participants were told that the rheumatoid arthritis and diabetes mellitus cases are relatively complex and that there are multiple reasonable treatment options for each case. We asked each participant to complete this stage in no more than 10 minutes in order to simulate the time pressure of a real patient visit [1,40]. We notified participants when there were 3 minutes and 1 minute left to finish the session. Within each case vignette session, participants could go back to the vignette description at any time.

For PubMed, participants were given a hyperlink that gave access to a *search results* page with the 10 RCTs in PubMed's default search results display format (see [Multimedia Appendix 1](#), Figure A1). No washout time was provided between the two case vignettes. At the end of the *problem-solving* stage, we asked each participant to provide a summary of the evidence they found and their decision about the treatment for the patient.

Stage Three: Participant Information-Seeking Experience Questionnaire

In this stage, we asked participants to complete an online REDCap [41] questionnaire regarding their information-seeking experiences with the tools. The questions (see [Multimedia Appendix 1](#), "Post Evaluation Survey") were adapted from the System Usability Scale [42], the National Aeronautics and Space Administration (NASA) Task Load Index (NASA-TLX) tool [43], and from Slager et al [22]. Two versions of the questionnaire were used, depending on which tool the participant was randomly assigned to use first. Each participant only needed to complete one survey.

The questionnaire started with items about participants' demographics, *experience with cases in the domain of the*

vignettes, and *experience with literature searching*. Next, using 17 Likert scale items, participants were asked to rate interactive visual displays versus narrative abstracts according to *perceived efficiency*, *effectiveness*, *effort*, *user experience*, and *preference*. The anchors for each question juxtaposed narrative abstracts in PubMed on one end and interactive visual displays on the other with the direction of the anchor randomized. For example, the hypothesis that there would be a difference in *perceived effectiveness* for users between interactive visual displays and narrative abstracts was assessed by four survey items: (1) *comprehend the meaning of the information presented well*, (2) *identify relevant information to understand the study*, (3) *effectively identify relevant RCTs from the search results*, and (4) *accomplish tasks with minimal frustration*. The results were constructed so that there were separate ratings for narrative abstracts and interactive visual displays for each question by centering the scores for both displays separately. Participants were then asked to rate the *intention to use* and *learnability* of using interactive visual displays on a 1 ("Strongly Disagree") to 9 ("Strongly Agree") scale. Scales were created for each of the constructs as the sum of the ratings given to each of the items in the construct. We reported internal reliability for the scales using Cronbach alpha ([Table 2](#)). Last, participants were asked to provide suggestions for improving interactive visual displays.

Data Analysis

We analyzed the *usability* results and the Likert scale items in *problem solving* to address the following questions below. We performed all statistical analyses using IBM SPSS Statistics Premium 24 [44].

Is the Interface Usable?

We conducted both qualitative and quantitative analyses to answer this question. We employed a qualitative analysis software (ATLAS.ti [45]) to code, categorize, and analyze users' verbalizations in the usability stage. To establish reliability for success and time measures, two authors (JB and DB) developed and tested coding protocols and employed the Cohen kappa and Pearson's correlation coefficient (PCC) to measure the interrater agreements. When both measurement metrics reached 0.80, we split the remaining sessions between coders in order to reduce the workload. For each task, we reported the mean, standard deviation, median, and range for the time spent and reported the success rate. We analyzed the correlation between the *experience with literature searching* and success rate of the usability tasks, and also categorized the open-ended comments and reported the descriptive statistics.

Table 2. Construct items and Cronbach alpha.

Construct and items	Cronbach alpha
Experience with cases in the domain of the vignettes	.793
Dealing with patients in the same clinical domain of the narrative abstracts case vignette	
Dealing with cases with similar clinical complexity as in the case presented in the narrative abstracts vignette	
Dealing with patients in the clinical domain of the interactive visual display case vignette	
Dealing with cases with similar clinical complexity as in the case presented in the interactive visual display vignette	
Experience with literature searching	.870
Experience in using computers for work activities	
Experience in using medical literature search tools in general (eg, PubMed, UpToDate)	
Experience in using PubMed for medical literature search	
Efficiency	.877
Scan the information quickly	
Quickly obtain the gist of the study findings	
Locate information rapidly	
Interpret individual RCT ^a results quickly	
Quickly compare the results of multiple RCTs	
Quickly determine study relevance for the case vignette	
Effectiveness	.921
Comprehend the meaning of the information presented well	
Identify relevant information to understand the study	
Effectively identify relevant RCTs from the search results	
Accomplish tasks with minimal frustration	
Effort	.823
Spend the least degree of mental effort	
Accomplish task effortlessly	
User experience	.921
Be satisfied with the presentation (ie, format of the display) of the information	
Easily use the user interface	
Enjoy exploring information	
Have fun seeking information to find answers	
Intention to use	.971
Help me with clinical decisions for specific patients	
Find evidence during patient consultations	
Find evidence after patient consultations	
Prepare for patient appointments	
Prepare for patient rounds	
Prepare for teaching	

^aRCT: randomized controlled trial.

Is There a Difference in Perceived Efficiency, Effort, Effectiveness, User Experience and Preference Between Interactive Visual Displays and Narrative Abstracts?

We employed the paired *t*-test to assess differences in ratings for each variable. We also assessed if there was a difference

between interactive visual displays and narrative abstracts after controlling for years of expertise, tool presentation order, clinical role, experience with literature searching, and experience with cases in the domain of the vignettes.

Do Clinicians' Perceived User Experience, Efficacy, Effort, and Effectiveness Predict Their Intention to Use Interactive Visual Displays?

To answer this question, we regressed intention to use on user experience, efficacy, effort, and effectiveness.

Results

Interactive Visual Displays

Following Shneiderman's principles, the interactive visual displays provided information *overviews* and *filters* with the option to retrieve *details on-demand*. These principles guided the design of each of the features in Table 3. These features are operationalized in one of five information displays, which are *article list*, *text summary*, *comparison table*, *efficacy graph*, and *side effects graph*. The displays can be launched for each case vignette by clicking on the "i" icons at our website [37]. Figures 2-4 depict the features listed in Table 3. A drop-down menu is used for switching between five information displays. *Article list* and *text summary* information displays aim to help users judge the relevance of an RCT based on the study patient

characteristics and the interventions under investigation. The *comparison table*, *efficacy graph*, and *side effects graph* information displays allow users to compare the results of relevant RCTs side-by-side. To avoid visual cluttering, we limited the maximum number of studies to *four* that can be displayed in the *comparison table*, *efficacy graph*, and *side effects graph* information displays.

Formative Evaluation

The formative evaluation results are structured according to the research questions.

The Interface is Usable

After two WebEx recording sessions (one resident and one attending) were coded and analyzed by DB and JB independently, the interrater agreement Cohen kappa (1.00) and PCC (0.92) were higher than the threshold established *a priori*, so we split the coding of the remaining recording sessions between two coders. Overall, the participants were able to solve each of the 19 *usability* tasks within a *median time* of 3 to 11 seconds, and a *success rate* of 78% to 100% (Table 4).

Table 3. Design principles that inspired each feature in the interactive visual displays.

Information display	Feature	Design principle
<i>Article list</i>	Information about study population and interventions	<i>Information scent</i>
<i>Article list</i>	Hyperlink to full abstract within PubMed	<i>Details on demand</i>
<i>Article list</i>	Ability to select specific, most relevant studies for further visualization	<i>Information patch enrichment, filter and zoom</i>
<i>Comparison table</i>	Population, Intervention, Comparison, and Outcome (PICO) table structure	<i>Information scent</i>
<i>Comparison table</i>	Hyperlink to full abstract within PubMed and hyperlink to efficacy and side effect graph	<i>Details on demand and zoom</i>
<i>Efficacy graph/side effects graph</i>	Ability to choose different outcome measures or side effects	<i>Information scent and zoom</i>

Figure 2. *Article list* table with trials on various treatments for diabetes mellitus. This display is the landing page of the information displays. It provides a table with the *title*, *patient population*, and *study arms* of each study. The goal is to allow clinicians to quickly scan each study and select relevant ones for further review.

DISPLAY FORMAT: Article list (MAIN MENU)

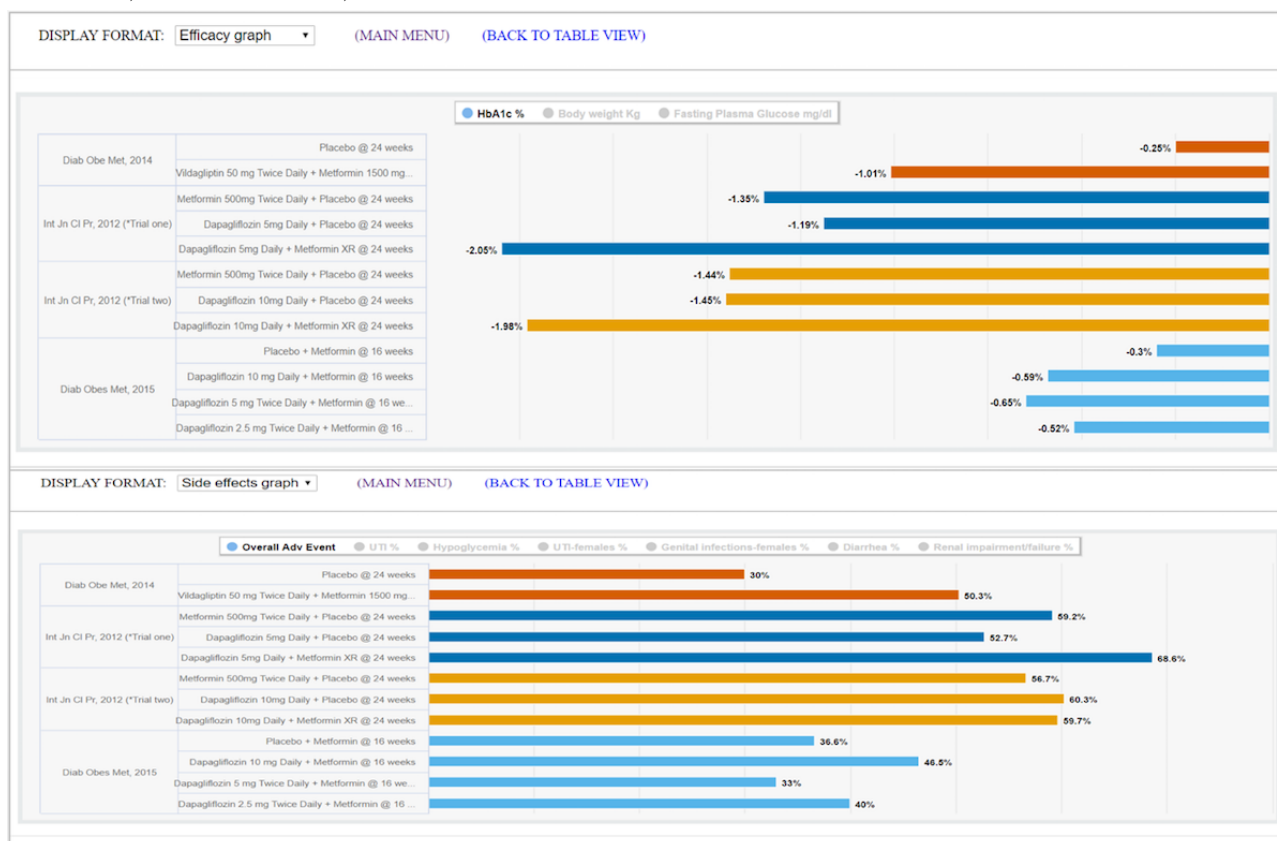
SELECT	TITLE	POPULATION	STUDY_ARMS
<input checked="" type="checkbox"/>	Efficacy and safety of vildagliptin in patients with type 2 diabetes mellitus inadequately controlled with dual combination of metformin and sulphonylurea.	Type 2 diabetes mellitus (DM2) on Metformin (1500 mg/day or more) plus Glimepiride (4mg/day or more)	1. Placebo 2. Vildagliptin 50 mg Twice Daily + Metformin 1500 mg/day or more + Glimepiride 4 mg/day or more
<input checked="" type="checkbox"/>	Dapagliflozin, Metformin XR, or both: initial pharmacotherapy for Type 2 Diabetes, a randomised control trial (*This study includes two trials)	DM2 with hemoglobin A1c (HbA1c) 7.5-12%	Trial One: 1. Metformin 500mg Twice Daily + Placebo 2. Dapagliflozin 5mg Daily + Placebo 3. Dapagliflozin 5mg Daily + Metformin XR Trial Two: 1. Metformin 500mg Twice Daily + Placebo 2. Dapagliflozin 10mg Daily + Placebo 3. Dapagliflozin 10mg Daily + Metformin XR
<input checked="" type="checkbox"/>	Twice-daily dapagliflozin co-administered with metformin in type 2 diabetes: a 16-week randomized, placebo-controlled clinical trial	DM2 on Metformin (1500mg/day or more) HbA1c: 6.7 - 10.5%	1. Placebo + Metformin 2. Dapagliflozin 10 mg Daily + Metformin 3. Dapagliflozin 5 mg Twice Daily + Metformin 4. Dapagliflozin 2.5 mg Twice Daily + Metformin
<input type="checkbox"/>	Efficacy and safety of liraglutide versus sitagliptin, both in combination with metformin, in Chinese patients with type 2 diabetes: a 26-week, open-label, randomised, active comparator clinical trial	DM2 on Metformin (1000mg/day or more) HbA1c: 7.0-10.0 %	1. Liraglutide (inj) 1.8 mg Daily + Metformin 2. Sitagliptin 100mg Daily + Metformin
<input type="checkbox"/>	Long-term safety and efficacy of empagliflozin, sitagliptin, and metformin: an active-controlled, parallel-group, randomized, 78-week open-label extension study in patients with type 2 diabetes.	DM2 who completed one of 12 week double blind randomized controlled trial (DBRCT) HbA1c: 7-10 %	1. Metformin 2. Empagliflozin 10 mg + Metformin 3. Empagliflozin 25 mg + Metformin 4. Sitagliptin + Metformin 5. Empagliflozin 10 mg

Figure 3. Comparison table display with four trials on various treatments for diabetes mellitus. This display contains key elements of selected studies in a tabular format according to the Population, Intervention, Comparison, Outcomes (PICO) framework [29-32]. Studies are displayed in columns, and attributes of studies are displayed in rows. Study results for primary outcomes and adverse events are represented in bar graphs [46]. Hovering over a bar brings up a callout with details on the intervention of the selected study arm. The scale of each measure is normalized across all studies to enable direct visual comparison. An illustration of the comparison table display for randomized controlled trials on rheumatoid arthritis is available in [Multimedia Appendix 1](#), Figure A2.

DISPLAY FORMAT: Comparison table ▾ (MAIN MENU)

	Efficacy and safety of vildagliptin in patients with type 2 diabetes mellitus inadequately controlled with dual combinat... [Diab Obe Met, 2014]	Dapagliflozin, Metformin XR, or both... [Int Jn CI Pr, 2012] (*Trial one of this study)	Dapagliflozin, Metformin XR, or both... [Int Jn CI Pr, 2012] (*Trial two of this study)	Twice-daily dapagliflozin co-administered with metformin in type 2 diabetes... [Diab Obes Met, 2015]
POPULATION				
Inclusion Criteria	Type 2 diabetes mellitus (DM2) on Metformin (1500 mg/day or more) plus Glimepiride (4mg/day or more)	DM2 with hemoglobin A1c (HbA1c) 7.5-12%	DM2 with HbA1c 7.5-12%	DM2 on Metformin (1500mg/day or more) HbA1c: 6.7 - 10.5%
Sample Size (completed/randomized (%))	299/318 (94%)	518/603 (86%)	552/641 (86%)	370/400 (93%)
INTERVENTION				
Arm 1	Placebo	Metformin 500mg Twice Daily + Placebo	Metformin 500mg Twice Daily + Placebo	Placebo + Metformin
Arm 2	Vildagliptin 50 mg Twice Daily + Metformin 1500 mg/day or more + Glimepiride 4 mg/day or more	Dapagliflozin 5mg Daily + Placebo	Dapagliflozin 10mg Daily + Placebo	Dapagliflozin 10 mg Daily + Metformin
Arm 3 more		Dapagliflozin 5mg Daily + Metformin XR	Dapagliflozin 10mg Daily + Metformin XR	Dapagliflozin 5 mg Twice Daily + Metformin
RESULTS (Efficacy Chart)				
HbA1c (%)				
FPG (mg/dl)				
Weight Change (kg)				
Conclusion	Vildagliptin significantly improved glycaemic control in patients with T2DM inadequately controlled with metformin plus glimepiride combination. more	In treatment-naïve patients with T2D, dapagliflozin plus metformin was generally well tolerated and effective in reducing HbA1c, FPG and weight. more	In treatment-naïve patients with T2D, dapagliflozin plus metformin was generally well tolerated and effective in reducing HbA1c, FPG and weight. more	Dapagliflozin 2.5 or 5 mg twice daily added to metformin was effective in reducing glycaemic levels in patients with type 2 diabetes inadequately controlled with metformin alone. more
ADVERSE EFFECTS (Side Effects Chart)				
Overall Adverse Effect (AE) (%)				

Figure 4. Efficacy graph display (top) and side effects graph display (bottom) with four trials on various treatments for diabetes mellitus. These two displays provide graphical comparisons of study primary outcomes and adverse effects respectively. Users can choose to set the bar graph for a specific outcome measure, overall adverse effects, or the most common adverse effect across all the arms of the selected studies.



Experience with literature searching was modestly correlated with success for the usability tasks ($r=0.417$, $P=.10$). A total of 14 out of 20 participants responded to the open-ended comments section, which we analyzed into categories. “Great tool” was the most frequent comment category (6 out of 14 participants), followed by “allow more than 4 studies for comparison” (5 out of 14 participants). Other less frequent comment categories included: request for more features, request for more information, request for clearer display, and prefer narrative abstracts in PubMed (see Multimedia Appendix 1, Table A2).

Clinicians Favored Interactive Visual Displays Over Narrative Abstracts on Perceived Efficiency, Effectiveness, Effort, User Experience, and Preference

The paired t -test results showed that clinicians favored interactive visual displays over narrative abstracts on all of the variables: efficiency $t_{(18)}=10.43$ (mean 7.86 vs 2.14, respectively), effectiveness $t_{(19)}=6.90$ (mean 7.36 vs 2.64), effort $t_{(19)}=8.24$ (mean 7.50 vs 2.50), user experience $t_{(19)}=7.94$ (mean 7.51 vs 2.49), and preference $t_{(19)}=8.62$ (mean 8.00 vs 2.00).

All differences were significant ($P<.001$). Figure 5 displays the comparison results.

In addition, participants’ years of expertise, tool presentation order, clinical role, experience with literature searching, and experience with cases in the domain of the vignettes were not correlated with any of the participants’ perception variables (all P values $>.05$; Multimedia Appendix 1, Table A3), which indicates that there is no need to control for these factors when comparing the difference between the two tools.

Do Clinicians’ Perceived User Experience, Efficacy, Effort, and Effectiveness Predict Their Intention to Use Interactive Visual Displays?

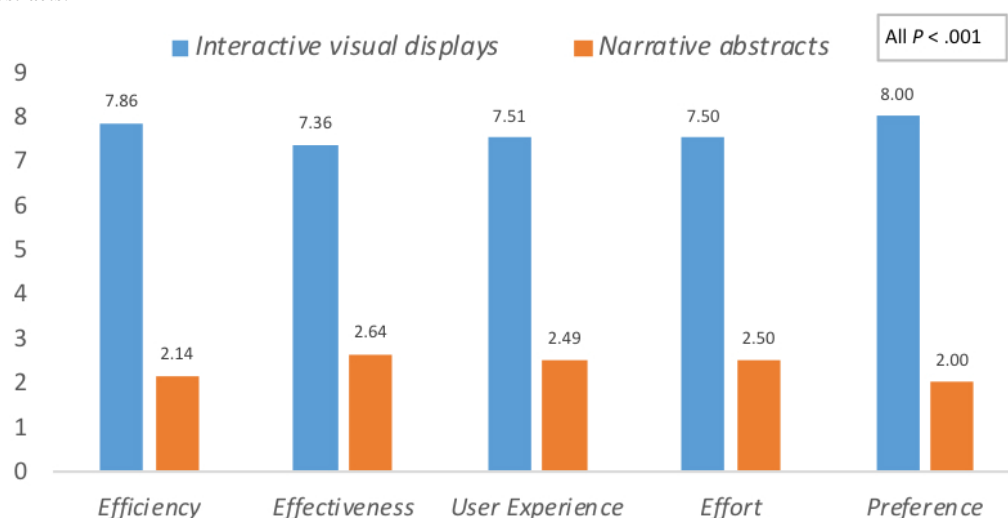
A scale for intention to use was created from six variables and had a Cronbach alpha of .971 (Table 2). We regressed intention to use on user experience, efficacy, effort, and effectiveness. The stepwise linear regression, which removes the variable with the highest beta weight sequentially, showed that efficiency was the only item that entered the prediction model ($R^2_{(16)}=0.661$, $t=5.59$, $F_{(16)}=31.26$ and $P<.001$) after controlling for all others.

Table 4. Time to completion (in seconds) and completion success rate for 19 usability tasks.

Usability task	Average time, seconds (SD)	Median time, seconds (range)	Success rate (%)
On the <i>Article list</i> format			
Highlight the study arms of the first study	7 (5)	5 (3-18)	83
Highlight the population of the second study	7 (14)	3 (1-59)	89
Access the PubMed abstract of the first study	7 (6)	4 (3-21)	82
This tool provides a textual summary of RCTs ^a . Please find out how to switch to the text summary of the two listed studies.	4 (1)	4 (2-6)	83
On the <i>Text summary</i> format			
What is the RCT publication journal and year of the first study?	7 (6)	5 (2-26)	100
Highlight the aim and conclusion of the second study	3 (1)	4 (1-6)	100
This tool also provides comparison views for multiple RCTs. Please switch to the comparison view for the two listed studies.	4 (3)	3 (1-13)	94
On the <i>Comparison table</i> format			
Highlight the study population of the first study	6 (2)	6 (2-10)	83
Highlight the study with the largest sample size	4 (3)	4 (1-14)	100
Highlight the research arms of the second study	5 (2)	5 (2-9)	94
Identify one of the study endpoints reported in both studies	8 (5)	7 (3-20)	83
Within trial 1, which <i>drug therapy</i> resulted in greatest total cholesterol reduction?	11 (10)	9 (3-44)	78
Which <i>drug therapy</i> across the two trials showed the best response in terms of total cholesterol reduction?	6 (4)	5 (2-17)	78
Which <i>drug therapy</i> across the two trials showed the best response in terms of high density lipoprotein increase?	12 (7)	11 (4-25)	100
Highlight the conclusion of the first trial	4 (5)	3 (1-23)	100
This tool also provides graphical visualization of RCTs. Please switch to the graphical view.	5 (3)	5 (2-12)	89
On the <i>Efficacy graph</i> format			
Set the graph to show LDL ^b outcomes	4 (2)	3 (1-8)	83
Which drug regimen across the two trials showed the greatest reduction in LDL?	4 (2)	3 (1-10)	100
Switch back to the main menu	3 (1)	3 (1-6)	100

^aRCT: randomized controlled trial.^bLDL: low density lipoprotein.

Figure 5. Mean differences for participants' *perceived efficiency, effectiveness, effort, user experience* and *preference* of interactive visual displays versus narrative abstracts.



Discussion

Summary

The goal of this work was to design a novel information display to help clinicians interpret, compare, and apply evidence from RCTs in clinical settings. We previously investigated a static structured PICO table for representing clinical trial reports and found that clinicians preferred a tabular PICO display over PubMed's default search results display [22]. We choose PubMed's default search results display as a baseline based on two reasons: (1) PubMed is the most widely used resource for browsing the biomedical literature, including RCT publications; and (2) PubMed is representative of resources in the same category (eg, Ovid, EBSCO, Scopus) since biomedical literature databases rely on the same narrative abstracts provided by biomedical journals. In this current study, we added graphical and interactive features to the previous static structured display and conducted a formative evaluation with 20 physicians in a simulation setting with case vignettes. Our results showed that when interpreting and applying research findings to patient care, physicians strongly preferred interactive visual displays that enable direct comparison of the results from multiple RCTs over narrative abstracts.

Information Processing Issues

Our findings suggest that the cognitive tasks involved in reviewing the literature are perhaps more complex than we had previously been aware. The tasks may involve a compilation of information processing goals (epistemic goals) that vary according to the clinical situation. Human information processing is essentially goal-oriented, so tailoring information to address specific goals is important [47]. Prior work in this area has found that task *problem-solving* is the most common information need in this context [48]. Our work suggests that displays that show adverse events, results by specific outcomes, and population descriptions by experimental arm match the information processing goals of clinicians seeking research information for medication decision-making.

In addition to exploring information processing goals, our results also suggest the need for further exploration of risky decision-making in work settings. One area that might be particularly fruitful is the well-established and robust findings from research in the "description-experience gap" [49]. This body of research has found large differences in decision-making between choices based on experience versus choices based on the provision of descriptive information. In general, physicians may weigh the probability of a loss (adverse events) and gain (treatment effectiveness) differently when being presented evidence rather than from their experience. Examining how displays can improve the accuracy of decision-making probability estimates is also an area of further research [49].

Information Foraging and information visualization principles (listed in Table 3) guided this study. Participants' preference for interactive visual displays can be attributed to the following reasons. First, interactive displays, as the central piece of visual analytics [50], provide clinicians with multiple advantages [51,52]. For example, with the interactive functions, only relevant information is presented up-front, and further details can be provided on demand. Second, the use of graphics reduces clinicians' cognitive effort when interpreting the results of multiple clinical trials [53]. In our displays, users can make direct comparisons both between and within clinical trials on the same display, thereby minimizing working memory overload [54]. Third, the PICO framework has been recommended to clinicians when formulating evidence-based clinical questions [25]. Therefore, our PICO tabular displays provide a consistent structure that is compatible with clinicians' mental models, facilitating their understanding of the gist of the evidence presented in RCTs [55].

Technology Adoption

According to the technology acceptance model proposed by Davis [56] and expanded on by the unified theory of acceptance and use of technology [57], *perceived ease-of-use* (PEOU), performance beliefs (ie, how well does it help me do my task), *perceived effort*, and social norms predict the actual use of a new technology. Findings from our *usability* study (PEOU) suggest that the prototype is easy to use. Most participants

completed the *usability* task correctly within a short period of time, with minimal training. In the clinical *problem-solving* session (performance beliefs and effort), participants' preference of the interactive visual displays was significantly higher than the narrative displays according to several perceived ability measurements. The performance of perceived ability measurements was not correlated with any of the clinicians' characteristics, suggesting that our finding is generalizable to a different range of users. The within-subjects design with randomized vignette assignment and tool presentation order minimized the impact of the participant's individual differences. We did not measure social norms. In sum, the interactive visual displays have the potential to ease clinicians' effort to interpret evidence from the primary literature at the point of care.

The stepwise regression analysis of the clinical *problem-solving* stage showed that *efficiency* was the only factor that predicted *intention to use*. Multicollinearity analysis also showed that only one dimension exists, which means that all predicting factors are correlated with each other. It is likely that *perceived efficiency* or *effectiveness* is the most general latent variable. It is possible that all of the poststudy questions measure the participants' general attitude towards the tool with little distinction among factors.

Implications for the Reporting of RCTs

Our study findings add to the growing evidence supporting alternative information display formats to convey the gist of clinical studies [11-22], suggesting that the standard format of scientific reporting, especially for article abstracts, is worth reconsidering. The ideal abstract display format should match clinicians' mental models to reduce cognitive workload in interpreting clinical study results. Much progress has been made with the increased adoption of structured abstracts, which are more readable, easier to search, preferred by readers, and easier to peer review than traditional unstructured abstracts [58,59]. Our findings suggest that interactive visual displays could further improve the presentation of summaries of clinical studies.

One important challenge in enabling interactive visual displays of clinical studies is the lack of a widely adopted standard data model for reporting study methods and results in a computable format. National clinical trial registries such as ClinicalTrials.gov have taken an important step towards the implementation of structured reporting. However, several challenges still exist, such as automatically extracting key study

data from clinical trial registries [60], incomplete linkage between clinical trial publications and clinical trial registration [61], and time delay between clinical trial publication and reporting of results in clinical trial registries [60]. Increasing requirements for structured reporting of clinical trials could be a possible solution. For example, core clinical journals could adopt and require structured reporting of clinical trial results using a common computable data model.

Limitations

This study has several limitations. First, we have not analyzed how much time participants spent looking at each component or piece of information in the information displays. Methods based on eye-tracking devices can be used in future studies to provide deeper insight into how users process the information presented on the screen. Second, the case vignettes did not have a single right or wrong answer, so it was impossible to measure the effect of the interactive visual displays on the accuracy of clinical decisions. Nevertheless, the vignettes were purposefully complex to stimulate a challenging information-seeking experience. Third, in this simulation study, we limited the information displays to 10 studies per case vignette. In real search sessions, the number of studies in a search result can be much higher.

Future Work

The RCT data under the interactive visual displays were manually extracted from a limited set of hand-selected RCTs. Future work is needed to automate the RCT data extraction process, leveraging resources such as ClinicalTrials.gov or RCT data extraction algorithms [60,62-64]. This work is underway, with a prototype currently available. Future studies should also implement the interactive visual displays in clinical settings and investigate their effect on clinicians' patient care decisions and clinical outcomes.

Conclusion

This study shows that when interpreting and applying research findings to patient care, physicians preferred graphical, interactive, and PICO-framework-based information displays that enable direct comparison of the results from multiple RCTs compared to the traditional narrative format of article abstracts. Future studies should investigate the use of these displays in clinical care settings and their effect on improving clinicians' patient care decisions and clinical outcomes.

Acknowledgments

This project was supported by National Library of Medicine grants 1R01LM011416-01 and T15LM007124, and National Cancer Institute Grant U24CA204800. We would also like to thank the Richard A Fay and Carol M Fay Endowed Graduate Fellowship for funding support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables and figures, case vignettes, and post evaluation survey.

[PDF File (Adobe PDF File), 704KB - [jmir_v20i6e10507_app1.pdf](#)]

References

1. Del Fiore G, Workman T, Gorman P. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern Med* 2014 May;174(5):710-718. [doi: [10.1001/jamainternmed.2014.368](#)] [Medline: [24663331](#)]
2. Leape L, Bates D, Cullen D, Cooper J, Demonaco H, Gallivan T, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA* 1995 Jul 05;274(1):35-43. [Medline: [7791256](#)]
3. Maviglia S, Yoon C, Bates D, Kuperman G. KnowledgeLink: impact of context-sensitive information retrieval on clinicians' information needs. *J Am Med Inform Assoc* 2006;13(1):67-73 [FREE Full text] [doi: [10.1197/jamia.M1861](#)] [Medline: [16221942](#)]
4. Ioannidis J. Why most published research findings are false. *PLoS Med* 2005 Aug;2(8):e124 [FREE Full text] [doi: [10.1371/journal.pmed.0020124](#)] [Medline: [16060722](#)]
5. Cook D, Sorensen K, Hersch W, Berger R, Wilkinson J. Features of effective medical knowledge resources to support point of care learning: a focus group study. *PLoS One* 2013;8(11):e80318 [FREE Full text] [doi: [10.1371/journal.pone.0080318](#)] [Medline: [24282535](#)]
6. Rennels G, Shortliffe E, Stockdale F, Miller P. A computational model of reasoning from the clinical literature. *Comput Methods Programs Biomed* 1987 Apr;24(2):139-149. [Medline: [3297479](#)]
7. Coumou H, Meijman F. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc* 2006 Jan;94(1):55-60 [FREE Full text] [Medline: [16404470](#)]
8. Hoogendam A, Stalenhoef A, Robbé PF, Overbeke A. Answers to questions posed during daily patient care are more likely to be answered by UpToDate than PubMed. *J Med Internet Res* 2008 Oct 03;10(4):e29 [FREE Full text] [doi: [10.2196/jmir.1012](#)] [Medline: [18926978](#)]
9. Thiele RH, Poirer NC, Scalzo DC, Nemergut EC. Speed, accuracy, and confidence in Google, Ovid, PubMed, and UpToDate: results of a randomised trial. *Postgrad Med J* 2010 Aug;86(1018):459-465. [doi: [10.1136/pgmj.2010.098053](#)] [Medline: [20709767](#)]
10. Balas E, Boren S. Managing clinical knowledge for health care improvement. *Yearb Med Inform* 2000(1):65-70. [Medline: [27699347](#)]
11. Perrier L, Persaud N, Thorpe K, Straus S. Using a systematic review in clinical decision making: a pilot parallel, randomized controlled trial. *Implement Sci* 2015 Aug 15;10:118 [FREE Full text] [doi: [10.1186/s13012-015-0303-4](#)] [Medline: [26276278](#)]
12. Perrier L, Kealey M, Straus S. An iterative evaluation of two shortened systematic review formats for clinicians: a focus group study. *J Am Med Inform Assoc* 2014 Oct;21(e2):e341-e346 [FREE Full text] [doi: [10.1136/amiainjnl-2014-002660](#)] [Medline: [24786378](#)]
13. Perrier L, Kealey M, Straus S. A usability study of two formats of a shortened systematic review for clinicians. *BMJ Open* 2014 Dec 23;4(12):e005919 [FREE Full text] [doi: [10.1136/bmjopen-2014-005919](#)] [Medline: [25537782](#)]
14. Guyatt G, Oxman A, Akl E, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011 Apr;64(4):383-394. [doi: [10.1016/j.jclinepi.2010.04.026](#)] [Medline: [21195583](#)]
15. Guyatt G, Oxman A, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: [10.1136/bmj.39489.470347.AD](#)] [Medline: [18436948](#)]
16. Carrasco-Labra A, Brignardello-Petersen R, Santesso N, Neumann I, Mustafa R, Mbuagbaw L, et al. Comparison between the standard and a new alternative format of the summary-of-findings tables in Cochrane review users: study protocol for a randomized controlled trial. *Trials* 2015 Apr 16;16:164 [FREE Full text] [doi: [10.1186/s13063-015-0649-6](#)] [Medline: [25873338](#)]
17. Langendam M, Akl E, Dahm P, Glasziou P, Guyatt G, Schünemann HJ. Assessing and presenting summaries of evidence in Cochrane Reviews. *Syst Rev* 2013 Sep 23;2:81 [FREE Full text] [doi: [10.1186/2046-4053-2-81](#)] [Medline: [24059250](#)]
18. Mustafa R, Wiercioch W, Santesso N, Cheung A, Prediger B, Baldeh T, et al. Decision-making about healthcare related tests and diagnostic strategies: user testing of GRADE evidence tables. *PLoS One* 2015;10(10):e0134553 [FREE Full text] [doi: [10.1371/journal.pone.0134553](#)] [Medline: [26474310](#)]
19. Welch V, Petticrew M, Petkovic J, Moher D, Waters E, White H, PRISMA-Equity Bellagio group. Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration. *J Clin Epidemiol* 2016 Feb;70:68-89 [FREE Full text] [doi: [10.1016/j.jclinepi.2015.09.001](#)] [Medline: [26348799](#)]
20. Crick K, Wingert A, Williams K, Fernandes R, Thomson D, Hartling L. An evaluation of harvest plots to display results of meta-analyses in overviews of reviews: a cross-sectional study. *BMC Med Res Methodol* 2015 Oct 26;15:91 [FREE Full text] [doi: [10.1186/s12874-015-0084-0](#)] [Medline: [26502717](#)]
21. Tong M, Hsu W, Taira R. Evaluating a novel summary visualization for clinical trial reports: a usability study. *AMIA Annu Symp Proc* 2016;2016:2007-2015 [FREE Full text] [Medline: [28269960](#)]

22. Slager S, Weir C, Kim H, Mostafa J, Del Fiol G. Physicians' perception of alternative displays of clinical research evidence for clinical decision support - a study with case vignettes. *J Biomed Inform* 2017 Jul;71S:S53-S59. [doi: [10.1016/j.jbi.2017.01.007](https://doi.org/10.1016/j.jbi.2017.01.007)] [Medline: [28089913](https://pubmed.ncbi.nlm.nih.gov/28089913/)]
23. Pirolli P, Card S. Information foraging. *Psychological Review* 1999;106(4):643-675. [doi: [10.1037/0033-295X.106.4.643](https://doi.org/10.1037/0033-295X.106.4.643)]
24. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. : IEEE Comput Soc Press; 1996 Sep 03 Presented at: Proceedings 1996 IEEE Symposium on Visual Languages; 3-6 Sept. 1996; Boulder, CO, USA p. 336-343. [doi: [10.1109/VL.1996.545307](https://doi.org/10.1109/VL.1996.545307)]
25. Schardt C, Adams M, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007 Jun 15;7:16 [FREE Full text] [doi: [10.1186/1472-6947-7-16](https://doi.org/10.1186/1472-6947-7-16)] [Medline: [17573961](https://pubmed.ncbi.nlm.nih.gov/17573961/)]
26. Huberman BA, Pirolli PL, Pitkow JE, Lukose RM. Strong regularities in world wide web surfing. *Science* 1998 Apr 03;280(5360):95-97 [FREE Full text] [Medline: [9525865](https://pubmed.ncbi.nlm.nih.gov/9525865/)]
27. Holling C. Some characteristics of simple types of predation and parasitism. *Can Entomol* 2012 May 31;91(07):385-398. [doi: [10.4039/ent91385-7](https://doi.org/10.4039/ent91385-7)]
28. Burns P, Rohrich R, Chung K. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg* 2011 Jul;128(1):305-310 [FREE Full text] [doi: [10.1097/PRS.0b013e318219c171](https://doi.org/10.1097/PRS.0b013e318219c171)] [Medline: [21701348](https://pubmed.ncbi.nlm.nih.gov/21701348/)]
29. Moher D, Schulz K, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001 Apr 14;357(9263):1191-1194. [Medline: [11323066](https://pubmed.ncbi.nlm.nih.gov/11323066/)]
30. Flaherty R. A simple method for evaluating the clinical literature. *Fam Pract Manag* 2004 May;11(5):47-52 [FREE Full text] [Medline: [15162611](https://pubmed.ncbi.nlm.nih.gov/15162611/)]
31. Dawes M, Pluye P, Shea L, Grad R, Greenberg A, Nie JY. The identification of clinically important elements within medical journal abstracts: Patient-Population-Problem, Exposure-Intervention, Comparison, Outcome, Duration and Results (PECODR). *Inform Prim Care* 2007;15(1):9-16 [FREE Full text] [Medline: [17612476](https://pubmed.ncbi.nlm.nih.gov/17612476/)]
32. Kim SN, Martinez D, Cavedon L, Yencken L. Automatic classification of sentences to support evidence based medicine. *BMC Bioinformatics* 2011 Mar 29;12 Suppl 2:S5 [FREE Full text] [doi: [10.1186/1471-2105-12-S2-S5](https://doi.org/10.1186/1471-2105-12-S2-S5)] [Medline: [21489224](https://pubmed.ncbi.nlm.nih.gov/21489224/)]
33. MEDLINE/PubMed via PICO. 2018. URL: <http://pubmedhh.nlm.nih.gov/nlmd/pico/piconew.php> [accessed 2013-02-15] [WebCite Cache ID 6ESdceen]
34. NinjaMock. 2018. URL: <https://ninjamock.com/s/XCTXS> [accessed 2018-03-21] [WebCite Cache ID 6y4kB0NWa]
35. Wix. Interactive visual displays in Wix. 2018. URL: <http://drprasad30.wixsite.com/ksviz> [accessed 2018-03-21] [WebCite Cache ID 6y4kKKIGK]
36. Highcharts API. 2018. URL: <http://www.highcharts.com/> [accessed 2012-11-05] [WebCite Cache ID 6Bxf3x2rw]
37. Interactive visual displays website. 2018. URL: <http://lite.bmi.utah.edu/rctcomp/index.html> [accessed 2018-03-21] [WebCite Cache ID 6y4kNHLKB]
38. Medscape. 2016. Options for patients with newly diagnosed rheumatoid arthritis who do not adequately respond to methotrexate URL: <http://www.medscape.org/viewarticle/574579> [WebCite Cache ID 6y4kUGy8K]
39. Hirsch I, Molitch M. Clinical decisions. Glycemic management in a patient with type 2 diabetes. *N Engl J Med* 2013 Oct 03;369(14):1370-1372. [doi: [10.1056/NEJMcldc1311497](https://doi.org/10.1056/NEJMcldc1311497)] [Medline: [24088099](https://pubmed.ncbi.nlm.nih.gov/24088099/)]
40. Ely J, Osherooff J, Ebell M, Bergus G, Levy B, Chambliss M, et al. Analysis of questions asked by family doctors regarding patient care. *BMJ* 1999 Aug 07;319(7206):358-361 [FREE Full text] [Medline: [10435959](https://pubmed.ncbi.nlm.nih.gov/10435959/)]
41. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009 Apr;42(2):377-381 [FREE Full text] [doi: [10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
42. Brooke J. SUS - a quick and dirty usability scale usability and context. 1996. URL: <https://hell.meiert.org/core/pdf/sus.pdf> [accessed 2018-03-21] [WebCite Cache ID 6y4kaByg6]
43. Hart S, Staveland L. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock P, Meshkati N, editors. *Human Mental Workload*. Amsterdam, Netherlands: Elsevier; 1988:139-183. [doi: [10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)]
44. IBM Corp. 2018. IBM SPSS Statistics for Windows, Version 24 URL: <http://www-01.ibm.com/support/docview.wss?uid=swg24041224> [accessed 2018-04-23] [WebCite Cache ID 6yu0z0eb8]
45. ATLAS.ti. 2018. URL: <https://atlasti.com/> [accessed 2018-06-06] [WebCite Cache ID 6zz4hfuyY]
46. Spence I, Lewandowsky S. Displaying proportions and percentages. *Appl Cognit Psychol* 1991 Jan;5(1):61-77. [doi: [10.1002/acp.2350050106](https://doi.org/10.1002/acp.2350050106)]
47. Xie I. Dimensions of tasks: influences on information - seeking and retrieving process. *J Doc* 2009 Apr 24;65(3):339-366. [doi: [10.1108/00220410910952384](https://doi.org/10.1108/00220410910952384)]
48. Agarwal N. Exploring context in information behavior: seeker, situation, surroundings, and shared identities. In: *Synthesis Lectures on Information Concepts, Retrieval, and Services*. San Rafael, CA: Morgan & Claypool Publishers; Dec 08, 2017:i-163. [doi: [10.2200/S00807ED1V01Y201710ICR061](https://doi.org/10.2200/S00807ED1V01Y201710ICR061)]
49. Lejarraga T, Pachur T, Frey R, Hertwig R. Decisions from experience: from monetary to medical gambles. *J Behav Dec Making* 2015 Apr 17;29(1):67-77. [doi: [10.1002/bdm.1877](https://doi.org/10.1002/bdm.1877)]

50. Thomas J, Cook K, editors. Illuminating the Path: The Research and Development Agenda for Visual Analytics. Richland, WA: National Visualization and Analytics Center; 2005.
51. Card S, Mackinlay J, Shneiderman B. Readings in Information Visualization: Using Vision to Think. San Francisco, CA: Morgan Kaufmann Publishers; Jan 1999.
52. Robertson G, Czerwinski M, Fisher D, Lee B. Selected human factors issues in information visualization. *Rev Hum Factors Ergon* 2009 Sep;5(1):41-81. [doi: [10.1518/155723409X448017](https://doi.org/10.1518/155723409X448017)]
53. Vekiri I. What is the value of graphical displays in learning? *Educ Psychol Rev* 2002;14(3):261-312. [doi: [10.1023/A:1016064429161](https://doi.org/10.1023/A:1016064429161)]
54. Mignini L, Champaneria R, Mishanina E, Khan K, EBM-CONNECT Collaboration. Graphical displays for effective reporting of evidence quality tables in research syntheses. *Reprod Health* 2016 Mar 09;13:21 [FREE Full text] [doi: [10.1186/s12978-016-0130-3](https://doi.org/10.1186/s12978-016-0130-3)] [Medline: [26957125](https://pubmed.ncbi.nlm.nih.gov/26957125/)]
55. Pirolli P. Information Foraging Theory: Adaptive Interaction with Information. Oxford, United Kingdom: Oxford University Press, Inc; 2007. [doi: [10.1093/acprof:oso/9780195173321.001.0001](https://doi.org/10.1093/acprof:oso/9780195173321.001.0001)]
56. Davis FDD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 1989 Sep;13(3):319. [doi: [10.2307/249008](https://doi.org/10.2307/249008)]
57. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. *MIS Quarterly* 2003;27(3):425. [doi: [10.2307/30036540](https://doi.org/10.2307/30036540)]
58. Hartley J. Current findings from research on structured abstracts. *J Med Libr Assoc* 2004 Jul;92(3):368-371 [FREE Full text] [Medline: [15243644](https://pubmed.ncbi.nlm.nih.gov/15243644/)]
59. Hartley J. Current findings from research on structured abstracts: an update. *J Med Libr Assoc* 2014 Jul;102(3):146-148 [FREE Full text] [doi: [10.3163/1536-5050.102.3.002](https://doi.org/10.3163/1536-5050.102.3.002)] [Medline: [25031553](https://pubmed.ncbi.nlm.nih.gov/25031553/)]
60. Kim H, Bian J, Mostafa J, Jonnalagadda S, Del Fiol G. Feasibility of extracting key elements from ClinicalTrials.gov to support clinicians' patient care decisions. *AMIA Annu Symp Proc* 2016;2016:705-714 [FREE Full text] [Medline: [28269867](https://pubmed.ncbi.nlm.nih.gov/28269867/)]
61. Huser V, Cimino JJ. Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials. *PLoS One* 2013;8(7):e68409 [FREE Full text] [doi: [10.1371/journal.pone.0068409](https://doi.org/10.1371/journal.pone.0068409)] [Medline: [23874614](https://pubmed.ncbi.nlm.nih.gov/23874614/)]
62. Hsu W, Speier W, Taira RK. Automated extraction of reported statistical analyses: towards a logical representation of clinical trial literature. *AMIA Annu Symp Proc* 2012;2012:350-359 [FREE Full text] [Medline: [23304305](https://pubmed.ncbi.nlm.nih.gov/23304305/)]
63. Summerscales R, Argamon S, Bai S, Hupert J, Schwartz A. Automatic summarization of results from clinical trials. : IEEE; 2011 Nov 12 Presented at: IEEE Int Conf Bioinforma Biomed Internet IEEE; 2011; Atlanta, GA, USA p. 372-377. [doi: [10.1109/BIBM.2011.72](https://doi.org/10.1109/BIBM.2011.72)]
64. Kang T, Zhang S, Tang Y, Hrubby GW, Rusanov A, Elhadad N, et al. EliIE: An open-source information extraction system for clinical trial eligibility criteria. *J Am Med Inform Assoc* 2017 Nov 01;24(6):1062-1071. [doi: [10.1093/jamia/ocx019](https://doi.org/10.1093/jamia/ocx019)] [Medline: [28379377](https://pubmed.ncbi.nlm.nih.gov/28379377/)]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
LDL: low density lipoprotein
MeSH: medical subject heading
NASA: National Aeronautics and Space Administration
PCC: Pearson's correlation coefficient
PEOU: perceived ease-of-use
PICO: Population, Intervention, Comparison, and Outcome
RCT: randomized controlled trial
URL: Uniform Resource Locator
XML: Extensible Markup Language

Edited by M Focsa; submitted 26.03.18; peer-reviewed by M Basit, Z Reis; comments to author 23.04.18; revised version received 30.04.18; accepted 03.05.18; published 25.06.18.

Please cite as:

Bian J, Weir C, Unni P, Borbolla D, Reese T, Wan YKJ, Del Fiol G
 Interactive Visual Displays for Interpreting the Results of Clinical Trials: Formative Evaluation With Case Vignettes
J Med Internet Res 2018;20(6):e10507
 URL: <http://www.jmir.org/2018/6/e10507/>
 doi: [10.2196/10507](https://doi.org/10.2196/10507)
 PMID: [29941416](https://pubmed.ncbi.nlm.nih.gov/29941416/)

©Jiantao Bian, Charlene Weir, Prasad Unni, Damian Borbolla, Thomas Reese, Yik-Ki Jacob Wan, Guilherme Del Fiol. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 25.06.2018. 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.

Original Paper

Misleading Claims About Tobacco Products in YouTube Videos: Experimental Effects of Misinformation on Unhealthy Attitudes

Dolores Albarracin¹, PhD; Daniel Romer², PhD; Christopher Jones², MD; Kathleen Hall Jamieson², PhD; Patrick Jamieson², PhD

¹University of Illinois at Urbana Champaign, Champaign, IL, United States

²Annenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:

Dolores Albarracin, PhD

University of Illinois at Urbana Champaign

603 E Daniel Street

Champaign, IL, 61822

United States

Phone: 1 2178402383

Email: dalbarra@illinois.edu

Abstract

Background: Recent content analyses of YouTube postings reveal a proliferation of user generated videos with misleading statements about the health consequences of various types of nontraditional tobacco use (eg, electronic cigarettes; e-cigarettes).

Objective: This research was aimed at obtaining evidence about the potential effects of YouTube postings about tobacco products on viewers' attitudes toward these products.

Methods: A sample of young adults recruited online (N=350) viewed one of four highly viewed YouTube videos containing misleading health statements about chewing tobacco, e-cigarettes, hookahs, and pipe smoking, as well as a control YouTube video unrelated to tobacco products.

Results: The videos about e-cigarettes and hookahs led to more positive attitudes toward the featured products than did control videos. However, these effects did not fully translate into attitudes toward combustible cigarette smoking, although the pipe video led to more positive attitudes toward combustible smoking than did the chewing and the hookah videos, and the e-cigarette video led to more positive attitudes toward combustible cigarette smoking than did the chewing video.

Conclusions: This research revealed young people's reactions to misleading claims about tobacco products featured in popular YouTube videos. Policy implications are discussed.

(*J Med Internet Res* 2018;20(6):e229) doi:[10.2196/jmir.9959](https://doi.org/10.2196/jmir.9959)

KEYWORDS

health communication; tobacco

Introduction

The study of media influences on smoking among adolescents and young adults has a long history of uncovering significant health threats [1,2]. Although these findings, along with regulatory efforts, have contributed to the decline of tobacco portrayal (mostly cigarettes) in cinema and on television since 1950 [3,4], emerging media such as the internet remain largely unregulated [5,6]. For example, many noncommercial internet materials generated by community members minimize or misrepresent the negative health consequences of tobacco use, either through omission (eg, not noting the negative health consequences [5,7-9]), or through commission (eg, asserting

that smoking is safe or even has health benefits [9-11]). The question guiding this research is:

What is the likely effect of such tobacco- friendly communications disseminated informally on the internet?

There is a rising suspicion that online exposure to user-generated content on YouTube shapes young people's perceptions of tobacco [5-15]. There are at least three reasons for this concern. First, there are large numbers of tobacco-related messages on YouTube, with more messages presenting favorable rather than unfavorable views on tobacco [6-10,12,13,15-17]. Second, YouTube reports over one billion users who collectively watch

hundreds of millions of hours of video per day [18]. Third, according to a recent survey, half of today's teens cite YouTube as their favorite website [19]. Despite the potential influence of this large number of regularly viewed messages from an outlet that young people trust, little systematic research has elucidated the degree to which YouTube messages influence attitudes toward tobacco products. Such evidence is critical for future policy decisions about tobacco-related content presented on the internet [11,12,20].

Prior studies of tobacco messages on YouTube have provided invaluable qualitative analyses of content [6,7,10,12,13,16] and determined prevalence of tobacco messages [21]. The next step, however, is to ascertain if these messages can promote favorable attitudes toward tobacco products such as electronic cigarettes (e-cigarettes) and combustible cigarette smoking in young viewers. Research on alcohol use portrayals in social media has already shown harmful influences of internet content [22-24]. In the arena of tobacco, the most likely targets of influence are products of ambiguous health consequences in the eyes of the public. Recent surveys suggest that young adults regularly use one or more tobacco products such as hookahs and e-cigarettes [25-27], even though combustible cigarette use has declined [3,27]. Thus, we identified popular user-generated YouTube videos that contained misleading messages about products such as chewing tobacco, hookahs, and e-cigarettes. We then experimentally examined whether these videos create favorable attitudes toward the featured product. We focused on four different tobacco products: (1) chewing tobacco, (2) e-cigarettes, (3) hookahs, and (4) pipe smoking. Four highly viewed messages were selected and presented online to a sample of 18-to-24-year olds (N=350) with varied prior use of tobacco products.

Methods

Sample and Experimental Design

Four hundred and thirty participants aged 18-24 years in the United States completed the 15-minute study via Amazon Mechanical Turk [28]. The study was approved by the Institutional Review Board of the University of Pennsylvania. Participants were compensated US \$1 for study completion. Researchers have found psychometric indicators of quality of Mechanical Turk data to be comparable to subject pools at research universities [29]. Nevertheless, we included two checks on participation quality [30]. To ensure that participants read the instructions, they were required to answer a question about their favorite color by clicking green and pink regardless of their actual preferences. Failing to follow this instruction indicates that this, and possibly other instructions of questions, were not read. A second check ensured that participants actually watched the videos by indicating what was discussed. Participants responded to a checklist of products including the tobacco product that appeared in the video and indicated which products appeared or were mentioned. Participants failing either check (80/430, 18.6%) were excluded, producing a final sample size of 350. Comparisons between the excluded and retained participants indicated no significant differences in tobacco consumption or demographics.

Selection of Videos

A search of popular videos on YouTube using 136 tobacco-related search terms identified over 8000 videos after removing unrelated content. Search terms included "smoke," "smoking," "tobacco," "cigarillo," and colloquial terms for products such as "shisha" for hookah tobacco. Criteria for exclusion included the video having fewer than 20,000 views at the time of download, the video being in a language other than English or having no audio, a video not containing tobacco content, or a video not being retrievable due to a broken or inactive link. Using an Excel random number generator, of the 8000 eligible videos, 200 were selected for further coding. Three coders met a Krippendorff alpha reliability of $K_{\alpha} > .91$ for the classifications of videos into different types of claims. This coding identified four major types of misleading health messages from this sample of YouTube videos: (1) rejection of science (ie, evidence supporting the harmfulness of a tobacco product is faulty), (2) assertion of benefit (eg, tobacco can be healthy), (3) denial of harm (eg, tobacco is not harmful after all), or (4) presence of acceptable risks (eg, using tobacco is no riskier than other common activities).

We pinpointed 37 videos containing misleading portrayals of tobacco's health consequences that lacked discernable brand affiliation or sponsorship. To ensure a varied sample of contents, we selected four videos representing each of the misleading categories. Within this selected set, two of the videos featured young adult white males and two featured adult white males. Within the selected set, videos were also representative of the major categories of misleading health claims that reached large audiences. In the selected videos, the source: (1) claimed that drinking green tea prevents mouth disease from chewing tobacco (denial of harm), (2) expressed skepticism toward scientific evidence that shisha contains harmful additives and that water filtration *does not* eliminate carcinogens (rejection of science), (3) suggested routine tasks like driving a car entail risks similar to pipe smoking (relative risk), or (4) asserted his status as a fitness expert while vaping (assertion of benefit). We selected videos that were popular without presenting expert sources. This selection allowed for the most stringent test of possible consequences of seemingly harmless amateur videos posted on YouTube. In addition to the lack of connection to a brand or sponsor, this selection also ensured a low probability of capturing commercial content.

The control video was not related to health. This YouTube video featured a demonstration about replacing a shower faucet and was similar in duration and features to the experimental videos. Videos were cropped to minimize background and edited into brief (approximately 20 seconds) segments. Screenshots and links to the videos appear in Figure 1.

Design

The conditions included five videos—four experimental videos and the control video—randomized between participants. Attitudes were measured after participants viewed one of the five videos. Specifically, in the experimental video conditions, we measured attitudes towards (1) chewing tobacco, (2) e-cigarettes, (3) hookahs, and (4) pipe smoking, depending on which was featured in the video. In the control video condition,

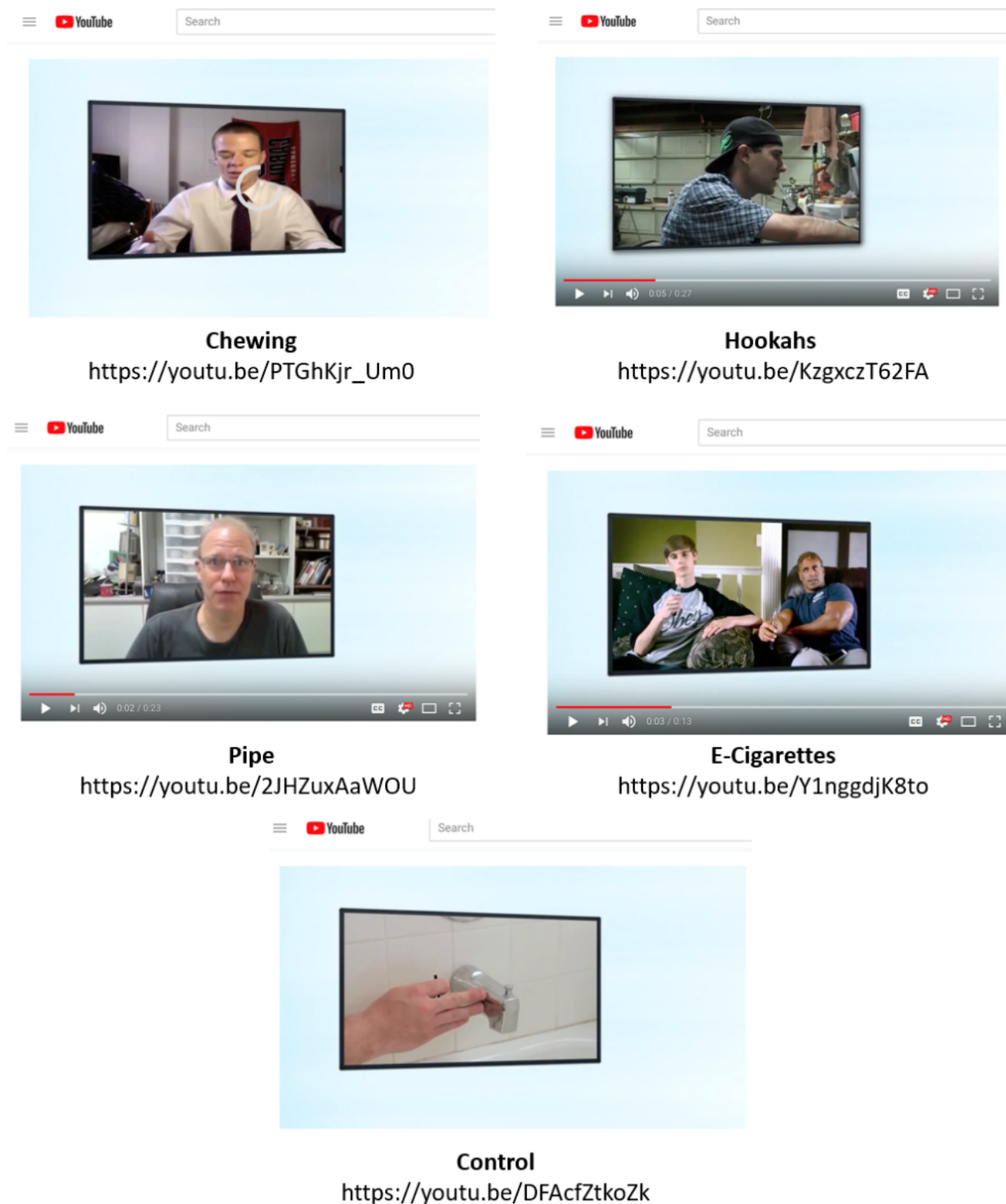
we measured attitudes towards (1) chewing tobacco, (2) e-cigarettes, (3) hookahs, and (4) pipe smoking in all cases. Attitudes toward combustible cigarette smoking were measured in all conditions even though combustible cigarettes were not featured.

Measures

Tobacco Use

We assessed use for the following products: cigarettes, cigars, pipe tobacco, chewing tobacco, hookah, and e-cigarettes. Participants first indicated which products they had ever tried. These measures were used to verify that the groups randomly assigned to conditions were indeed similar in experience.

Figure 1. YouTube videos.



Attitudes Toward Featured Tobacco Products

Attitudes toward tobacco use were measured with six semantic differential scales from 1 to 7, namely: *harmful/beneficial*, *wise/foolish*, *healthy/less healthy*, *enjoyable/unenjoyable*, *pleasant/unpleasant*, and *bad/good*. Participants in experimental conditions reported their attitudes toward the product mentioned in their video condition. Participants in control conditions reported their attitudes toward each of the five products mentioned in the experimental videos. Negative items were reverse scored. Attitude scales had good internal consistency ($\alpha > .70$) and were averaged to form overall indexes of attitudes toward the featured product in experimental conditions and toward each product in control conditions. In addition to the overall attitude scale, we created scales for only positively worded items, and scales for only negatively worded reverse-scored items. The attitudes towards chewing tobacco, pipes, hookahs, and e-cigarettes in the control condition were averaged to compare with the attitudes in experimental conditions.

Attitudes Toward Combustive Cigarette Smoking

Combustive cigarettes were not the focal product in any of the videos but attitudes toward combustive cigarette smoking may still be indirectly promoted in videos featuring nontraditional forms of tobacco use. Thus, participants in experimental and control conditions were asked to report their attitudes toward smoking combustive cigarettes on scales from 1 to 7 anchored on the following adjectives: *harmful/beneficial*, *wise/foolish* (*R*), *healthy/less healthy* (*R*), *enjoyable/unenjoyable*, *pleasant/unpleasant* (*R*), and *bad/good*. The overall attitude scales after reverse-scoring the negative items had good internal consistency ($\alpha > .70$) and were averaged to represent attitudes toward combustive cigarette smoking. In addition, we created a scale for only positively worded items, and another for only negatively worded reverse-scored items.

Credibility

Perceived credibility of the spokespersons with audio was measured with a 5-item scale with the following items: “To what extent do you agree that the person was speaking sincerely?”; “To what extent do you agree that the person is not worth listening to in the future?”; “To what extent do you agree that the person is a person who influenced my thinking on the matter?”; “To what extent do you agree that the person was communicating clearly?”; and “To what extent do you agree that the person is an expert on the topic?” These items had adequate internal consistency ($\alpha = .63$) and were averaged as a measure of credibility.

Results

Sample Description and Comparability of Conditions

For descriptive purposes, we assessed use for the following products: cigarettes, pipe tobacco, chewing tobacco, hookahs, and e-cigarettes. We specifically asked participants, “Have you

ever used any of the following tobacco products?”, after which they checked products that they had used. Table 1 contains the demographic and tobacco product usage description of the sample. The sample had considerable experience smoking combustive cigarettes, using hookahs, and using e-cigarettes but had little experience chewing tobacco and smoking pipes. As shown by the inferential statistics used to compare across experimental groups, there were no significant differences across the five video conditions in any of these characteristics. These analyses thus suggested that any experimental effects were due to the videos rather than a priori differences among groups of participants.

Level of Credibility of the Spokesperson in Experimental Videos

The level of credibility of the spokesperson in the experimental videos was low to moderate, as judged by a mean credibility of 2.52 (SD 0.56), which differed significantly from the midpoint (3) of the scale, $t(279) = -14.24$, $P < .001$. The means and 95% confidence intervals were 2.30 (2.18-2.43) for chewing tobacco, 2.50 (2.38-2.62) for hookahs, 2.88 (2.76-3.01) for pipes, and 2.41 (2.29-2.53) for e-cigarettes; all suggestive of the low to moderate credibility of amateur sources.

Message Effects on Attitudes

We estimated attitudes toward featured products across each specific experimental video and the control condition. As explained above, each experimental condition measured attitudes toward a different featured product, whereas the control condition measured attitudes towards all products, which were averaged for comparison. Table 2 presents these analyses; post hoc least significant difference (LSD) contrasts are represented with different subscripts. As shown, attitudes toward e-cigarettes and hookahs were more favorable following the experimental videos compared to the control video. In addition, the video about chewing tobacco produced more negative attitudes toward chewing tobacco than did the control video, and the video about pipe smoking did not differ from the control.

The effects on attitudes were investigated by conducting analyses of variance of attitudes toward the use of combustive cigarettes (never featured in the presented videos) as a function of experimental condition. Results from these analyses appear in the lower panel of Table 2 and show a significant effect of condition for both overall attitudes and positively worded attitude items. The omnibus effects on these attitude measures can be attributed to significant differences between the chewing and pipe videos, between the chewing and the e-cigarette videos, and between the hookah and the pipe videos. Although none of the videos differed significantly from the control videos, the pipe video produced the most favorable attitudes toward combustive cigarette smoking. Specifically, the pipe video led to more positive attitudes toward combustive smoking than did the chewing and the hookah videos, and the e-cigarette video led to more positive attitudes toward combustive cigarette smoking than did the chewing video.

Table 1. Descriptive statistics and comparisons across conditions.

Parameter	All	Chew	Hookah	Pipe	E-cigarettes	Control	Between-group statistic
Age (years), mean (SD)	21.98 (1.72)	22.15 (1.59)	22.04 (1.70)	22.02 (1.79)	21.80 (1.69)	21.87 (1.86)	0.47 ^{a,b}
Male, n (%)	154 (44)	161 (46)	151 (43)	165 (47)	161 (46)	144 (41)	0.69 ^{b,c}
Race, n (%)							19.37 ^{b,d}
White	256 (73)	207 (59)	273 (78)	259 (74)	280 (80)	224 (64)	
Black	42 (12)	56 (16)	35 (10)	49 (14)	11 (3)	67 (19)	
Native American	7 (2)	4 (1)	0 (0)	7 (2)	14 (4)	4 (1)	
Asian	32 (9)	39 (11)	21 (6)	39 (11)	35 (10)	35 (10)	
Other	14 (4)	11 (3)	25 (7)	7 (2)	11 (3)	21 (6)	
Hispanic origin, n (%)							2.87 ^{b,c}
Yes	42 (12)	35 (10)	14 (4)	28 (8)	32 (9)	39 (11)	
No	308 (88)	315 (90)	336 (96)	322 (92)	319 (91)	312 (89)	
Ever chew, n (%)	42 (12)	46 (13)	35 (10)	60 (17)	39 (11)	35 (10)	2.08 ^{b,c}
Ever hookah, n (%)	186 (53)	179 (51)	172 (49)	196 (56)	175 (50)	200 (57)	1.50 ^{b,c}
Ever pipe, n (%)	42 (12)	39 (11)	42 (12)	39 (11)	49 (14)	46 (13)	0.52 ^{b,c}
Ever e-cigarettes, n (%)	154 (44)	133 (38)	144 (41)	172 (49)	165 (47)	165 (47)	2.36 ^{b,c}
Ever cigarettes, n (%)	196 (56)	179 (51)	182 (52)	217 (62)	210 (60)	200 (57)	2.76 ^{b,c}

^a $F_{(4,345)}$.^b ns: not statistically significant.^c $\chi^2_{(1,4)}$.^d $\chi^2_{(1,16)}$.**Table 2.** Means (95% CI) for attitudes as a function of video.

Attitudes	Video conditions					F _(4,315)
	Chew	Hookah	Pipe	E-cigarette	Control	
Attitudes towards products featured in videos						
Overall	1.53 ^a (1.27-1.80)	2.99 ^b (2.73-3.25)	2.49 ^c (2.21-2.76)	3.21 ^d (2.94-3.47)	2.57 ^{b,c,e} (2.21-2.93)	23.12 ^f
Positive items	1.46 ^a (1.18-1.74)	3.07 ^b (2.79-3.45)	2.40 ^c (2.11-2.70)	3.22 ^d (2.94-3.51)	2.57 ^{b,c,e} (2.19-2.95)	23.42 ^f
Negatively worded items (reversed-scored)	1.60 ^a 1.32-1.68)	2.92 ^b (2.65-3.19)	2.57 ^c (2.28-2.85)	3.19 ^d (2.92-3.47)	2.45 ^{b,c,e} (2.08-2.82)	18.51 ^f
Attitudes towards smoking						
Overall	1.89 ^a (1.67-2.12)	2.04 ^{a,d} (1.82-2.26)	2.39 ^{b,d} (2.16-2.62)	2.23 ^{c,d} (2.00-2.46)	2.13 ^{a,d} (1.91-2.36)	2.59 ^g
Positively worded items	1.85 ^a (1.55-2.06)	1.95 ^{a,d} (1.70-2.20)	2.33 ^{b,d} (2.07-2.60)	2.23 ^{c,d} (1.97-2.49)	2.06 ^{a,d} (1.80-2.32)	2.58 ^g
Negatively worded items (reversed-scored)	1.99 ^a (1.75-2.24)	2.13 ^a (1.89-2.38)	2.45 ^b (2.19-2.71)	2.24 ^a (1.99-2.49)	2.21 ^a (1.96-2.46)	1.70

^{a-e} Within a row, different subscripts indicate statistically significant differences between cell means.^f $P < .001$.^g $P < .05$.

Discussion

Principal Findings

We examined the responses of young adults aged 18-24 years reacted to four misleading portrayals of tobacco's health consequences in popular YouTube videos. Results indicated that such material can increase positivity toward the featured products, such as e-cigarettes or hookah smoking (standardized difference Cohen g vs control video in each case was 0.38 and 0.37).

Although our study illustrates how potentially harmful content on social media may be studied, it has some limitations. We cannot generalize our findings to the many other videos that populate YouTube. We attempted to cover the major types of claims made in those videos, but there may well be others that are even more persuasive than the ones we identified. In addition, our findings with young adults may not generalize to adolescents, who may be even more susceptible to the claims made in these videos. Further research will be needed to assess this possibility.

This tendency for young tobacco consumers to respond credulously to misinformation on YouTube raises the possibility of tobacco use and exposure to misleading media exerting mutually reinforcing effects [31]. Finding tobacco-friendly material convincing and gratifying, recipients might seek out similar content, further bolstering self-justifying beliefs and prompting further selective exposure. In this regard, the aforementioned abundance of material on YouTube is cause for concern.

The present findings highlight the need to further study how new media sources such as YouTube affect tobacco knowledge, attitudes, and behavior. Although this study investigated misleading health portrayals only, content might also affect perceptions of tobacco products in other ways, both blatant and subtle, such as by modeling consumption [1-3], associating

products with sex [16], or facilitating product acquisition [20]. Furthermore, the participatory, interactive, and self-selected nature of social media may enhance pro-tobacco media effects more readily and perniciously than was possible with traditional media.

Policy Implications

The findings of YouTube video effects on positive attitudes toward hookahs and e-cigarettes should alert the public to the potential threats that these widely viewed videos can pose to youth and the health of the population. Attitudes toward hookahs and e-cigarettes can predict engagement in the behavior in the future [32-35], and essentially mimic the long-standing strategy of the tobacco industry to create favorable impressions of their products despite the harm they cause [36,37].

Difficulties Including User-Generated Postings Under the Total Ban on Tobacco Advertising, Promotion, and Sponsorship

The notion of regulating tobacco advertising is not new and stems from a large body of evidence on the powerful effects of the media on acceptance and use of tobacco. The Framework Convention on Tobacco Control organized by the World Health Organization has examined the media-effects evidence and recommended a total ban on tobacco advertising, promotion, and sponsorship [38]. The total ban is based on the principles listed in [Textbox 1](#). A quick inspection of these principles highlights the difficulties that regulating Web 2.0 practices would pose. Web 2.0 is defined by an online environment in which users share information and build networks of users [39,40]. Thus, many of the user-generated social media postings are probably developed by independent citizens more likely to be motivated by the goal of achieving fame than by payments from the tobacco industry. In our study, the videos we selected had no associations with either a brand or a company and appeared to be amateur, and were judged to have limited credibility. Therefore, a direct connection between these developers and the tobacco industry is unlikely.

Textbox 1. Principles underlying a total ban on tobacco advertising, promotion, and sponsorship [53].

1. It is well documented that tobacco advertising, promotion, and sponsorship increase tobacco use and that comprehensive bans on tobacco advertising, promotion, and sponsorship decrease tobacco use.
2. An effective ban on tobacco advertising, promotion, and sponsorship should, as recognized by Parties to the Convention in Articles 13.1 and 13.2, be comprehensive and applicable to all tobacco advertising, promotion, and sponsorship.
3. According to the definitions in Article 1 of the Convention, a comprehensive ban on all tobacco advertising, promotion, and sponsorship applies to all forms of commercial communication, recommendation, or action and all forms of contribution to any event, activity, or individual with the aim, effect, or likely effect of promoting a tobacco product or tobacco use either directly or indirectly.
4. A comprehensive ban on tobacco advertising, promotion, and sponsorship should include cross-border advertising, promotion, and sponsorship. This includes both outflowing advertising, promotion, and sponsorship (originating from a Party's territory) and in-flowing advertising, promotion, and sponsorship (entering a Party's territory).
5. To be effective, a comprehensive ban should address all persons or entities involved in the production, placement, and/or dissemination of tobacco advertising, promotion, and sponsorship.
6. Effective monitoring, enforcement, and sanctions supported and facilitated by strong public education and community awareness programs are essential for implementation of a comprehensive ban on tobacco advertising, promotion, and sponsorship.
7. Civil society has a central role in building support for, developing, and ensuring compliance with laws addressing tobacco advertising, promotion, and sponsorship, and it should be included as an active partner in this process.
8. Effective international cooperation is fundamental to the elimination of both domestic and cross-border tobacco advertising, promotion, and sponsorship.

Obstacles to Banning Internet Contents in the United States

Most countries impose some regulations on internet contents following the principle of protecting children from potentially harmful contents [41]. However, in the United States, restricting online interactions is perceived as violating freedom of speech and as, “throwing the baby with the bath water.” That is, restricting freedom of speech in any way may open the door to censorship, and in fact some of the countries that control internet contents are motivated by political reasons [42]. The Fifth Amendment would also be violated by imposing a total ban on internet tobacco contents. The Fifth Amendment protects the right to act in ways that are not self-incriminating and thus make all internet use private.

Voluntary Ban by YouTube

A more promising measure to reduce postings that promote tobacco use is for social media platforms to restrict postings. Verifying that amateur videos can still have measurable impacts on attitudes toward tobacco use should be a sufficient demonstration that responsible businesses must advocate for socially beneficial behavior in their premises. For example, in 2015, Facebook expanded the list of contents that the company is allowed to remove, which includes violent materials and postings that are degrading to specific social groups (eg, women) [43]. Following in Facebook's steps, in 2017 YouTube updated the guidelines of acceptable contents as follows:

Chiefly, the video site will not show advertising of “hateful” content that “promotes discrimination or disparages or humiliates an individual or group of people,” it said. Also barred from running ads are videos that involve “family entertainment characters” engaging in inappropriate behavior, and those that carry messages that demean or are incendiary. [44]

Moving from these guidelines to banning advertising of such products with no known benefits (eg, tobacco) would be easy and beneficial to society.

User Norms and Boycotting Sites

User boycott of sites that allow tobacco promotion would be another reasonable step. In 2016, several proposals were

suggested to boycott Facebook following the company's involvement in dissemination of fake news during the American presidential campaign [45]. These proposals were likely influenced by reports that 67% of Americans believed that Facebook should have done more to prevent the spread of fake news during the 2016 US presidential election on their site [46]. These public responses prompted the company's decision to limit fake news, although the success of these measures remains to be seen [47]. Even more relevant to our analysis, during the first half of 2017, YouTube faced a boycott from advertisers who refused to have their ads displayed next to the many hate speech videos that populate the site [48]. The boycott was effective at influencing their 2017 posting policies, which we covered above.

Electronic Filtering or Control Methods

Content-limited (or filtered) internet service providers allow subscribers to opt into specific websites or set mandatory restrictions for all subscribers. These are the most extreme forms of filtering and enable government, organizational, or parental control over the contents viewed by subscribers [49]. Less dramatic methods involve network-based filtering, in which software is installed to control content within a network such as a home or school [50]. Filtering can also be done by domain, by user, or by a combination of user and domain, all of which may be implemented for tobacco-related materials [50].

Debunking and Recipient Training Approaches

Future research should also explore ways to debunk the misleading claims found in YouTube videos to render these materials ineffective. A recent meta-analysis of debunking messages [51,52] suggested that detailed forms of debunking, as well as an active audience, are key to effective corrections [53]. Furthermore, resistance training has been identified as key to reducing susceptibility to peer influences on smoking [54]. In the context of our findings, young adults should be trained in identifying and resisting contents that appear on social media, but share many of the characteristics of peer pressure to use unhealthy products.

Conflicts of Interest

None declared.

References

1. Sargent JD, Beach ML, Adachi-Mejia AM, Gibson JJ, Titus-Ernstoff LT, Carusi CP, et al. Exposure to movie smoking: its relation to smoking initiation among US adolescents. *Pediatrics* 2005 Nov;116(5):1183-1191. [doi: [10.1542/peds.2005-0714](https://doi.org/10.1542/peds.2005-0714)] [Medline: [16264007](https://pubmed.ncbi.nlm.nih.gov/16264007/)]
2. National Cancer Institute. The role of the media in promoting and reducing tobacco use. Executive summary. *New Directions for Youth Development* 2010 Jul 20;2010(126):7-12. [doi: [10.1002/nd.346](https://doi.org/10.1002/nd.346)]
3. Jamieson PE, Romer D. Trends in US movie tobacco portrayal since 1950: a historical analysis. *Tob Control* 2010 Jun;19(3):179-184. [doi: [10.1136/tc.2009.034736](https://doi.org/10.1136/tc.2009.034736)] [Medline: [20395408](https://pubmed.ncbi.nlm.nih.gov/20395408/)]
4. Jamieson PE, Romer D. Portrayal of tobacco use in prime-time TV dramas: trends and associations with adult cigarette consumption--USA, 1955-2010. *Tob Control* 2015 May;24(3):243-248. [doi: [10.1136/tobaccocontrol-2012-050896](https://doi.org/10.1136/tobaccocontrol-2012-050896)] [Medline: [24699183](https://pubmed.ncbi.nlm.nih.gov/24699183/)]

5. Hong T, Cody MJ. Presence of pro-tobacco messages on the Web. *J Health Commun* 2002 Jul;7(4):273-307. [doi: [10.1080/10810730290088148](https://doi.org/10.1080/10810730290088148)] [Medline: [12356288](#)]
6. Elkin L, Thomson G, Wilson N. Connecting world youth with tobacco brands: YouTube and the Internet policy vacuum on Web 2.0. *Tob Control* 2010 Oct;19(5):361-366. [doi: [10.1136/tc.2010.035949](https://doi.org/10.1136/tc.2010.035949)] [Medline: [20739706](#)]
7. Bromberg JE, Augustson EM, Backinger CL. Portrayal of smokeless tobacco in YouTube videos. *Nicotine Tob Res* 2012 Apr;14(4):455-462 [FREE Full text] [doi: [10.1093/ntr/ntr235](https://doi.org/10.1093/ntr/ntr235)] [Medline: [22080585](#)]
8. Richardson A, Vallone DM. YouTube: a promotional vehicle for little cigars and cigarillos? *Tob Control* 2014 Jan;23(1):21-26. [doi: [10.1136/tobaccocontrol-2012-050562](https://doi.org/10.1136/tobaccocontrol-2012-050562)] [Medline: [23047887](#)]
9. Seidenberg AB, Rodgers EJ, Rees VW, Connolly GN. Youth access, creation, and content of smokeless tobacco (“dip”) videos in social media. *J Adolesc Health* 2012 Apr;50(4):334-338. [doi: [10.1016/j.jadohealth.2011.09.003](https://doi.org/10.1016/j.jadohealth.2011.09.003)] [Medline: [22443835](#)]
10. Luo C, Zheng X, Zeng DD, Leischow S, Cui K, Zhang Z, et al. Portrayal of electronic cigarettes on YouTube. In: Zeng DD, editor. *Smart Health. ICSH 2013. Lecture Notes in Computer Science*. Berlin: Springer; 2013.
11. Freeman B. New media and tobacco control. *Tob Control* 2012 Mar;21(2):139-144. [doi: [10.1136/tobaccocontrol-2011-050193](https://doi.org/10.1136/tobaccocontrol-2011-050193)] [Medline: [22345236](#)]
12. Carroll MV, Shensa A, Primack BA. A comparison of cigarette- and hookah-related videos on YouTube. *Tob Control* 2013 Sep;22(5):319-323 [FREE Full text] [doi: [10.1136/tobaccocontrol-2011-050253](https://doi.org/10.1136/tobaccocontrol-2011-050253)] [Medline: [22363069](#)]
13. Forsyth SR, Malone RE. “I’ll be your cigarette—light me up and get on with it”: examining smoking imagery on YouTube. *Nicotine Tob Res* 2010 Aug;12(8):810-816 [FREE Full text] [doi: [10.1093/ntr/ntq101](https://doi.org/10.1093/ntr/ntq101)] [Medline: [20634267](#)]
14. Peters K, Kashima Y. A multimodal theory of affect diffusion. *Psychol Bull* 2015 Sep;141(5):966-992. [doi: [10.1037/bul0000020](https://doi.org/10.1037/bul0000020)] [Medline: [26011791](#)]
15. Paek H, Kim S, Hove T, Huh JY. Reduced harm or another gateway to smoking? Source, message, and information characteristics of e-cigarette videos on YouTube. *J Health Commun* 2014 Oct;19(5):545-560. [doi: [10.1080/10810730.2013.821560](https://doi.org/10.1080/10810730.2013.821560)] [Medline: [24117370](#)]
16. Kim K, Paek H, Lynn J. A content analysis of smoking fetish videos on YouTube: regulatory implications for tobacco control. *Health Commun* 2010 Mar;25(2):97-106. [doi: [10.1080/10410230903544415](https://doi.org/10.1080/10410230903544415)] [Medline: [20390676](#)]
17. Freeman B, Chapman S. Is “YouTube” telling or selling you something? Tobacco content on the YouTube video-sharing website. *Tob Control* 2007 Jun;16(3):207-210 [FREE Full text] [doi: [10.1136/tc.2007.020024](https://doi.org/10.1136/tc.2007.020024)] [Medline: [17565142](#)]
18. YouTube Press Pages. 2018. URL: <http://www.youtube.com/yt/press/en-GB/statistics.html> [accessed 2018-05-30] [WebCite Cache ID 6zo4148C5]
19. Wasserman T. Mashable. 2013 Nov 05. Report says YouTube overtakes Facebook among teens URL: http://mashable.com/2013/11/05/teens-facebook-youtube-most-popular/#qCZ0q_GYXakL [accessed 2018-05-30] [WebCite Cache ID 6zoAQQtIT]
20. Ribisl KM, Jo C. Tobacco control is losing ground in the Web 2.0 era: invited commentary. *Tob Control* 2012 Mar;21(2):145-146. [doi: [10.1136/tobaccocontrol-2011-050360](https://doi.org/10.1136/tobaccocontrol-2011-050360)] [Medline: [22345237](#)]
21. Hua M, Yip H, Talbot P. Mining data on usage of electronic nicotine delivery systems (ENDS) from YouTube videos. *Tob Control* 2013 Mar;22(2):103-106. [doi: [10.1136/tobaccocontrol-2011-050226](https://doi.org/10.1136/tobaccocontrol-2011-050226)] [Medline: [22116832](#)]
22. Litt DM, Stock ML. Adolescent alcohol-related risk cognitions: the roles of social norms and social networking sites. *Psychol Addict Behav* 2011 Dec;25(4):708-713. [doi: [10.1037/a0024226](https://doi.org/10.1037/a0024226)] [Medline: [21644803](#)]
23. Moreno MA, Christakis DA, Egan KG, Brockman LN, Becker T. Associations between displayed alcohol references on Facebook and problem drinking among college students. *Arch Pediatr Adolesc Med* 2012 Feb;166(2):157-163 [FREE Full text] [doi: [10.1001/archpediatrics.2011.180](https://doi.org/10.1001/archpediatrics.2011.180)] [Medline: [21969360](#)]
24. Romer D, Moreno M. Digital media and risks for adolescent substance abuse and problematic gambling. *Pediatrics* 2017 Nov;140(Suppl 2):S102-S106. [doi: [10.1542/peds.2016-1758L](https://doi.org/10.1542/peds.2016-1758L)] [Medline: [29093042](#)]
25. Jamal A, Gentzke A, Hu SS, Cullen KA, Apelberg BJ, Homa DM, et al. Tobacco use among middle and high school students - United States, 2011-2016. *MMWR Morb Mortal Wkly Rep* 2017 Jun 16;66(23):597-603 [FREE Full text] [doi: [10.15585/mmwr.mm6623a1](https://doi.org/10.15585/mmwr.mm6623a1)] [Medline: [28617771](#)]
26. Kasza KA, Ambrose BK, Conway KP, Borek N, Taylor K, Goniewicz ML, et al. Tobacco-product use by adults and youths in the United States in 2013 and 2014. *N Engl J Med* 2017 Dec 26;376(4):342-353 [FREE Full text] [doi: [10.1056/NEJMsa1607538](https://doi.org/10.1056/NEJMsa1607538)] [Medline: [28121512](#)]
27. Meich R, Johnston L, O'Malley P, Bachman J, Schulenberg J. Monitoring the Future national survey results on drug use, 1975-2014: Volume I, Secondary school students. In: Institute for Social Research. Ann Arbor, Michigan: University of Michigan; 2015.
28. Amazon Mechanical Turk. 2018. URL: <https://www.mturk.com/> [accessed 2018-06-09] [WebCite Cache ID 703ACRDe7]
29. 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](https://doi.org/10.1177/1745691610393980)]
30. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: detecting satisficing to increase statistical power. *J Exp Soc Psychol* 2009 Jul;45(4):867-872. [doi: [10.1016/j.jesp.2009.03.009](https://doi.org/10.1016/j.jesp.2009.03.009)]
31. Slater M. Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Commun Theory* 2007;17(3):281-303. [doi: [10.1111/j.1468-2885.2007.00296.x](https://doi.org/10.1111/j.1468-2885.2007.00296.x)]

32. Aizen I, Fishbein M, Lohmann S, Albarracín D. The influence of attitudes on behavior. In: *Handbook of Attitudes*. 2nd Edition. New York: Springer; 2018.
33. Ajzen I, Albarracín D, Hornik R. Predicting and changing behavior: a reasoned action approach. In: *Prediction and changing behavior: Applying the theory of reasoned action approach*. Mahwah, NJ: Lawrence Erlbaum; 2007:3-21.
34. Gold GJ. Review of predicting and changing behavior: the reasoned action approach. *J Soc Psychol* 2011 Apr 21;151(3):382-385. [doi: [10.1080/00224545.2011.563209](https://doi.org/10.1080/00224545.2011.563209)]
35. Glasman LR, Albarracín D. Forming attitudes that predict future behavior: a meta-analysis of the attitude-behavior relation. *Psychol Bull* 2006 Sep;132(5):778-822 [FREE Full text] [doi: [10.1037/0033-2909.132.5.778](https://doi.org/10.1037/0033-2909.132.5.778)] [Medline: [16910754](https://pubmed.ncbi.nlm.nih.gov/16910754/)]
36. Slovic P. *The Perception of Risk*. London, UK: Earthscan Publications; 2000.
37. McIntosh M. The growth of racketeering. *Econ Soc* 1973 Feb;2(1):35-69. [doi: [10.1080/03085147300000002](https://doi.org/10.1080/03085147300000002)]
38. World Health Organization Framework Convention on Tobacco Control. 2018. URL: <http://www.who.int/fctc/en/> [accessed 2018-06-09] [WebCite Cache ID 703A7U6Xk]
39. Paul Graham. 2005 Nov. Web 2.0 URL: <http://www.paulgraham.com/web20.html> [accessed 2018-05-30] [WebCite Cache ID 6zo9vbZUC]
40. O'Reilly T. O'Reilly Network. 2005 Sep 30. What Is Web 2.0? URL: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?> [accessed 2018-06-09] [WebCite Cache ID 703CLrU4b]
41. Broadcasting Standards Commission. Office of Communication. 1999. Regulating the Internet URL: <https://www.ofcom.org.uk/about-ofcom/website/regulator-archives> [accessed 2018-05-30] [WebCite Cache ID 6zo9WBRm2]
42. British Broadcasting Corporation. 2012 Nov 01. Russia Internet blacklist law takes effect URL: <https://www.bbc.com/news/technology-20096274> [accessed 2018-06-09] [WebCite Cache ID 703APdTaX]
43. Facebook. 2018. Community standards URL: <https://www.facebook.com/communitystandards> [accessed 2018-05-30] [WebCite Cache ID 6zo9CDuiP]
44. TechCrunch. 2017 Jun 01. YouTube bans "hateful" videos from making money via its advertising network URL: <https://techcrunch.com/2017/06/01/youtube-bans-hateful-videos-from-making-money-via-its-advertising-network/> [accessed 2018-05-30] [WebCite Cache ID 6zo99lyHv]
45. TomCat. Politics Plus. 2017 Sep 07. Boycott Facebook URL: <http://httpwww.politicsplus.org/blog/2017/09/07/boycott-facebook/> [accessed 2018-05-30] [WebCite Cache ID 6zo8qfjvr]
46. USA Today. 2017 Oct 03. USA today poll: Facebook faulted for not doing more about Russian ads during the 2016 race URL: <https://www.usatoday.com/story/news/2017/10/03/usa-today-poll-facebook-faulted-not-doing-more-russian-ads-2016-race/722306001/> [accessed 2018-06-09] [WebCite Cache ID 703CfiZyb]
47. Chan M, Morales A, Farhadloo M, Palmer R, Albarracín D. Social Psychological Research Methods: Social Psychological Measurement. 2017. Harvesting and harnessing social media data for psychological research URL: https://www.researchgate.net/publication/318215050_Harvesting_and_Harnessing_Social_Media_Data_for_Psychological_Research [accessed 2018-06-09] [WebCite Cache ID 703DH5FQ7]
48. The Guardian. 2017 Mar 25. Google's bad week: YouTube loses millions as advertising row reaches US URL: <https://www.theguardian.com/technology/2017/mar/25/google-youtube-advertising-extremist-content-att-verizon> [accessed 2018-05-30] [WebCite Cache ID 6zo5DoOfq]
49. Xu X, Mao Z, Halderman J. Internet censorship in China: where does the filtering occur? *Passive and Active Measurement* 2011;133-142. [doi: [10.1007/978-3-642-19260-9_14](https://doi.org/10.1007/978-3-642-19260-9_14)]
50. Wikipedia. 2018. Control software URL: https://en.wikipedia.org/wiki/Content-control_software [accessed 2018-05-30] [WebCite Cache ID 6zo4ySAA9]
51. Chan MS, Jones CR, Hall JK, Albarracín D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol Sci* 2017 Nov;28(11):1531-1546. [doi: [10.1177/0956797617714579](https://doi.org/10.1177/0956797617714579)] [Medline: [28895452](https://pubmed.ncbi.nlm.nih.gov/28895452/)]
52. Chan M, Jones C, Albarracín D. Countering false beliefs: an analysis of the evidence and recommendations of best practices for the retraction and correction of scientific misinformation. In: *Handbook of the Science of Science Communication*. Cambridge, UK: Oxford Handbooks; 2017.
53. Romer D, Jamieson PE, Jamieson KH, Jones C, Sherr S. Counteracting the influence of peer smoking on YouTube. *J Health Commun* 2017 Apr;22(4):337-345 [FREE Full text] [doi: [10.1080/10810730.2017.1290164](https://doi.org/10.1080/10810730.2017.1290164)] [Medline: [28306481](https://pubmed.ncbi.nlm.nih.gov/28306481/)]
54. Hansen WB, Graham JW. Preventing alcohol, marijuana, and cigarette use among adolescents: peer pressure resistance training versus establishing conservative norms. *Prev Med (Baltim)* 1991 May;20(3):414-430. [doi: [10.1016/0091-7435\(91\)90039-7](https://doi.org/10.1016/0091-7435(91)90039-7)]

Abbreviations

e-cigarette: electronic cigarette
LSD: least significant difference
ns: not statistically significant

Edited by G Eysenbach; submitted 26.01.18; peer-reviewed by L Glasman, T Basanez; comments to author 01.03.18; revised version received 23.04.18; accepted 12.05.18; published 29.06.18.

Please cite as:

Albarracin D, Romer D, Jones C, Hall Jamieson K, Jamieson P

Misleading Claims About Tobacco Products in YouTube Videos: Experimental Effects of Misinformation on Unhealthy Attitudes
J Med Internet Res 2018;20(6):e229

URL: <http://www.jmir.org/2018/6/e229/>

doi: [10.2196/jmir.9959](https://doi.org/10.2196/jmir.9959)

PMID: [29959113](https://pubmed.ncbi.nlm.nih.gov/29959113/)

©Dolores Albarracin, Daniel Romer, Christopher Jones, Kathleen Hall Jamieson, Patrick Jamieson. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.06.2018. 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.

Original Paper

Dissemination of a Web-Based Tool for Supporting Health Insurance Plan Decisions (Show Me Health Plans): Cross-Sectional Observational Study

Jingsong Zhao¹, MPH; Nageen Mir², MPH; Nicole Ackermann², MPH; Kimberly A Kaphingst^{1,3}, ScD; Mary C Politi², PhD

¹Huntsman Cancer Institute, Salt Lake City, UT, United States

²Division of Public Health Sciences, Department of Surgery, Washington University in Saint Louis, St. Louis, MO, United States

³Department of Communication, University of Utah, Salt Lake City, UT, United States

Corresponding Author:

Mary C Politi, PhD

Division of Public Health Sciences

Department of Surgery

Washington University in Saint Louis

660 South Euclid Ave

St. Louis, MO, 63110

United States

Phone: 1 (314) 747 1967

Email: mpoliti@wustl.edu

Abstract

Background: The rate of uninsured people has decreased dramatically since the Affordable Care Act was passed. To make an informed decision, consumers need assistance to understand the advantages and disadvantages of health insurance plans. The Show Me Health Plans Web-based decision support tool was developed to improve the quality of health insurance selection. In response to the promising effectiveness of Show Me Health Plans in a randomized controlled trial (RCT) and the growing need for Web-based health insurance decision support, the study team used expert recommendations for dissemination and implementation, engaged external stakeholders, and made the Show Me Health Plans tool available to the public.

Objective: The purpose of this study was to implement the public dissemination of the Show Me Health Plans tool in the state of Missouri and to evaluate its impact compared to the RCT.

Methods: This study used a cross-sectional observational design. Dissemination phase users were compared with users in the RCT study across the same outcome measures. Time spent using the Show Me Health Plans tool, knowledge, importance rating of 9 health insurance features, and intended plan choice match with algorithm predictions were examined.

Results: During the dissemination phase (November 2016 to January 2017), 10,180 individuals visited the SMHP website, and the 1069 users who stayed on the tool for more than one second were included in our analyses. Dissemination phase users were more likely to live outside St. Louis City or County ($P<.001$), were less likely to be below the federal poverty level ($P<.001$), and had a higher income ($P=.03$). Overall, Show Me Health Plans users from St. Louis City or County spent more time on the Show Me Health Plans tool than those from other Missouri counties ($P=.04$); this association was not observed in the RCT. Total time spent on the tool was not correlated with knowledge scores, which were associated with lower poverty levels ($P=.009$). The users from the RCT phase were more likely to select an insurance plan that matched the tool's recommendations ($P<.001$) compared with the dissemination phase users.

Conclusions: The study suggests that a higher income population may be more likely to seek information and online help when making a health insurance plan decision. We found that Show Me Health Plans users in the dissemination phase were more selective in the information they reviewed. This study illustrates one way of disseminating and implementing an empirically tested Web-based decision aid tool. Distributing Web-based tools is feasible and may attract a large number of potential users, educate them on basic health insurance information, and make recommendations based on personal information and preference. However, using Web-based tools may differ according to the demographics of the general public compared to research study participants.

KEYWORDS

Affordable Care Act; health insurance decision aid; health literacy; public health

Introduction

A key strategic goal of the Affordable Care Act (ACA) is to extend affordable coverage to the uninsured [1]. Since the enactment of the ACA, 20 million consumers have gained health insurance coverage and the uninsured rate dropped to a historic low of 8.6% [2,3]. In addition, the ACA improved access to primary care and medications, decreased mortality, and overall improved health outcomes among vulnerable populations [4].

Although the ACA led to better access to care for many, some consumers, especially those new to health insurance, had difficulty understanding health insurance details and using selected health insurance plans [5,6]. Making well-informed decisions about health insurance in the ACA marketplace requires individuals to understand the complex benefits and trade-offs of each insurance plan option and compare them to select the best choice for them. Recognizing a need for better consumer support, many national and state-wide organizations drafted recommendations that organizations should adopt to help consumers identify effective plans [7-9]. In addition, marketplace enrollment assistance programs were created to support consumer choices in-person, by telephone, and through outreach events in most states.

Despite public education and outreach efforts, health insurance literacy remains a critical barrier impacting enrollment decisions [10]. Limited knowledge about health insurance may hinder consumers' abilities to select a suitable health plan and use it to obtain health care. Many in-person assistance programs have very large caseloads, and are not able to reach everyone who might need guidance selecting insurance [11]. Given the budget cuts in the 2018 enrollment cycle [12], providing comprehensive in-person assistance has become even more challenging.

Web-based resources may be an effective way to supplement in-person assistance for learning about health insurance and plan selection. Web-based support may be particularly important for reaching disadvantaged populations with limited access to in-person outreach, or those who require more guidance than can be accomplished in face-to-face meetings; online tools may promote health equity among these groups [13,14]. However, developing and disseminating Web-based resources are different from in-person support workshops and may require systematic dissemination and implementation plans to reach target populations.

In order to distribute information on federal health plans and assist consumers with health insurance plan selection in the ACA marketplace, we created the Show Me Health Plans (SMHP) Web-based decision support tool. Designed for those with limited health insurance literacy, SMHP delivered health insurance education and calculated estimated annual costs for each user based on Medical Expenditure Panel Survey (MEPS) data for their age, gender, and self-reported health conditions.

It then sorted health insurance plans in the ACA marketplace by lowest to highest annual cost, and provided ranges of potential annual costs across the marketplace. The effectiveness of the SMHP tool was examined in a randomized controlled trial (RCT), and described in previous publications [15,16]. Study participants who used SMHP had higher health insurance knowledge, decision self-efficacy, confidence in their health plan choice, and improved health insurance literacy compared to participants using the federal health insurance exchange website [15,16]. They were also more likely to select plans that better matched their health care needs.

After the RCT, in response to the promising effectiveness of SMHP in a research context and the growing need for Web-based health insurance decision support, the study team released the SMHP tool so that it was available to the general public. We used principles of dissemination and implementation science to guide efforts in translation and adoption of research evidence to the target population throughout the state of Missouri. There is increasing interest and investment in disseminating and translating effective study interventions more broadly to target populations [17-19]. Political support, funding agency priorities, capacity of the dissemination organization, and researchers' knowledge on applying study findings can influence the success of study translation [20-23]. Although many researchers have dissemination expectations from funders [20], the majority of public health professionals spend little time and effort on program dissemination [24], and few studies reported sustainability and challenges of disseminating an online decision aid tool from an effective intervention study. To address this gap, we examined the use of the SMHP tool during its dissemination to the general public in Missouri and compared this with use of the tool during the RCT.

Methods

Overview

This study used a cross-sectional observational design to examine the public dissemination of the SMHP tool across Missouri. Dissemination phase users were compared with users from a RCT study previously carried out by the authors. In brief, the RCT recruited English-speaking participants, aged 18 to 64 years, not eligible for Medicaid, and living within 90 miles of St. Louis, Missouri. Enrolled participants (n=328) were randomly allocated to the SMHP intervention group or to the HealthCare.gov control group [15,16]. In the dissemination phase, the SMHP tool had information on page one about users who might benefit most from the SMHP tool (ie, people living in Missouri eligible for the ACA marketplace), but anyone could access the information without a login ID or access criteria. On the first page of the tool, visitors were notified that some of the information they entered would be used for scientific research, but all of their information was anonymous and not connected to identifying information.

Table 1. Key expert recommendations for implementing evidence-based interventions.

Recommendation	Strategies	Implementation examples
Develop stakeholder interrelationships	<ul style="list-style-type: none"> Identify and prepare champions Build a coalition Identify and prepare champions Use advisory boards and workgroups 	<ul style="list-style-type: none"> The study team built up coalitions with local health nonprofit organizations, community action agencies, and local health care centers and departments throughout the randomized controlled trial and dissemination phases These community partners were regularly updated on study findings, received updated website information for their use, and continued to communicate with the study team about health insurance reform and decision support
Train and educate stakeholders	<ul style="list-style-type: none"> Conduct educational meetings Conduct educational outreach visits Develop educational materials Inform local opinion leaders 	<ul style="list-style-type: none"> Several strategies were utilized to educate stakeholders, including developing educational materials, conducting educational meetings and outreach visits, and informing local opinion leaders
Use evaluative and iterative strategies	<ul style="list-style-type: none"> Assess for readiness and identify barriers and facilitators Conduct local needs assessment 	<ul style="list-style-type: none"> Local needs assessments were conducted to collect information on the SMHP^a website The study team assessed the likelihood of adoption and implementation of the website, along with potential barriers and facilitators to implementation
Adapt and tailor to context	<ul style="list-style-type: none"> Promote adaptability Tailor strategies 	<ul style="list-style-type: none"> Website changes made to the tool adapted based on collected feedback
Engage consumers	<ul style="list-style-type: none"> Use mass media 	<ul style="list-style-type: none"> SMHP was featured on local television news, shared via social media, and shared via electronic newsletters to reach large number of consumers and health policy experts

^aSMHP: Show Me Health Plans.

Dissemination of the SMHP Tool

The study team relied on several expert recommendations for implementation of evidence-based interventions when disseminating the SMHP tool [25,26]. Key strategies are summarized and displayed in Table 1.

Tool Content Changes

The organization of the SMHP tool during the RCT phase included five sections: (1) *Welcome* (to introduce the goals of the tool); (2) *Let's Learn* (to educate users on different topics important to know prior to purchasing health insurance coverage); (3) *Let's Review* (to measure the user's knowledge/understanding of key terms); (4) *Eligibility* (to assess whether or not the user is eligible for Marketplace plans based on the information provided); and (5) *Your Plans* (to display good-fit plans based on the information the user inputs in the eligibility section). In the trial, users had to view all sections in order. During the dissemination phase, users could reach sections in any order except the last section. This allowed SMHP users to skip sections and choose the ones that they wanted to view.

Four types of changes were made between the RCT and dissemination phases based on stakeholder feedback, including design changes, content changes, page section changes, and wording changes. Design changes (eg, darken text, increase size of image, label a *Next* button rather than simply display an arrow) were to help SMHP users navigate through the tool effectively. Content changes included reiterating statements on

preventive care, pre-existing conditions, and out-of-pocket maximum; adding a new pregnancy question to better calculate Medicaid eligibility; updating the list of health conditions assessed to generate a more precise cost estimate of health care expenses for each user; and adding a link pointing to additional information resources. Content changes were made based on new ACA policies and suggestions from stakeholders and community members. A *Simple Choice Plan* page was added based on new policy changes in the marketplace, and a *Gateway to Better Health* program page was added to inform those in St. Louis City and County about a bridge program to provide limited coverage to those who were ineligible for Medicaid since Missouri did not expand the program. Furthermore, RCT study information was deleted and wording changes were made to add clarity to the tool content.

Measures

For this analysis, demographic information, including age, gender, income, federal poverty level, number of chronic conditions, and county of residence were collected. During the analysis, users in the dissemination phase were divided into two groups: those who started using the tool but did not finish all the sections (ie, started group), while the finished tool group was defined as those who finished the tool and saw their health insurance plan options.

Use of the Tool

Each time a user logged on to the tool, a unique session ID was generated, which enabled each visit to be tracked. Each time a page on the tool was accessed, a tracking database stored the

session ID, date and time, as well as user actions (ie, whether they logged in, logged out, viewed the page, or redirected). Session IDs were randomly generated and created so that users could not be identified.

Knowledge

Knowledge was measured using eight questions in the *Let's Review* section. The scale was developed based on our past work assessing health insurance knowledge [15,16]. Knowledge was assessed based on the percentage of people answering each item correctly.

Importance Rating

In the *Let's Review* section, users were asked to rate the importance of nine insurance features from least to most important on a scale of 1 (not at all important) to 5 (very important). The nine features were: cost of health insurance premium, cost of deductible, cost of doctor visits, cost of prescription pills or medicine, choice of doctors (including some that are out-of-network), cost of out-of-network care, fixed costs for tests or care, out-of-pocket maximum, and formulary.

Match With Preferences and Algorithm Predictions

In the *Your Plans* section, the tool recommended three good-fit insurance plans based on participant's eligibility and estimated costs across available plans. It also invited users to select an intended plan choice from the entire list of available plan options. Match with preferences was assessed by comparing the participant's intended plan choice with the most important features. Matches were categorized as good, moderate, or poor, using methodology described in prior papers [15,16]. Good matches included plan selections that included features participants rated as important to them. For example, if participants rated premiums as "very important" (ie, a rating of 4 or 5), and ranked premium cost as most important to their plan choice, and chose a plan with a premium in the lowest 25% out of plans available, it was considered a good match. Moderate matches would include plans with some features rated as important to participants (eg, rating premium as "very important" and most important to plan choice, then selecting a plan with a premium in the lowest 50% of all plans). Poor matches were plans that did not include many features participants rated or ranked as important. Match with algorithmic predictions was calculated by comparing how many SMHP users selected one of the plans that was displayed as a "good fit" plan based on their demographics and health care needs.

Data Analysis

Descriptive statistics were calculated for demographics, time spent using the tool, importance ratings, and plan choice match stratified by phase of tool use. Means and standard deviations or frequencies and percentages are presented for all variables. Additionally, range is presented for time spent using the tool, and medians and interquartile ranges are presented for importance rankings. We conducted bivariate analyses to test for associations between phase of tool use (RCT versus dissemination phase) and demographics; between those who began the eligibility portion and those who completed the eligibility portion to view plans; time spent using the tool,

importance ratings, and plan choice match. Chi-square tests were used for categorical variables, and *t*-tests were used for parametric continuous variables, using a Satterthwaite adjustment for inequality of variances when appropriate, or a Wilcoxon rank sum test or Kruskal-Wallis test for nonparametric data. For correlations obtained for time usage between RCT and dissemination phases across demographics and knowledge scores, Pearson's correlation was obtained for parametric data, and Spearman's correlation was obtained in instances where data were nonparametric. SAS version 9.4 was used for analyses.

Results

Participant Characteristics

During the dissemination phase (November 9, 2016 to January 31, 2017), 10,180 individuals visited the SMHP tool. Of those 10,180 individuals, 1069 stayed on the tool for at least one second, suggesting that they did not exit after briefly viewing the home page. The mean age of SMHP users ($n=386$), who began the eligibility section in the dissemination phase, was 43.6 years (SD 14.4), more than half were female (212/374, 56.7%), 52% came from St. Louis City or County (196/374), and 57% had one or more chronic conditions (201/350, see Table 2). Comparing the dissemination phase users to the RCT users, the first group was more likely to live outside St. Louis City or County ($P<.001$), were less likely to be below the federal poverty level (FPL; $P<.001$), and had a higher income (mean US \$40,523 versus US \$30,407, $P=.03$).

We compared the characteristics of users who began the eligibility section of the tool ($n=56$) to those who completed the eligibility section and saw plans ($n=330$) for the dissemination phase. No significant differences were found between the 2 groups in terms of age, gender, county of residence, and FPL, but those who finished the tool had higher income (mean US \$41,085 versus US \$14,000, Wilcoxon Rank Sum $|Z|$ approximation=2.8, $P=.005$), and were significantly more likely to have a chronic condition (59.1% versus 30.0%; $\chi^2=6.5$, $P=.01$).

Use of the Show Me Health Plans (SMHP) Tool

The median total time for tool usage for the 1069 dissemination phase users was 0.9 minutes (range 0.02-189.1; Table 3). Three-quarters of participants spent 7.5 minutes or less on the tool. One hundred and thirty SMHP users viewed each page of the tool's five sections with a median time of 17 minutes (range 3.5-189.1). This was significantly lower than time spent by RCT users ($P=.001$), who had a median time of 21.5 minutes (range 6.3-175.1). Compared to the RCT phase, users spent less time on the *Welcome* ($P<.001$), *Let's Learn* ($P<.001$), *Let's Review* ($P<.001$), *Eligibility* ($P<.001$), and *Your Plans* ($P<.001$) sections. All users in the dissemination phase started on the *Welcome* section (1069/1069, 100%) and of those users, 46% (488/1069) went on to view the *Let's Learn* section, 34% (362/1069) went on to view the *Let's Review* section, 43% (459/1069) went on to view the *Eligibility* section, and 31% (331/1069) went on to view the *Your Plans* section. Users in the RCT phase were required to go through the entire tool so their usage did not differ.

Table 2. Demographics of users from the dissemination and randomized controlled trial (RCT) phases.

Variable	Dissemination phase	RCT phase	Test statistic	P value
Age ^a (n=386), mean (SD)	43.6 (14.4)	43.1 (13.2)	-0.4 ^b	.69
Gender (n= 374), n (%)				
Male	162 (43.3)	67 (40.9)		
Female	212 (56.7)	97 (59.1)	0.28 ^c	.59
County (n=374), n (%)				
St. Louis City or County	196 (52.4)	152 (92.7)		
Other	178 (47.6)	12 (7.3)	81.0 ^c	<.0001
Income ^d (n= 337), mean (SD)	40,523.20 (38,867)	30,407.01 (54,402)	-2.13 ^b	.03
Federal poverty level (n=334), n (%)				
<100%	45 (13.5)	72 (43.9)		
100%-249%	152 (45.5)	64 (39.0)		
250%-399%	79 (23.7)	15 (9.2)		
≥	58 (17.4)	13 (7.9)	63.6 ^c	<.0001
Number of chronic conditions (n= 350), n (%)				
0	149 (42.6)	68 (41.5)		
≥1	201 (57.4)	96 (58.5)	0.06 ^c	.81

^an=386 (dissemination phase); n=164 (RCT).^bRefers to *t* values.^cRefers to χ^2 values.^dn=337 (dissemination phase); n=164 (RCT). All values in US \$.**Table 3.** Time spent using the Show Me Health Plans (SMHP) tool by users in the dissemination and the randomized controlled trial (RCT) phases.

Variable	Dissemination phase ^a			RCT phase			Wilcoxon rank sum, Z approximation	<i>P</i> value ^c
	n	Median time in sec (IQR ^b)	Range (sec)	n	Median time in sec (IQR)	Range (sec)		
Section								
<i>Welcome</i>	1069	7.0 (20.0)	1-1751	164	19.0 (22.5)	5-260	8.78	<.001
<i>Let's Learn</i>	488	75.5 (228.5)	1-4967	164	284.5 (245.0)	26-2328	9.67	<.001
<i>Let's Review</i>	362	116.0 (217.0)	1-3081	164	383.5 (217.5)	144-1685	14.38	<.001
<i>Eligibility</i>	459	132 (167.0)	1-4085	164	279.0 (171.5)	94-1013	10.19	<.001
<i>Your Plans</i>	331	138 (324)	1-8319	164	180.5 (292.5)	18-7662	4.07	<.001
All included ^d	1069	0.9 (7.5)	0.02-189.1	164	21.5 (15.1)	6.3-175.1	17.0	<.001
Completed entire tool ^d	130	17.0 (17.8)	3.5-189.1	164	21.5 (15.1)	6.3-175.1	3.24	.0012
At least started each section ^d	229	12.2 (15.6)	1.0-189.1	164	21.5 (15.1)	6.3-175.1	7.27	<.001

^aOnly users with greater than zero seconds time are included in time calculations (ie, those who just clicked are not included).^bIQR: interquartile range.^cTesting difference in time between dissemination and RCT phases.^dValues refer to overall time; times are expressed in minutes.

SMHP users from St. Louis City or County spent more time overall on the tool than those from other Missouri counties in the dissemination phase (Wilcoxon Rank Sum |Z| approximation=2.01, *P*=.04); this association was not observed

in the RCT group. Age (*r*=.24, *P*<.001) was positively correlated to overall time spent using SMHP in both phases; however, it had a stronger positive correlation with overall time spent using SMHP in the RCT phase (*r*=.32, *P*<.001). The number of

chronic conditions was only positively associated with overall time spent using SMHP in the RCT phase ($r=.23$, $P=.003$).

Knowledge Scores

In the dissemination phase, SMHP users had a mean knowledge score of 89.5% (SD 15.3), compared to 77.4% (SD 18.2) for users in the RCT. The total time spent on the tool was not correlated with knowledge scores. In both phases, the knowledge score was directly associated with the percentage of FPL (Kruskal-Wallis $\chi^2=9.0$, $P=.03$ [RCT] and Kruskal-Wallis $\chi^2=11.5$, $P=.009$ [dissemination phase]).

Importance Ranking

Out of nine categories, cost of health insurance premium (126/182, 69.2%) and choice of doctors (110/171, 64.3%) were ranked as the most important factors when considering a health insurance plan in the dissemination phase, while costs of out-of-network care (29/173, 16.9%) received the highest percentage of not important rankings. In contrast, out-of-pocket maximum (111/164, 67.7%) and cost of health insurance

premium (108/164, 65.9%) received the highest percentage of most important rankings in the RCT, while choice of doctors (8/164, 4.9%) and cost of out-of-network care (7/164, 4.3%) received the highest percentages of not important rankings. The mean importance rankings of cost of doctors' visits ($P=.002$), cost of prescription pills or medicine ($P=.002$), cost of doctors ($P=.004$), cost of out-of-network care ($P<.001$), fixed cost for tests or care ($P<.001$) and out-of-pocket maximum ($P=.01$) were different in the dissemination phase and RCT phase (Table 4).

Match with Preferences and Algorithm Predictions

Only 39 SMHP users selected a plan choice during the dissemination phase by "starring" a plan on the website. Of these, 97% (38/39) of selected plans were good or moderate matches (good matches: 17/39, 44%; moderate matches: 21/39, 54%), and 22 matched one of the algorithm recommendations (22/39, 56%). The users in the RCT phase were more likely to have a match in the algorithm recommendations ($P<.001$) but not in the match score ($P=.52$; see Table 5), compared with the dissemination phase users.

Table 4. Importance ranking between the dissemination and randomized controlled trial (RCT) phases.

Question	Dissemination phase			RCT phase			Wilcoxon rank sum, [Z] approximation	P value
	n	Mean ^a (SD)	Median (IQR ^b)	n	Mean ^a (SD)	Median (IQR)		
Cost of health insurance premium	182	4.5 (0.9)	5.0 (1.0)	164	4.5 (0.9)	5.0 (1.0)	0.49	.62
Cost of deductible	182	4.3 (1.0)	5.0 (1.0)	164	4.4 (0.9)	5.0 (1.0)	0.13	.89
Cost of doctor visits	173	4.0 (1.1)	4.0 (2.0)	164	4.3 (1.0)	5.0 (1.0)	2.99	.003
Cost of prescription pills or medicine	173	4.1 (1.1)	4.0 (2.0)	164	4.4 (1.0)	5.0 (1.0)	3.01	.003
Choice of doctors, including some that are out-of-network	171	4.3 (1.1)	5.0 (1.0)	164	4.0 (1.2)	5.0 (2.0)	2.82	.005
Cost of out-of-network care	173	3.2 (1.4)	3.0 (3.0)	164	4.0 (1.2)	5.0 (2.0)	5.45	<.001
Fixed cost for tests or care	171	4.0 (1.1)	4.0 (2.0)	164	4.4 (0.9)	5.0 (1.0)	3.75	<.001
Out-of-pocket maximum	173	4.3 (1.0)	5.0 (1.0)	164	4.6 (0.8)	5.0 (1.0)	2.49	.01
Formulary	171	4.2 (1.2)	5.0 (2.0)	164	4.3 (1.1)	5.0 (1.0)	0.72	.47

^aImportance of features ranked from least to most important on a scale of 1 (not at all important) to 5 (very important).

^bIQR: interquartile range.

Table 5. Match with preferences and algorithm predictions between the dissemination and randomized controlled trial (RCT) phases.

Variable	Dissemination phase, n (%)	RCT phase, n (%)	χ^2	P value ^a
Choice match algorithm			12.5	<.001
Yes	22 (56.4)	134 (82.7)		
No	17 (43.6)	28 (17.3)		
Match score			1.29	.52
Good match	17 (43.6)	85 (52.5)		
Moderate match	21 (53.9)	71 (43.8)		
Poor match	1 (2.6)	6 (3.7)		

^aTesting difference between dissemination and RCT phases.

Discussion

Principal Results

Dissemination of the Web-based SMHP health insurance decision support tool successfully reached a large number of users in the state of Missouri. Although this phase was successful in reaching many users, there were key lessons learned from this process. First, the dissemination tool users had a significantly higher income level compared to those in the RCT phase, suggesting that a higher income population may be more likely to seek information and online help when making a health insurance plan decision. Barriers that may hinder low income populations from seeking online insurance help could be limited time to access the internet, lack of interest seeking online information, lack of familiarity with tools or discomfort entering personal information online [27-29]. To better facilitate the dissemination of this tool to a larger target audience and increase its visibility, future work could incorporate the tool information into other assistant programs and websites. For example, since people are more likely to pay attention to personally relevant information, marketing the tailored tool on insurance enrollment websites could encourage users to engage with the tool. However, there are challenges associated with keeping the tool sustainable and updated, including the costs of advertising and dissemination as well as website maintenance.

In addition, our findings suggest that SMHP users were more selective in the information they reviewed and spent less time on all the sections in the dissemination phase compared with the RCT phase. In the dissemination phase, only a small number of SMHP users filled out their personal and family information and reviewed the final health insurance plan recommendations. This may indicate that people are reluctant to disclose personal health information online [30]. Prior studies have found that people's perceived health information sensitivity influences their intention to disclose health information [31]. Online users chose to share private information when the perceived benefits outweighed the perceived risks [32,33]. For example, patients were willing to share electronic health data with their health care professionals, but were less inclined to permit secondary data use when there was a greater risk of confidentiality loss [34]. Collecting some private information is necessary for the implementation of an online tool that generates tailored health insurance recommendations, but the process of collecting personal information or the level of detail may be improved by promoting a model of trust and safety with information technology.

Low completion rate might also be linked to the current insurance status of SMHP users. For instance, SMHP users who were already enrolled in health insurance in the 2016 cycle may have completed the educational sections to learn more about health insurance. In this case, the *Eligibility* and *Your Plans* sections would not have been relevant to these users. In addition, users from outside the state of Missouri would not have benefited from the cost calculator, which was specific to Missouri plans; this may explain why some users did not complete the tool.

Furthermore, this study actively integrated evidence-based strategies for implementing change [25,26], built coalitions with local stakeholders and communities, and created a feedback loop in developing tool content to transform a research tool into a publicly available online tool. Damschroder et al [35] suggested that implementing a research tool in a particular setting requires the preservation of the essential and core elements of the tool as well as modification of adaptable elements based on the dissemination settings. The evidence-based strategies [25,26] were helpful for engaging stakeholders and community members to identify the core and adaptable elements in order to assess the likelihood of adoption of SMHP, along with potential barriers and facilitators to implementation. In addition, formatting visual and written context in a research study is different because research participants are more engaged in the study procedures and interventions, thus seeking feedback from stakeholders could potentially reduce user burden and make the tool robust across broader populations, especially for an audience with low health literacy.

However, even with these extensive strategies, many more users are available in the marketplace across the state, and additional work may need to be done in-person to promote the routine adoption of tools like SMHP. When a public health program proceeds to the dissemination phase, there are many confounding factors that might affect the utilization of the tool. For example, participants in the RCT phase had a better match score on health insurance plan recommendations compared to the users in the dissemination phase. Therefore, we cannot determine if any other factors impacted a user's final choice in the dissemination phase.

Limitations

One primary limitation of this study is that dissemination phase users were less likely to complete demographic questions. Sample sizes for demographic variables varied, as users often stopped before completing the entire tool. SMHP users were allowed to skip the sections that required them to enter their personal information, which is typical of dissemination in a real-world setting. In addition, we did not track SMHP users who used the tool under a different user ID as the user ID generated by the tool was not linked to an IP address or any other identifiable information. When we compared the users who began the *Eligibility* section of the tool to those who completed the *Eligibility* section and saw plans, fewer data were available from sections later in the tool. More SMHP users, therefore, reported age and county, which were asked earlier in the tool, but fewer reported income.

We assume that anyone who stayed on the website for one or more seconds was categorized as a SMHP user, indicating that those users did not open and then immediately close the tool. The reason for this assumption is based on page hits, as we were able to calculate time for anyone who clicked past the first page. This could lead to sampling bias given that we were only collecting data from SMHP users who may favor the SMHP tool. Additionally, because of skewed distributions in many instances, nonparametric tests were used for analyses.

Conclusions

This study provides an example of the dissemination and implementation of an empirically tested Web-based decision aid tool for the general public. From this experience, we can conclude that disseminating this tool is feasible as it was able to attract potential users, educate them on basic health insurance terms, and make recommendations based on personal

information and preference. In addition, this also serves as an example of a successful adoption of evidence-based recommendations for implementing change [25,26]. However, future research is needed to investigate the factors that impact online users' information-seeking behaviors when using a public health information tool, as well as explore strategies that may engage low income populations.

Conflicts of Interest

MCP has a research contract (2017-2018) from Merck & Do on a topic unrelated to the content of this manuscript.

References

1. US Department of Health & Human Services. Strategic Plan FY 2014 -2018. 2014. URL: <https://www.hhs.gov/about/strategic-plan/index.html> [accessed 2017-11-29] [WebCite Cache ID 6wB4Oq3dn]
2. Uberoi N, Finegold K, Gee E. Health Insurance Coverage and the Affordable Care Act, 2010-2016. 2016. URL: <https://aspe.hhs.gov/system/files/pdf/187551/ACA2010-2016.pdf> [accessed 2017-11-29] [WebCite Cache ID 6wB4oAzAn]
3. Avery K, Finegold K, Whitman A. Affordable Care Act Has Led to Historic, Widespread Increase in Health Insurance Coverage. 2016. URL: <https://aspe.hhs.gov/system/files/pdf/207946/ACAHistoricIncreaseCoverage.pdf> [accessed 2017-11-29] [WebCite Cache ID 6wB55RB9X]
4. Sommers BD, Gunja MZ, Finegold K, Musco T. Changes in Self-reported Insurance Coverage, Access to Care, and Health Under the Affordable Care Act. JAMA 2015 Jul 28;314(4):366-374. [doi: [10.1001/jama.2015.8421](https://doi.org/10.1001/jama.2015.8421)] [Medline: [26219054](https://pubmed.ncbi.nlm.nih.gov/26219054/)]
5. Long SK, Kenney GM, Zuckerman S, Goin DE, Wissoker D, Blavin F, et al. The health reform monitoring survey: addressing data gaps to provide timely insights into the affordable care act. Health Aff (Millwood) 2014 Jan;33(1):161-167. [doi: [10.1377/hlthaff.2013.0934](https://doi.org/10.1377/hlthaff.2013.0934)] [Medline: [24352654](https://pubmed.ncbi.nlm.nih.gov/24352654/)]
6. Long SK, Shartz A, Politi M. Low Levels of Self-Reported Literacy and Numeracy Create Barriers to Obtaining and Using Health Insurance Coverage. 2014. URL: <http://hrms.urban.org/briefs/Low-Levels-of-Self-Reported-Literacy-and-Numeracy.html> [accessed 2017-11-28] [WebCite Cache ID 6wB5G47Yd]
7. National Partnership for Women & Families. Affordable Care Act. URL: <http://www.nationalpartnership.org/> [accessed 2017-11-29] [WebCite Cache ID 6zgeTtd11]
8. Rand Corporation. Affordable Care Act. 2017. URL: <https://www.rand.org/> [accessed 2017-11-29] [WebCite Cache ID 6wOgHUBdH]
9. Cover Missouri. About the Missouri Health Insurance Marketplace. 2017. URL: <http://covermissouri.org/rules/marketplace/> [accessed 2017-11-29] [WebCite Cache ID 6wOfxGv16]
10. Giovannelli J, Curran E. Factors Affecting Health Insurance Enrollment Through the State Marketplaces: Observations on the ACA's Third Open Enrollment Period. Issue Brief (Commonw Fund) 2016 Jul;19:1-12. [Medline: [27459742](https://pubmed.ncbi.nlm.nih.gov/27459742/)]
11. Pollitz K, Tolbert J, Semansky A. 2016 Survey of Health Insurance Marketplace Assister Programs and Brokers.: The Henry J. Kaiser Family Foundation; 2016. URL: <http://files.kff.org/attachment/2016-Survey-of-Marketplace-Assister-Programs-and-Brokers> [accessed 2017-11-29] [WebCite Cache ID 6zgelyIMn]
12. Pollitz K, Tolbert J, Diaz M. Data Note: Changes in 2017 Federal Navigator Funding.: The Henry J. Kaiser Family Foundation; 2017. URL: <https://www.kff.org/health-reform/issue-brief/data-note-changes-in-2017-federal-navigator-funding/> [accessed 2017-11-29] [WebCite Cache ID 6wB5jWMm5]
13. 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 Apr 23;15(4):e85 [FREE Full text] [doi: [10.2196/jmir.1933](https://doi.org/10.2196/jmir.1933)] [Medline: [23615206](https://pubmed.ncbi.nlm.nih.gov/23615206/)]
14. Welch V, Petkovic J, Pardo PJ, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [Medline: [27077792](https://pubmed.ncbi.nlm.nih.gov/27077792/)]
15. Politi MC, Barker AR, Kaphingst KA, McBride T, Shacham E, Kebodeaux CS. Show Me My Health Plans: a study protocol of a randomized trial testing a decision support tool for the federal health insurance marketplace in Missouri. BMC Health Serv Res 2016 Feb 16;16:55 [FREE Full text] [doi: [10.1186/s12913-016-1314-9](https://doi.org/10.1186/s12913-016-1314-9)] [Medline: [26880251](https://pubmed.ncbi.nlm.nih.gov/26880251/)]
16. Politi MC, Kuzemchak MD, Liu J, Barker AR, Peters E, Ubel PA, et al. Show Me My Health Plans: Using a Decision Aid to Improve Decisions in the Federal Health Insurance Marketplace. MDM Policy Pract 2016 Jul;1:1-11 [FREE Full text] [Medline: [28804780](https://pubmed.ncbi.nlm.nih.gov/28804780/)]
17. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol 2015 Sep 16;3:32 [FREE Full text] [doi: [10.1186/s40359-015-0089-9](https://doi.org/10.1186/s40359-015-0089-9)] [Medline: [26376626](https://pubmed.ncbi.nlm.nih.gov/26376626/)]
18. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med 2012 Sep;43(3):337-350 [FREE Full text] [doi: [10.1016/j.amepre.2012.05.024](https://doi.org/10.1016/j.amepre.2012.05.024)] [Medline: [22898128](https://pubmed.ncbi.nlm.nih.gov/22898128/)]

19. Woolf SH. The meaning of translational research and why it matters. *JAMA* 2008 Jan 09;299(2):211-213. [doi: [10.1001/jama.2007.26](https://doi.org/10.1001/jama.2007.26)] [Medline: [18182604](#)]
20. Tabak RG, Stamatakis KA, Jacobs JA, Brownson RC. What predicts dissemination efforts among public health researchers in the United States? *Public Health Rep* 2014;129(4):361-368 [FREE Full text] [doi: [10.1177/003335491412900411](https://doi.org/10.1177/003335491412900411)] [Medline: [24982539](#)]
21. Tabak RG, Padek MM, Kerner JF, Stange KC, Proctor EK, Dobbins MJ, et al. Dissemination and Implementation Science Training Needs: Insights From Practitioners and Researchers. *Am J Prev Med* 2017 Mar;52(3S3):S322-S329 [FREE Full text] [doi: [10.1016/j.amepre.2016.10.005](https://doi.org/10.1016/j.amepre.2016.10.005)] [Medline: [28215389](#)]
22. Harris JR, Cheadle A, Hannon PA, Forehand M, Lichiello P, Mahoney E, et al. A framework for disseminating evidence-based health promotion practices. *Prev Chronic Dis* 2012;9:E22 [FREE Full text] [Medline: [22172189](#)]
23. Schell S, Luke D, Schooley M, Elliott MB, Herbers SH, Mueller NB, et al. Public health program capacity for sustainability: a new framework. *Implement Sci* 2013 Feb 01;8:15 [FREE Full text] [doi: [10.1186/1748-5908-8-15](https://doi.org/10.1186/1748-5908-8-15)] [Medline: [23375082](#)]
24. 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](https://doi.org/10.2105/AJPH.2012.301165)] [Medline: [23865659](#)]
25. Waltz TJ, Powell BJ, Chinman MJ, Smith JL, Matthieu MM, Proctor EK, et al. Expert Recommendations for Implementing Change (ERIC): protocol for a mixed methods study. *Implement Sci* 2014 Mar 26;9:39 [FREE Full text] [doi: [10.1186/1748-5908-9-39](https://doi.org/10.1186/1748-5908-9-39)] [Medline: [24669765](#)]
26. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci* 2015 Feb 12;10:21 [FREE Full text] [doi: [10.1186/s13012-015-0209-1](https://doi.org/10.1186/s13012-015-0209-1)] [Medline: [25889199](#)]
27. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. *J Am Med Inform Assoc* 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: [10.1093/jamia/ocv204](https://doi.org/10.1093/jamia/ocv204)] [Medline: [27206459](#)]
28. Kontos EZ, Bennett GG, Viswanath K. Barriers and facilitators to home computer and internet use among urban novice computer users of low socioeconomic position. *J Med Internet Res* 2007 Oct 22;9(4):e31 [FREE Full text] [doi: [10.2196/jmir.9.4.e31](https://doi.org/10.2196/jmir.9.4.e31)] [Medline: [17951215](#)]
29. 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 Jul 16;16(7):e172 [FREE Full text] [doi: [10.2196/jmir.3117](https://doi.org/10.2196/jmir.3117)] [Medline: [25048379](#)]
30. Marreiros H, Tonin M, Vlassopoulos M, Schraefel MC. "Now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization* 2017 Aug;140:1-17 [FREE Full text]
31. Bansal G, Zahedi FM, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems* 2010;49(2):138-150.
32. Debatin B, Lovejoy JP, Horn AK, Hughes BN. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication* 2009;15(1):83-108.
33. Lee H, Park H, Kim J. Why do people share their context information on Social Network Services? A qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk. *International Journal of Human-Computer Studies* 2013;71(9):862-877 [FREE Full text]
34. Perera G, Holbrook A, Thabane L, Foster G, Willison D. Views on health information sharing and privacy from primary care practices using electronic medical records. *Int J Med Inform* 2011 Feb;80(2):94-101. [doi: [10.1016/j.ijmedinf.2010.11.005](https://doi.org/10.1016/j.ijmedinf.2010.11.005)] [Medline: [21167771](#)]
35. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci* 2009 Aug 07;4:50 [FREE Full text] [doi: [10.1186/1748-5908-4-50](https://doi.org/10.1186/1748-5908-4-50)] [Medline: [19664226](#)]

Abbreviations

ACA: Affordable Care Act
MEPS: Medical Expenditure Panel Survey
RCT: randomized controlled trial
SMHP: Show Me Health Plans

Edited by G Eysenbach; submitted 17.01.18; peer-reviewed by B Braun, C Hoving; comments to author 30.03.18; revised version received 30.04.18; accepted 12.05.18; published 20.06.18.

Please cite as:

Zhao J, Mir N, Ackermann N, Kaphingst KA, Politi MC

Dissemination of a Web-Based Tool for Supporting Health Insurance Plan Decisions (Show Me Health Plans): Cross-Sectional Observational Study

J Med Internet Res 2018;20(6):e209

URL: <http://www.jmir.org/2018/6/e209/>

doi: [10.2196/jmir.9829](https://doi.org/10.2196/jmir.9829)

PMID: [29925498](https://pubmed.ncbi.nlm.nih.gov/29925498/)

©Jingsong Zhao, Nageen Mir, Nicole Ackermann, Kimberly A Kaphingst, Mary C Politi. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.06.2018. 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.

Short Paper

Using a Virtual Reality Social Network During Awake Craniotomy to Map Social Cognition: Prospective Trial

Florian Bernard^{1,2*}, MD; Jean-Michel Lemée^{1,3}, MD, PhD; Ghislaine Aubin^{4*}, PhD; Aram Ter Minassian^{5,6*}, MD, PhD; Philippe Menei^{1*}, MD, PhD

¹Neurosurgery, CHU Angers, Angers, France

²Laboratoire d'Anatomie, Faculté de Médecine d'Angers, Angers, France

³CRCINA, INSERM, Université de Nantes, Université d'Angers, Angers, France

⁴CHU Angers, Département d'Ortophonie, Angers, France

⁵CHU Angers, Département d'Anesthésie-Réanimation, Angers, France

⁶CHU Angers, LARIS EA 7315, Image Signal et Sciences du Vivant, Angers, France

*these authors contributed equally

Corresponding Author:

Philippe Menei, MD, PhD

Neurosurgery

CHU Angers

Angers,

France

Phone: 33 680353917

Email: phmenei@chu-angers.fr

Abstract

Background: In awake craniotomy, it is possible to temporarily inactivate regions of the brain using direct electrical stimulation, while the patient performs neuropsychological tasks. If the patient shows decreased performance in a given task, the neurosurgeon will not remove these regions, so as to maintain all brain functions.

Objective: The objective of our study was to describe our experience of using a virtual reality (VR) social network during awake craniotomy and discuss its future applications for perioperative mapping of nonverbal language, empathy, and theory of mind.

Methods: This was a single-center, prospective, unblinded trial. During wound closure, different VR experiences with a VR headset were proposed to the patient. This project sought to explore interactions with the neuropsychologist's avatar in virtual locations using a VR social network as an available experience.

Results: Three patients experienced VR. Despite some limitations due to patient positioning during the operation and the limitation of nonverbal cues inherent to the app, the neuropsychologist, as an avatar, could communicate with the patient and explore gesture communication while wearing a VR headset.

Conclusions: With some improvements, VR social networks can be used in the near future to map social cognition during awake craniotomy.

Trial Registration: ClinicalTrials.gov NCT03010943; <https://clinicaltrials.gov/ct2/show/NCT03010943> (Archived at WebCite at <http://www.webcitation.org/70CYDil0P>)

(*J Med Internet Res* 2018;20(6):e10332) doi:[10.2196/10332](https://doi.org/10.2196/10332)

KEYWORDS

virtual reality; neurosurgery; social cognition; awake surgery

Introduction

Social cognition includes all complex cognitive processes involved in social interaction such as nonverbal language (facial

and bodily nonverbal cues as affective prosody), empathy, and theory of mind (ToM). Following brain surgery, the impairment of nonverbal cue decoding, such as expression of facial emotions, eye gaze, body gestures, and prosody can lead to

ToM deficits. Patients often experience difficulties with understanding humor and as well as conceptualizing and understanding thoughts and beliefs, emotions, feelings and desires, behaviors, actions, and intentions of other people. In recent times, these sequelae were largely misunderstood by neurosurgeons and compared with postsurgical impairment of language or executive functions, with few evaluations concerning social cognition having been published [1-6].

As previously done for language, it is now possible to propose a substrate for social cognition based on parallel and large-scale interactive distributed brain networks [7]. However, unlike language, this substrate cannot be reliably localized based on anatomical criteria alone, mostly due to the individual variations. Individual brain mapping by direct electrical stimulation (DES) during awake craniotomy is therefore essential. The procedure has been well documented [8]. Briefly, it is possible to temporarily inactivate regions of the brain using DES, while patients perform neuropsychological tasks. If a patient shows decreased performance in a given task, the neurosurgeon will not remove these regions, so as to maintain brain function.

Compared with motor or language mapping, nonverbal language mapping has not been performed yet. This is due to the difficulties involved in adapting classic bedside tasks to awake surgery conditions.

In 2014, we started to explore the use of virtual reality (VR) during awake craniotomy with patients wearing a virtual reality headset (VRH). We previously developed an app for VRH to explore visuospatial cognition [9]. We are now performing a larger study evaluating the tolerance and safety of VRH and 3D immersive experiences in patients undergoing awake craniotomy and brain mapping by DES. Herein, we describe a VR experience, the interaction with an avatar using a social VR platform, and highlight its advantages, limitations, and future applications for perioperative mapping of social cognition.

Methods

This was a single-center, prospective, unblinded trial (ClinicalTrials.gov base identifier: NCT03010943), which was performed in compliance with all regulatory and ethical guidelines for clinical research. All patients signed a written informed consent.

The inclusion criterion was patients aged >18 years who were hospitalized for treatment of a tumor or any type of surgical lesion near the language region of the brain. The exclusion criteria were all contraindications to an awake surgery (cognitive impairment, aphasia, and morbid anxiety). The main objective was to assess procedural feasibility and safety.

This study was performed using a Samsung Gear VR combined with a Samsung S7 smartphone (android platform) and headphones. After general and local anesthesia, the patient was positioned lying on his side, with a rigid pin fixation of the head. Once the craniotomy was completed and the dura was opened, we awakened the patient. Electroencephalography signals were recorded using a subdural electrode. After the cortex was exposed, language mapping was performed by a neuropsychologist using an image denomination task on a digital

tablet. The mapping took place as previously described [10,11]. DES was applied with a bipolar electrode delivering a biphasic current (60Hz, 1 ms pulse width, current amplitude ranging from 2 to 8 mA over 2-3 s).

To prevent interference with the routine procedure of awake craniotomy and language brain mapping, we decided to duplicate the image naming task viewed in VRH (two dimensions, 2D) and then in stereoscopy (three dimensions, 3D; an app based on Unity 3D software with an interface allowing VRH communication via a computer and Bluetooth connection). Further VR experiences with a relaxing film were proposed at the end of the tumor resection while the wound was being closed. These options included interaction with the neuropsychologist's avatar in virtual locations; this option is the focus of this paper. For this experience, we used the vTime app, a social network in VR [12]. This app allows users to create an avatar and socialize with other people in virtual environments. The avatar can be piloted on a smartphone or in conjunction with a VRH using a game controller.

Results

A total of 3 patients used the vTime app during wound closure (2 males and one female; mean age, 54 years). Only 1 participant had a previous VR experience. Before the surgery, all the patients were trained without any issues.

Patients used a standard avatar and an account opened by the Department of Neurosurgery to preserve anonymity. They interacted with an avatar piloted by the neuropsychologist, who also wore a VRH, under the control of a physician who participated in the meeting and controlled the scene on a smartphone connected to the app. This allowed continual monitoring of the operation (Multimedia Appendix 1). The mean time of connection and interaction was 10 minutes.

During DES of the left inferior frontal gyrus (pars opercularis), all patients failed to perform the 2D and 3D language and motor tasks. All deficits disappeared when DES was stopped, and the patients were allowed to recover. The stimulated areas were not resected. Patients were neurologically intact.

During the social cognition experiences, the patients passively viewed the neuropsychologist's avatar and reproduced and commented on his gestures. Alternatively, they assumed more active roles, controlling their own avatars with a game controller in their hands (Multimedia Appendix 2).

Despite the discomfort associated with the awake surgery environment and other tasks completed with VRH, no patient experienced eye strain, nausea, or any sign of "VR disease." No seizures occurred while the patients looked at the VR experience.

Discussion

Principal Findings

As described previously, social cognition includes nonverbal language, empathy, and ToM. These functions are explored at the bedside by complex neuropsychological tasks batteries including story movies, comic strips, or interactive games that

depict a short story. These tasks require time to be performed, meaning they are not compatible with the brain mapping conditions (DES length inferior to 4 seconds, fast response, and no ambiguity in the answer).

VR approaches that allow interactions with an avatar are commonly used in cognitive neurosciences [13]. There is consistent evidence that avatars are perceived in a similar manner to real human beings and can be used to explore the complex processes of nonverbal language, empathy, and ToM [14]. In VR, the social interactions are governed by the same social norms as social interactions in the real world (for example social norms related to gender, interpersonal distance, and eye gaze) [15].

VR can imitate complex social situations, even for the patients undergoing awake craniotomies. The potential of VR lies in its increased real-life environment validity compared with screen-based studies. Rather than being a passive observer of stimuli on a computer screen, participants in virtual environments become part of the depicted scene. Although an increase in ecological validity often results in a decrease in experimental control, immersive VR has the potential to combine the naturalness of everyday interactions with experimental controls required during brain mapping procedures.

Instead of developing a specific app to test and map social cognition during awake brain surgery, we decided to test the potential of the available VR social networks. Several VR social platforms are already available, such as vTime [12], Oculus, Facebook Spaces, PLUTOVR, and AltspaceVR. Interestingly, these platforms take different approaches for conveying nonverbal language: arms, hands, head, and mouth movement and gaze.

For our trial, we chose Samsung VR, a low-cost, high-quality, customizable wireless device, with an optional pad control and a game controller. The VR social network vTime is compatible with the Samsung VRH [12]. The vTime app allows interaction with several avatars and positional control in different virtual environments. The avatar can point anywhere within the scene

and produce gestural expressions such as OK, Thumbs Up, Clap, Thumbs Down, Blow Kiss, etc [12]. It is also possible for the user to touch other avatars or to take control of his or her personal space.

We demonstrated that patients undergoing awake craniotomies can wear a VRH and interact with an avatar piloted by a neuropsychologist.

Limitations

We experienced some difficulties and limitations using vTime [12]. During an awake surgery, the patient is usually lying on his or her back or side with the head immobilized using a Mayfield skull clamp. The vTime app [12], as with most VR social experiences, is not well-adapted for use in this position and cannot make use of the 360 degree view. Further, there is no option to control the orientation of the virtual environment. In our experience, vTime can only be used with side-lying patients. Restrictions of neck, limbs, and face movements can affect psychological testing. If the patient does not lie on his or her side to visualize the neuropsychologist avatar, he or she cannot be tested. Despite this limitation, all awake surgeries were performed in our institution with the patient lying on his or her side. Mobility restrictions limit the patient's ability to explore the 360-degree environment, potentially limiting the use of vTime [12] to explore visuospatial cognition. Moreover, it is not possible to control facial expression and eye gaze, which are potent nonverbal language cues. An app dedicated to awake surgery is currently in development to overcome this limitation.

Conclusions

We showed that it is possible to use a VR social network during awake craniotomy and to test gesture communication. Progress in VR development is currently promising, and some VRHs even allow facial expressions to be captured and transferred to a virtual avatar in real time, opening a new level of virtual human interaction. We are convinced that these improvements could be applied to further research for awake craniotomy, nonverbal language, empathy, and ToM in the near future.

Acknowledgments

We thank Gwenaëlle Soulard (CHU, Angers) for her technical help.

Conflicts of Interest

None declared.

Multimedia Appendix 1

During awake brain surgery, the patient and the neuropsychologist (A) performing a language task; (B) Direct electrical stimulation and mapping of the cortex during the task; and (C) and (D) the patient and the neuropsychologist communicating with the VR social network.

[PNG File, 1MB - [jmir_v20i6e10332_app1.png](http://www.jmir.org/2018/6/e10332_app1.png)]

Multimedia Appendix 2

Video showing interaction between the patients and the neuropsychologist avatar using Vtime [12] in order to test social cognition during awake brain surgery.

[MP4 File (MP4 Video), 95MB - [jmir_v20i6e10332_app2.mp4](#)]

References

1. Richard-Mornas A, Mazziotti A, Koenig O, Borg C, Convers P, Thomas-Antérion C. Emergence of hyper empathy after right amygdalohippocampectomy. *Neurocase* 2014;20(6):666-670. [doi: [10.1080/13554794.2013.826695](#)] [Medline: [23944742](#)]
2. Happé F, Malhi GS, Checkley S. Acquired mind-blindness following frontal lobe surgery? A single case study of impaired 'theory of mind' in a patient treated with stereotactic anterior capsulotomy. *Neuropsychologia* 2001;39(1):83-90. [Medline: [11115657](#)]
3. Rowe AD, Bullock PR, Polkey CE, Morris RG. "Theory of mind" impairments and their relationship to executive functioning following frontal lobe excisions. *Brain* 2001 Mar;124(Pt 3):600-616. [Medline: [11222459](#)]
4. Milesi V, Cekic S, Péron J, Frühholz S, Cristinzio C, Seeck M, et al. Multimodal emotion perception after anterior temporal lobectomy (ATL). *Front Hum Neurosci* 2014;8:275 [FREE Full text] [doi: [10.3389/fnhum.2014.00275](#)] [Medline: [24839437](#)]
5. Nakajima R, Yordanova YN, Duffau H, Herbet G. Neuropsychological evidence for the crucial role of the right arcuate fasciculus in the face-based mentalizing network: A disconnection analysis. *Neuropsychologia* 2018 Jan 26:---. [doi: [10.1016/j.neuropsychologia.2018.01.024](#)] [Medline: [29360518](#)]
6. Herbet G, Lafargue G, Moritz-Gasser S, Menjot DCN, Costi E, Bonnetblanc F, et al. A disconnection account of subjective empathy impairments in diffuse low-grade glioma patients. *Neuropsychologia* 2015 Apr;70:165-176. [doi: [10.1016/j.neuropsychologia.2015.02.015](#)] [Medline: [25687031](#)]
7. Bernard F, Lemée J, Ter MA, Menei P. Right hemisphere cognitive functions: from clinical and anatomical bases to brain mapping during awake craniotomy Part I: Clinical and functional anatomy. *World Neurosurg* 2018:---.
8. Fernández CA, Moritz-Gasser S, Martino J, Martinoni M, Matsuda R, Duffau H. Selection of intraoperative tasks for awake mapping based on relationships between tumor location and functional networks. *J Neurosurg* 2013 Dec;119(6):1380-1394. [doi: [10.3171/2013.6.JNS122470](#)] [Medline: [24053503](#)]
9. Mazerand E, Le RM, Hue S, Lemée J, Klinger E, Menei P. Intraoperative Subcortical Electrical Mapping of the Optic Tract in Awake Surgery Using a Virtual Reality Headset. *World Neurosurg* 2017 Jan;97:424-430. [doi: [10.1016/j.wneu.2016.10.031](#)] [Medline: [27756671](#)]
10. Ojemann G, Ojemann J, Lettich E, Berger M. Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. *J Neurosurg* 1989 Sep;71(3):316-326. [doi: [10.3171/jns.1989.71.3.0316](#)] [Medline: [2769383](#)]
11. Duffau H, Capelle L, Denvil D, Sichez N, Gatignol P, Taillandier L, et al. Usefulness of intraoperative electrical subcortical mapping during surgery for low-grade gliomas located within eloquent brain regions: functional results in a consecutive series of 103 patients. *J Neurosurg* 2003 Apr;98(4):764-778. [doi: [10.3171/jns.2003.98.4.0764](#)] [Medline: [12691401](#)]
12. vTime: The VR Sociable Network - Out Now for Windows Mixed Reality, Gear VR, Oculus Rift, iPhone, Google Daydream, and Google Cardboard Internet URL: <https://vtime.net/> [accessed 2018-05-11] [WebCite Cache ID 6zKYC8j6Q]
13. Georgescu AL, Kuzmanovic B, Roth D, Bente G, Vogeley K. The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Front Hum Neurosci* 2014;8:807 [FREE Full text] [doi: [10.3389/fnhum.2014.00807](#)] [Medline: [25360098](#)]
14. de BAW, de GB. Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Front Psychol* 2015;6:576 [FREE Full text] [doi: [10.3389/fpsyg.2015.00576](#)] [Medline: [26029133](#)]
15. Yee N, Bailenson JN, Urbanek M, Chang F, Merget D. The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. *Cyberpsychol Behav* 2007 Feb;10(1):115-121. [doi: [10.1089/cpb.2006.9984](#)] [Medline: [17305457](#)]

Abbreviations

DES: direct electrical stimulation
RCT: randomized controlled trial
ToM: theory of mind
VR: virtual reality
VRH: virtual reality headset

Edited by G Eysenbach; submitted 07.03.18; peer-reviewed by L Moscote, A Thapa; comments to author 10.05.18; revised version received 14.05.18; accepted 19.05.18; published 26.06.18.

Please cite as:

Bernard F, Lemée JM, Aubin G, Ter Minassian A, Menei P

Using a Virtual Reality Social Network During Awake Craniotomy to Map Social Cognition: Prospective Trial

J Med Internet Res 2018;20(6):e10332

URL: <http://www.jmir.org/2018/6/e10332/>

doi: [10.2196/10332](https://doi.org/10.2196/10332)

PMID: [29945859](https://pubmed.ncbi.nlm.nih.gov/29945859/)

©Florian Bernard, Jean-Michel Lemée, Ghislaine Aubin, Aram Ter Minassian, Philippe Menei. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.06.2018. 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.

Original Paper

A Deep Learning Method to Automatically Identify Reports of Scientifically Rigorous Clinical Research from the Biomedical Literature: Comparative Analytic Study

Guilherme Del Fiol^{1*}, MD, PhD; Matthew Michelson^{2,3*}, PhD; Alfonso Iorio^{4,5}, MD, PhD; Chris Cotoi⁶, BEng, EMBA; R Brian Haynes^{4,5}, MD, PhD

¹University of Utah, Department of Biomedical Informatics, Salt Lake City, UT, United States

²Evid Science, Los Angeles, CA, United States

³InferLink Corporation, Los Angeles, CA, United States

⁴Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

⁵Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

⁶Health Information Research Unit, McMaster University, Hamilton, ON, Canada

*these authors contributed equally

Corresponding Author:

Guilherme Del Fiol, MD, PhD

University of Utah

Department of Biomedical Informatics

421 Wakara Way

Suite 140

Salt Lake City, UT, 84108

United States

Phone: 1 8015814080

Email: guilherme.delfiol@utah.edu

Abstract

Background: A major barrier to the practice of evidence-based medicine is efficiently finding scientifically sound studies on a given clinical topic.

Objective: To investigate a deep learning approach to retrieve scientifically sound treatment studies from the biomedical literature.

Methods: We trained a Convolutional Neural Network using a noisy dataset of 403,216 PubMed citations with title and abstract as features. The deep learning model was compared with state-of-the-art search filters, such as PubMed's Clinical Query Broad treatment filter, McMaster's textword search strategy (no Medical Subject Heading, MeSH, terms), and Clinical Query Balanced treatment filter. A previously annotated dataset (Clinical Hedges) was used as the gold standard.

Results: The deep learning model obtained significantly lower recall than the Clinical Queries Broad treatment filter (96.9% vs 98.4%; $P<.001$); and equivalent recall to McMaster's textword search (96.9% vs 97.1%; $P=.57$) and Clinical Queries Balanced filter (96.9% vs 97.0%; $P=.63$). Deep learning obtained significantly higher precision than the Clinical Queries Broad filter (34.6% vs 22.4%; $P<.001$) and McMaster's textword search (34.6% vs 11.8%; $P<.001$), but was significantly lower than the Clinical Queries Balanced filter (34.6% vs 40.9%; $P<.001$).

Conclusions: Deep learning performed well compared to state-of-the-art search filters, especially when citations were not indexed. Unlike previous machine learning approaches, the proposed deep learning model does not require feature engineering, or time-sensitive or proprietary features, such as MeSH terms and bibliometrics. Deep learning is a promising approach to identifying reports of scientifically rigorous clinical research. Further work is needed to optimize the deep learning model and to assess generalizability to other areas, such as diagnosis, etiology, and prognosis.

(*J Med Internet Res* 2018;20(6):e10281) doi:[10.2196/10281](https://doi.org/10.2196/10281)

KEYWORDS

information retrieval; evidence-based medicine; deep learning; machine learning; literature databases

Introduction

Background and Significance

With roughly 95 clinical trials published per day, the biomedical literature is increasing at a very rapid pace, imposing a significant challenge to the practice of evidence-based medicine. However, only 1% of studies in the biomedical literature meet minimum criteria for scientific quality [1] and most published research findings are eventually shown to be false [2]. As a result, a major barrier to the practice of evidence-based medicine is efficiently finding the relatively small number of scientifically sound studies on a given clinical topic. Systematic reviews and meta-analyses attempt to summarize the available evidence on a given clinical question aiming for near perfect recall. However, systematic reviews are often not available and become quickly outdated. Therefore, clinicians may benefit from access to the latest evidence from high-quality clinical trials before they are included in systematic reviews.

For over two decades, the *Clinical Query* filters have been the state-of-the-art approach to retrieve scientifically sound clinical studies from the primary literature, both for the development of systematic reviews and point-of-care decision support [3,4]. The Clinical Query filters consist of Boolean search strategies based on textwords and Medical Subject Headings (MeSH) terms that have been developed and validated through a systematic approach [5]. The search textwords and MeSH terms used in the Clinical Query filters reflect widely accepted criteria for scientifically sound clinical studies, such as “clinical trial,” “random allocation,” and “randomized controlled trial [Publication Type].” Although initially developed in the 1990s, the Clinical Query filters have been updated over time and the recall and precision of the filters developed in 2000 did not significantly change a decade later [6]. Clinical Query filters for several topics are available in PubMed and several other bibliographic biomedical databases, with focuses on areas such as therapy, diagnosis, etiology, and prognosis, and these are tuned for precision or recall. A limitation of the Clinical Query filters is their dependency on MeSH terms, which are added to PubMed citations 23 to 177 days after an article is published (according to a previous study [7]) and 17 to 328 days according to our more recent analysis. In addition, there is room for improvement, especially in terms of retrieval precision.

Previous studies investigated the use of machine learning approaches to automate the retrieval of scientifically sound studies [8-10]. Features used in those studies included bibliometrics (eg, citation count, impact factor), words in the article title and abstract, MeSH terms, Unified Medical Language System (UMLS) concepts, and semantic predications. Although the results of machine learning studies were promising, they had important limitations that precluded wide adoption in practice, such as a requirement for significant feature engineering (eg, UMLS concepts, bibliometrics), reliance on proprietary and time-dependent features (eg, MeSH index, citation counts), and potential overfitting to a particular dataset.

In the present study, we investigated a deep learning approach for the retrieval of scientifically sound treatment studies from PubMed. To overcome limitations of previous methods, we

focused on an approach that requires very little feature engineering and does not rely on proprietary or time-dependent features. We then compared the performance of a deep learning model with state-of-the-art PubMed search strategies against Clinical Hedges, a rigorous gold standard of over 50,000 studies that were systematically rated for scientific quality according to rigorous criteria [5].

Deep Machine Learning

Recent advances in machine learning have led to dramatic improvements in the abilities of computers to mimic human activities. Many of these improvements leverage “deep learning,” and embody neural-networks with many nodes that are fully connected across layers of the network. In the context of supervised deep learning, which we utilized here, such a network is trained by providing many examples of the objective to classify, as well as many counter examples.

Deep Neural Networks

A Deep Neural Network (DNN) is a fully connected set of “layers,” each of which contains a node that encodes information in the form of a weight associated with a particular feature of the input data. By “connected” we mean that the nodes of each layer connect with the nodes of the next. A DNN is considered “deep” because it can contain many such connected nodes and/or layers, thereby encoding a significant amount of information in the weights applied to the input of each layer.

In the case of text categorization, the input to the network is a set of words (or “word embeddings” described below). Each successive layer of the DNN applies some transformation to the words in the form of linear algebraic operations that progressively encodes more granular features of the data [11,12]. A supervised DNN, such as our approach, requires that each input (eg, set of words) is associated with a class such that the DNN will learn how to associate the words with each class in order to predict the class for newly unseen sets of words. As with most machine learning approaches, the input text can be transformed in a number of ways. In the case of text classification, such transformations could include adding extraneous information such as bibliographic and author information. This process of designing and applying features to optimize classification is known as *feature engineering*.

Although potentially useful, feature engineering is challenging: it may require significant manual effort and introduces the risk that certain features will be too specific to the training data or may even be unavailable. As we discuss below, leveraging the MeSH terms used to index articles in PubMed can certainly help in a task such as ours, but there is no guarantee that such information is available for an article in a timely manner.

Therefore, we opted for an “end-to-end” machine solution. In end-to-end approaches, the DNN is trained solely on the inputs and classes with minimal or no feature engineering. Minimal features are those that are task- and domain-agnostic, such as converting words to lower case, removing stopwords, and stemming. Potential advantages of such an approach include: (1) simpler design, therefore strong results are more likely indicative that the DNN is detecting textual signal, rather than an arcane feature; (2) no reliance on external factors, such as

features that may not be timely available; and (3) mitigation of concept drift, since the training features may misalign from those available when a model is deployed. Therefore, end-to-end systems provide a strong justification for a first approach in classification tasks.

Recurrent Neural Networks and Convolutional Neural Networks

In this study, we utilized a particular deep-learning neural-network known as Convolutional Neural Network (CNN), following the approach of Kim [13]. To some extent, Kim's CNN architecture has become a *de facto* standard for text classification. CNNs analyze text using sliding word windows of specified sizes. Each sliding word window generates a set of real-valued vectors. Generally, each word or even character is associated with a "word-embedding," which is a low-dimension real-valued vector that represents the semantic space for the word [14]. Therefore, as each term is associated with a vector, each sliding word window then represents a matrix. Each sliding word window is then passed through an activation function, and a "max pooling" is applied such that only the maximum value is kept from the set of values produced by the activation function, as applied to the window. That is, each window is associated with its single, maximal value outputted by the activation function. These maximal values are concatenated together to form their own vector representing the set of windows. This set of concatenated values forms the next layer, which is then passed to the final layer, which includes the decision-making activation function (such as Softmax, as described below).

An example of a CNN is shown in Figure 1. From left to right, we see one set of input words and their word embeddings, which forms the initial input matrix. This network uses two sets of sliding windows, one of size two and one of size three. These sets of sliding windows produce the convolutional layer, transforming the sliding window's features into new feature values, which are then pooled such that only the maximum value is kept (the "max pooling"). Finally, the max-pooled values are passed through the fully connected final (output) layer, which uses Softmax to assign a probability of class membership (shown as "yes" or "no" for binary class membership). While this approach may appear "shallow," it has been shown to be effective, becoming one of the most popular architectural choices for CNN [13].

Another popular approach for text-analysis tasks are Recurrent Neural Networks (RNNs). In contrast to CNN's sliding-windows, which treat phrases somewhat independently, RNNs are well suited for language tasks where the classification of a particular piece of text depends on the surrounding text. For instance, RNNs are well suited for part-of-speech tagging or machine translation, which have a strong dependency on the particular word order. However, because they must consider order dependencies, they are not as appropriate for tasks such as ours. In fact, in a head-to-head comparison between CNNs and RNNs for natural language processing tasks, Yin et al [15] found that CNNs are particularly well suited for so called

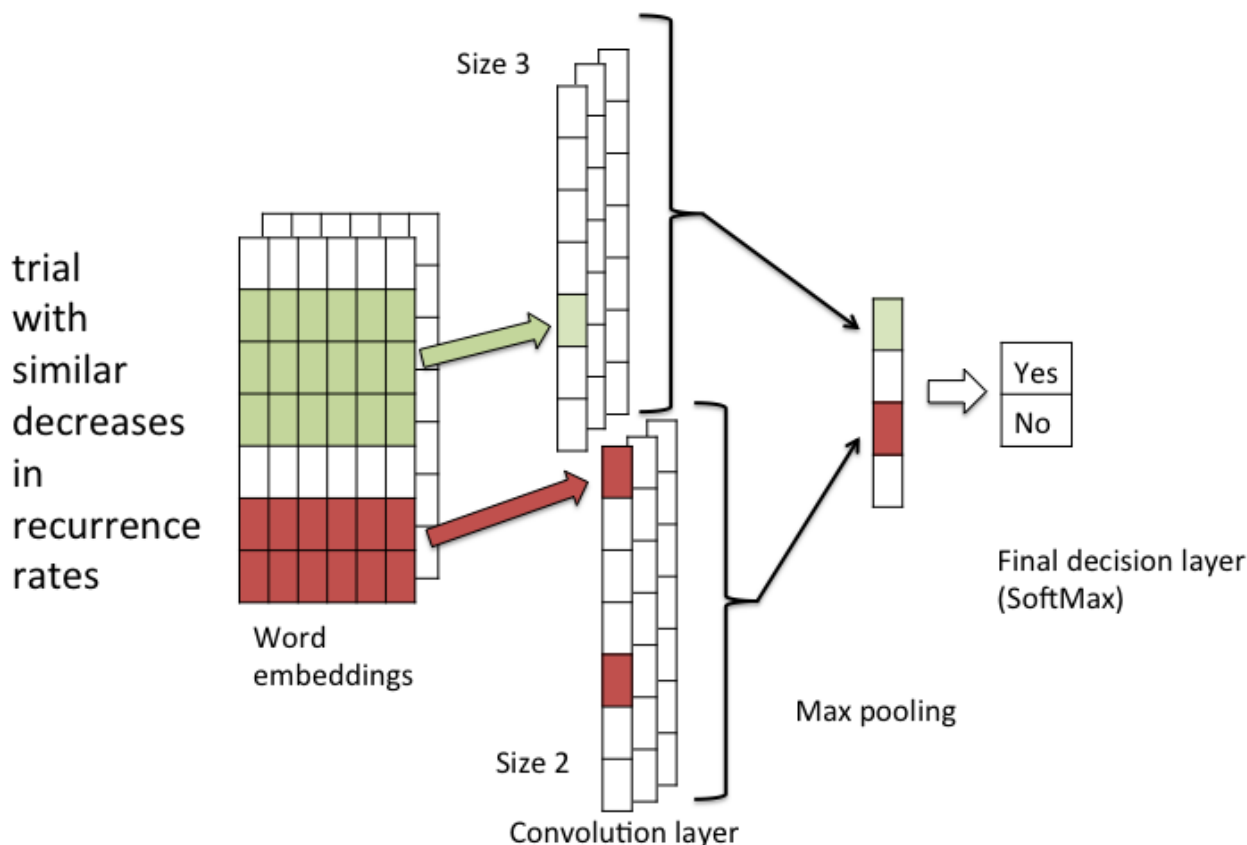
"keyphrase recognition" tasks such as text classifications or sentiment analyses. Furthermore, CNNs were found to be up to five times faster than RNNs [15], which is important in real-world tasks such as ours where the goal is to classify an extremely large corpus, such as PubMed, in a reasonable amount of time.

Deep Neural Network Optimization (How It Learns)

The main learning for a neural network involves "forward propagation" and "backward propagation." In forward propagation, inputs are translated into features by transforming the inputs into real-valued vectors of fixed sizes. These vectors (eg, "layers") are combined with weights and passed through an activation function that summarizes the contribution of each feature of the vector and its weight. Layers are connected to one another such that the values from the activation function of the current layer become the inputs to the next layer. Therefore, the "forward propagation" starts with input and passes activation values from layer to layer until the final layer, which outputs some decision vector. In our case, this final output function is a sigmoid activation function, which can assign probability to class membership. In "backward propagation" the final classification decision is compared with the known result from the training data and errors are propagated backward through the network, from the output layer to the input layer. Each weight is updated according to its contribution to the decision accuracy via gradient descent.

In the context of CNN, one can interpret the various passes through "forward" propagation as applying weights to different "chunks" of the text input, and "backwards" propagation as adjusting those weights to make the fewest errors in predicting the class of the input text. Within the context of DNN, since optimization is essentially a weight adjustment process, the higher the number of nodes and layers, the more weights must be adjusted to find the optimal classifier, which requires more training data. Conversely, more weights and layers may improve classification. Therefore, part of DNN design is to identify optimal parameter choices and how to deal with overfitting. In our case, we used a technique called dropout regularization, which randomly prevents nodes from participating in a classification decision for a given training input, so the model does not overfit by learning to simply rely on a particular node.

Other optimizations include which mathematical operations to choose for the propagation; this is called the "activation function" (ie, how a node produces a score given the weight and input). Different choices can result in different DNN behavior; some activation functions are more robust than others, while some can make the training process exceedingly long. We chose the Rectified Linear Unit (ReLU) for our activation function, as it provides an efficient mechanism to build robust and accurate CNNs. The choice of ReLU is quite common in tasks such as ours. Finally, within the context of CNN, it is common to provide a down-sampling between layers, which helps control overfitting and makes training more efficient. The most common approach is max pooling, which we use in our approach.

Figure 1. Example of a Convolutional Neural Network.

Therefore, the training of a network involves multiple passes of forward propagation followed by backward propagation. It is common to call each iteration over all the training data an “epoch.” The model generally stops this training process at a fixed number of epochs or when the metric of success appears to have reached some maximal value.

Clinical Hedges

Clinical Hedges is a database previously developed by the Hedges Group at McMaster University, used to develop and evaluate the Clinical Query filters [5] and previous machine learning approaches [8] that retrieve scientifically sound clinical studies from PubMed. The database has 50,594 articles published in 170 clinical journals. All articles were manually annotated by highly-calibrated information science experts according to type (eg, etiology, prognosis, diagnosis, prevention, therapy, clinical prediction) and whether or not each study met prespecified and experimentally validated methodological criteria for scientifically sound clinical research. The criteria and process used to rate the articles in Clinical Hedges are described elsewhere [5]. In summary, criteria for scientifically sound studies on treatment interventions include random allocation of study participants, clinically relevant outcomes, and at least 80% follow-up of study participants.

Methods

Overview of the Approach

Overall, our approach consisted of (1) training and testing deep learning models with a large and noisy dataset obtained

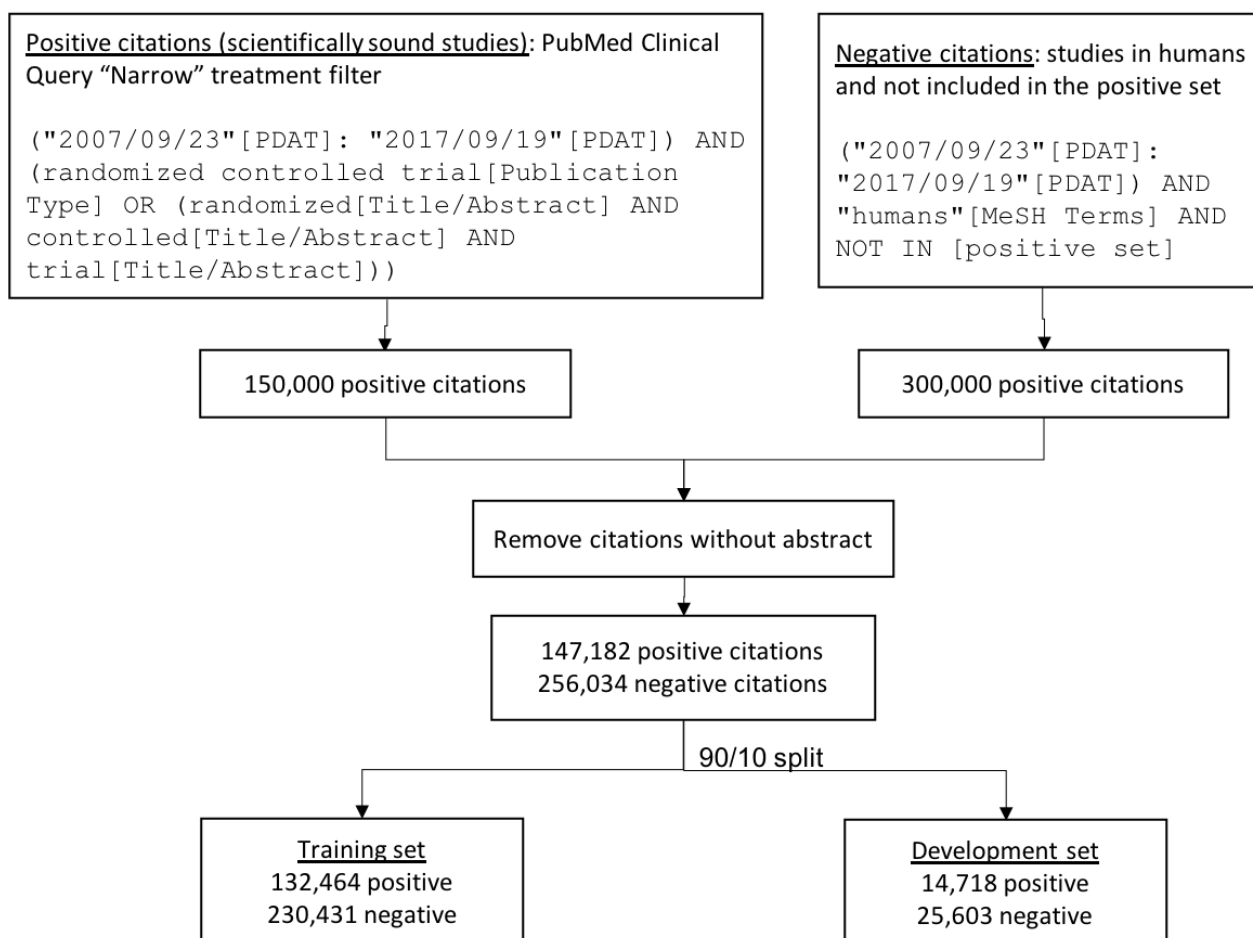
automatically through PubMed searches based on the Clinical Query treatment filter, and (2) evaluating the performance of the resulting model against Clinical Hedges as a gold standard.

Specifically, the study method consisted of the following steps, which are described in more detail in the sections below: (1) preparation of a dataset for training the deep learning models, (2) training and tuning deep learning models, (3) comparison of the deep learning approach with state-of-the-art search filters and McMaster’s textword filter in terms of precision and recall, and (4) analysis of deep learning performance in terms of precision at several levels of K retrieved citations.

Preparation of Training Dataset

The training/testing dataset consisted of 403,216 positive and negative citations retrieved from PubMed. To retrieve *positive studies* (ie, scientifically sound), we used the Clinical Queries treatment filter tuned for precision (“narrow” filter; Figure 2). In previous studies, this filter yielded 93% recall and 54% precision for scientifically sound treatment studies in the Clinical Hedges gold standard [5]. Therefore, this search strategy was used as a surrogate for retrieving a large dataset of scientifically sound studies that are similar to the ones in the Clinical Hedges gold standard. Although this approach produced a rather noisy training set (close to half of the positive samples were false-positives), the CNN approach is resilient to handle noisy data as long as there is sufficient training data. To retrieve *negative studies* (ie, not scientifically sound), we retrieved studies conducted in humans which were not retrieved by the “positive” search strategy above.

Figure 2. Datasets used for training and testing the deep learning models. The PubMed Clinical Query “Narrow” treatment filter was used as a surrogate to identify positive (scientifically sound) studies. The resulting dataset was split into training and development sets using a 90/10 ratio.



The strategies were limited to retrieve a maximum of 150,000 and 300,000 citations respectively to yield a dataset with one third positive and two thirds negative citations. Both strategies were limited to citations published between 2007 and 2017. Citations without an abstract were removed. The search strategies were executed with PubMed’s eUtils application program interface. The resulting dataset contained 147,182 positive and 256,034 negative citations (Figure 2).

Training and Tuning Deep Learning Models

Deep learning models were trained using 90% of the citations in the dataset, with the remaining 10% used as a “development” set (Figure 2). As the training/development split was randomly generated, the development set maintained a similar proportion of positive to negative instances as the training set. To build model inputs, we concatenated the title with the abstract, removed stopwords, and kept the first 650 tokens of the remaining words.

As mentioned in the *Deep Machine Learning* section, our model follows the well-accepted approach of applying CNNs for text classification. The first layer applies character embedding to the words, so that words outside of the known vocabulary can be included for prediction. The character embeddings are then combined with word embeddings (built from the training data), to capture semantic similarity. This input is passed into our model, which contains two convolutional layers: one for sliding

word windows of size two and one for word windows of size three. Each convolutional layer contains 512 filters associated with it. We apply a ReLU unit to the convolutional layers and pass them through a max pooling procedure. The resulting max-pooled features are then concatenated into a single layer. The max-pooled layer is passed to the next layer which consists of 512 units (fully connected), to which we apply a Softmax activation function to predict the probability of a citation belonging to either class. We then take the Argmax of the Softmax predictions as the predicted class. We ran this model with dropout regularization of 0.5 (to prevent overfitting) for 30 epochs. Hyper-parameters were chosen experimentally based on maximized precision on the training data.

Comparison of the Deep Learning Approach With State-of-the-Art PubMed Search Strategies

We tested three hypotheses that reflect the requirements imposed by different information retrieval scenarios. The first scenario consisted of search strategies to support the development of evidence-based syntheses, such as systematic reviews and clinical guidelines [16]. In this scenario, there is a requirement for near perfect recall. The hypothesis for this scenario was that the deep learning approach would yield equivalent recall with higher precision for scientifically sound treatment studies compared with the PubMed Clinical Queries Broad filter, which has almost perfect recall (Figure 3).

Figure 3. Search strategies used to retrieve scientifically sound treatment studies in comparison with the deep learning model.**Hypothesis #1 (evidence synthesis - maximize recall) - Clinical Query Treatment Broad filter**

```
((clinical[Title/Abstract] AND trial[Title/Abstract]) OR
clinical trials as topic[MeSH Terms] OR clinical
trial[Publication Type] OR random*[Title/Abstract] OR random
allocation[MeSH Terms] OR therapeutic use[MeSH Subheading])
```

Hypothesis #2 (literature surveillance - no MeSH terms, maximize recall) - McMaster's textword search

```
hasabstract[text] AND (effect*[Title/Abstract] OR
control[Title/Abstract] OR controlled[title/abstract] OR
random*[Title/Abstract] NOT (qualitative[title/abstract] OR
retrospective[title/abstract] OR mice[title/abstract] OR
rat[title/abstract] OR rats[title/abstract] OR editorial[pt]
OR letter[pt])
```

Hypothesis #3 (patient care decision making - balance between precision and recall) - McMaster's Balanced Clinical Query filter

```
((clinical[Title/Abstract] AND trial[Title/Abstract]) OR
random*[Title/Abstract])
```

The second scenario reflects the need to retrieve recent studies, such as in literature surveillance efforts to identify new evidence to update existing systematic reviews and clinical guidelines [17-19]. Since Clinical Query filters depend partially on MeSH terms and publication type, they are less effective for literature surveillance. Instead, search strategies based on terms in the citation title and abstract are preferred. The hypothesis for this scenario was that the deep learning approach would yield equivalent recall but higher precision for scientifically sound treatment studies compared with a textword search strategy provided by the Clinical Hedges group from McMaster University (Figure 3).

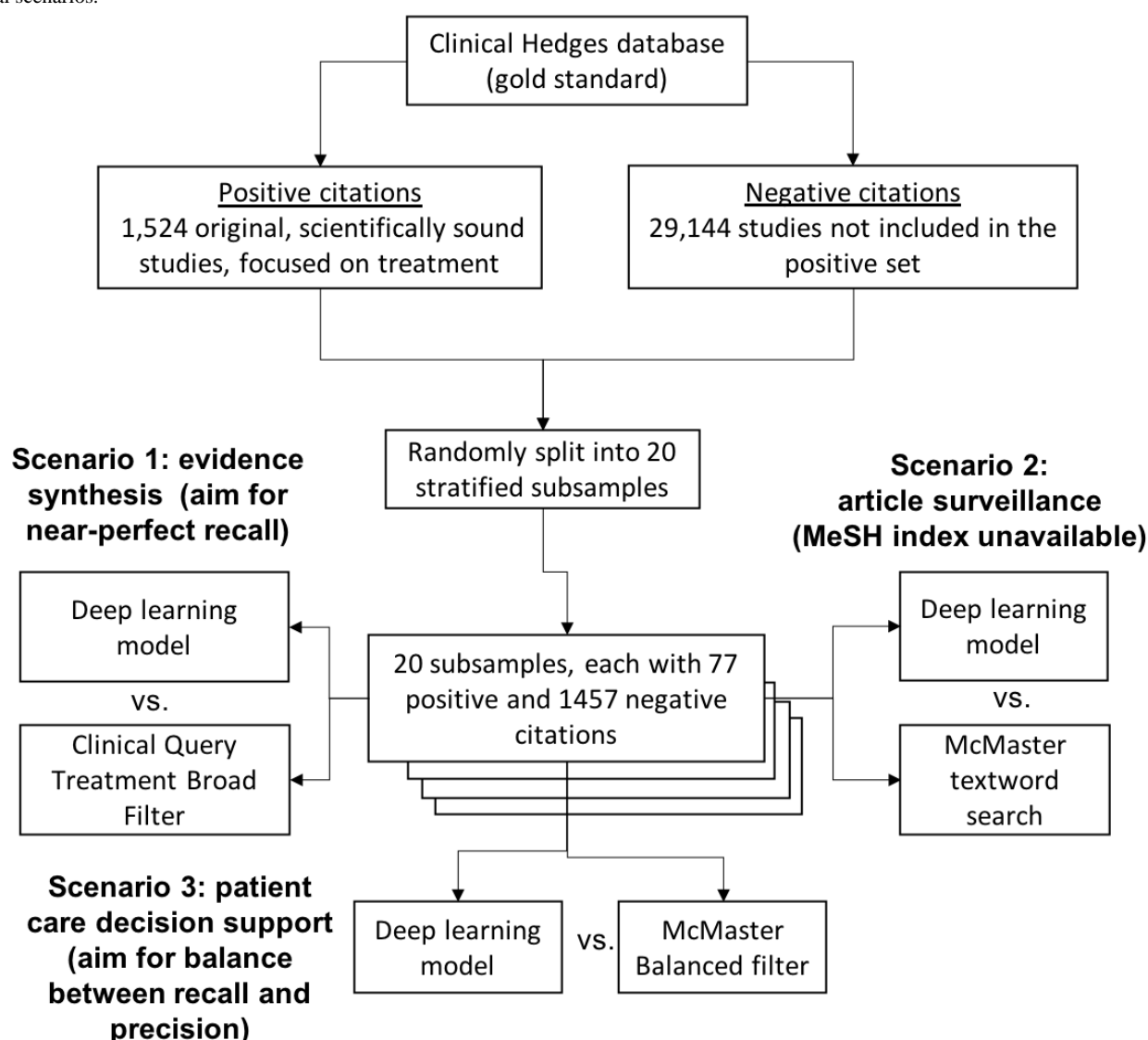
The third scenario represents clinicians searching the literature for evidence to meet clinicians' information needs that are raised in the care of a specific patient [20]. In this scenario, trading a small loss in recall for substantial gains in precision is acceptable. We hypothesized that the deep learning approach would yield equivalent recall but higher precision for scientifically sound treatment studies compared with McMaster's Balanced Clinical Query filter, which uses a combination of textwords, MeSH terms, and publication types (Figure 3).

The Clinical Hedges gold standard was used to test the three hypotheses. For positive citations, we retrieved 1524 original scientifically sound studies, with a focus on treatment, from the Clinical Hedges database. For negative citations, we retrieved 29,144 treatment studies from Clinical Hedges that were not in the positive set. For statistical analyses, we split the resulting dataset into 20 random subsamples, which were stratified to ensure a balanced ratio of positive and negative citations in each subsample. Measures of precision, recall, and F-measure were obtained for the four approaches on each of the 20 subsamples (Figure 4). Last, we ranked the output of the deep learning model according to its probability score and obtained measures of precision at several levels of top K citations (10, 20, 50, 100, 200, 300, and 500).

Statistical Analysis

Classification performance was measured according to the average precision and recall across 20 data samples. We used the paired Student t-test to test the significance of the differences in recall and precision between the two approaches in each experiment, with the significance level set at 0.05.

Figure 4. Evaluation method, including comparisons between the deep learning approach and Boolean searches focused on three different information retrieval scenarios.



Results

The results are organized according to the three information retrieval scenarios and study hypotheses.

Scenario 1 - Development of Evidence-Based Syntheses

Table 1 shows the results of the comparisons for Scenario 1, which requires near perfect recall. We tested the hypothesis that *the deep learning approach yields equivalent recall with higher precision for scientifically sound treatment studies compared with the PubMed Clinical Queries Broad filter*. The Clinical Queries Broad filter had statistically significantly higher recall than the deep learning model (98.4% vs 96.9%; $P=.002$), although the difference was small (-1.6%) and likely marginal in practice, depending on the use case. The deep learning model had significantly higher precision than the Clinical Queries Broad filter, with a +12.2% absolute difference (34.6% vs 22.4%; $P<.001$).

Scenario 2 - Literature Surveillance

Table 2 shows the results of the comparisons for Scenario 2, which requires retrieval of recent studies prior to MeSH indexing. We tested the hypothesis that *the deep learning approach yields equivalent recall but higher precision for scientifically sound treatment studies compared with a textword search strategy*. The deep learning model was equivalent to McMaster's textword search in terms of recall (97.1% vs 96.9%; $P=.57$); and had significantly higher precision than the textword search (34.6% vs 28.5%; $P<.001$).

Scenario 3 - Patient Care Decision Support

Table 3 shows the results of the comparisons for Scenario 3, in which trading a small loss in recall for gains in precision is acceptable. We tested the hypothesis that *the deep learning approach yields equivalent recall but higher precision for scientifically sound treatment studies compared with McMaster's Balanced Clinical Query filter*. Compared with the McMaster Balanced treatment filter, the deep learning model had similar recall (96.9% vs 97.0%; $P=.63$), but lower precision (34.6% vs 40.9%; $P<.001$; Table 3).

Precision at K

The precision at K curve for the ranked output of the deep learning model showed that precision ranged from 75.5% to

61% among the top 10 to top 100 citations and only decreased substantially after the top 200, 300, and 500 citations (Figure 5).

Table 1. Average recall, precision, and F-measure of the deep learning model and Clinical Query Broad filter according to the Clinical Hedges gold standard (N=20).

Parameter	Deep learning (%)	CQ ^a broad (%)	P value
Recall	96.9	98.4	<.001
Precision	34.6	22.4	<.001
F-measure	51.0	36.5	<.001

^aCQ: PubMed Clinical Query Treatment filter

Table 2. Average recall, precision, and F-measure of the deep learning model and McMaster's textword search according to the Clinical Hedges gold standard (N=20).

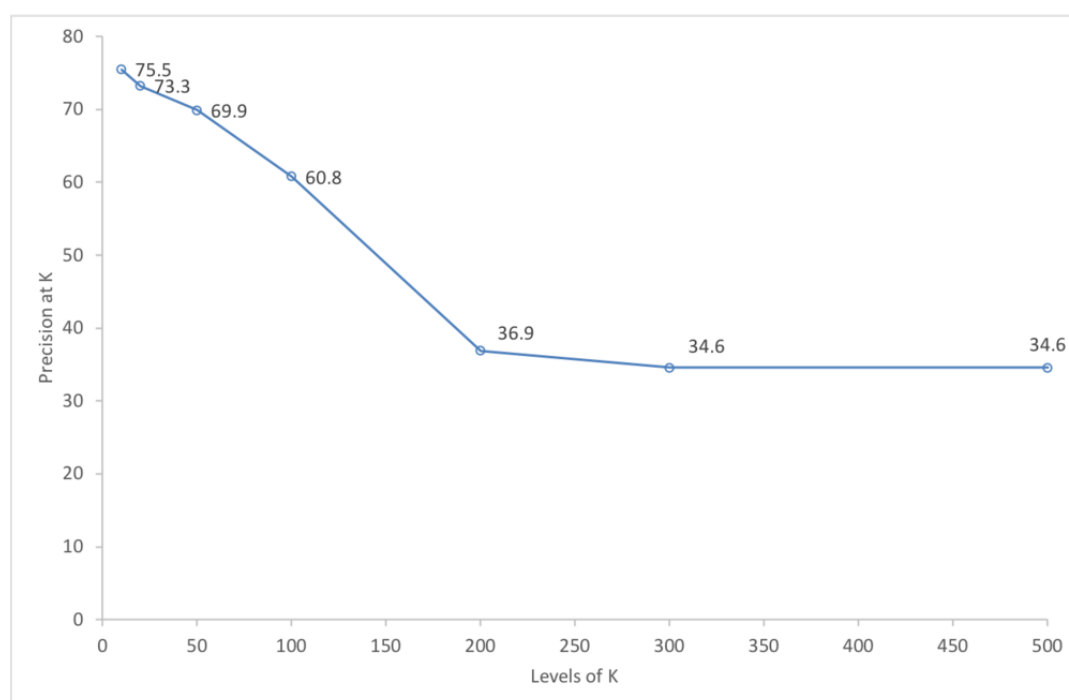
Parameter	Deep learning (%)	Textword search (%)	P value
Recall	96.9	97.1	.57
Precision	34.6	11.8	<.001
F-measure	51.0	21.0	<.001

Table 3. Average recall, precision, and F-measure of the deep learning approach and McMaster's Balanced Treatment filter according to the Clinical Hedges gold standard (N=20).

Measure	Deep learning (%)	McMaster's CQ ^a balanced filter (%)	P value
Recall	96.9	97.0	.63
Precision	34.6	40.9	<.001
F-measure	51.0	57.5	<.001

^aCQ: PubMed Clinical Query Treatment filter

Figure 5. Average precision of the deep learning model at different levels of top K citations.



Discussion

Significant Findings

To our knowledge, this is the first study to investigate the use of deep learning techniques to identify reports of scientifically sound studies in the biomedical literature in three different information-seeking scenarios. The deep learning approach performed reasonably well compared with state-of-the-art search filters, especially for literature surveillance. For evidence synthesis, the deep learning approach had slightly lower recall (-1.6%), but significantly higher precision (+12.2%) than the PubMed Clinical Query Broad treatment filter. For literature surveillance, the deep learning approach had equivalent recall to McMaster's textword filter, but significantly higher precision (+22.2%). For patient care decision-making, the deep learning model had similar recall, but lower precision (-6.3%) than McMaster's Balanced filter. Strengths of the study methodology include the use of a very large training set, comparison with state-of-the-art search strategies, and evaluation with a rigorous gold standard which was completely independent from the training set.

The proposed deep learning approach has three main potential benefits compared with previous approaches. First, unlike previous machine learning approaches, which depend on features that are not always openly and contemporaneously available (eg, MeSH terms, citation counts, journal impact factors), the proposed deep learning approach only uses citation title and abstract, which are available as soon as citations are entered in PubMed. Although full-text articles could be added as features in an attempt to improve performance, obtaining access to the full-text of all articles indexed in PubMed is impracticable since most journals do not provide open access to full-text. To assess the potential duration of delays for literature surveillance strategies based on MeSH filters, we determined the time between the date of creation of the article record in PubMed (CRDT) and the date of posting of MeSH terms (MHDA) for 107 journals (55,237 articles) in the McMaster PLUS database, from which the Clinical Hedges database was derived. The mean delay in MeSH indexing per journal was 162 days (95% CI 157-167), with a range of 17 to 328 days. Indexing intervals for journals were inversely correlated with journal impact factors (for 2016), but the correlation was relatively weak (-0.38; CI -0.199 to -0.517). As a second benefit, the deep learning model provides a ranked output with 70% or higher precision among the top 50 citations. This feature could be particularly useful for clinicians in busy clinical settings who are less likely to look beyond the top 20 citations that are displayed in PubMed searches [20,21]. In addition, citation ranking could help with systematic review development, since front-loading "eligible" citations can be used to help train and calibrate citation screeners and prioritize work [22]. Third, the deep learning model obtained reasonable performance despite being trained on a noisy dataset (an estimate of roughly 50% of the positive cases were false-positives). This finding confirms the robustness of the deep learning approach, which is known to be resilient to noisy training data [23].

Comparison With Prior Work

Previous work applied deep learning to classification tasks in the biomedical informatics domain. Lee [24] classified sentences as belonging to papers that would be included in a systematic review, or those that would not. However, because they did not employ a large-scale training procedure, as we devised here, their results were poor. It is also not clear whether the author focused solely on sentence classification, or document classification, as in our work [24]. Hughes et al [25] applied CNNs to classify sentences according to one of 26 categories, such as "Brain" or "Cancer," using a similar approach (though a different training procedure) to a different problem. Wang et al [26] used word, dependency, and abstract meaning representation embeddings to extract information on drug-drug interactions from the biomedical literature. Both Nguyen et al [27] and Che et al [28] utilized CNNs to predict risk outcomes, such as hospital readmission, using electronic health record data as an input. As with Hughes et al [25], although applied to different problems, the latter studies demonstrated precedent for using CNN in biomedical text classification.

A polynomial Support Vector Machine classifier based on MeSH terms, publication type, and title/abstract words obtained a recall of 96% and precision of 18% against a gold standard of internal medicine articles included in the American College of Physicians Journal Club [9]. A different study compared Clinical Query filters, machine learning, and algorithms based on citation count and the PageRank algorithm using a gold standard of important literature on common problems in surgical oncology [10]. The PageRank algorithm obtained a precision at the top 10, 20, 50, and 100 citations of 7.8%, 13.0%, 19.9%, and 26.3%, respectively [10]. Overall precision and recall were not reported. More recently, a study by Kilicoglu et al [8] investigated a set of classifiers using features such as MeSH terms, title/abstract words, UMLS concepts, and semantic predications. A Naïve Bayes classifier with these features obtained a recall and precision of 91.4% and 52.5% for treatment studies in the Clinical Hedges database [8]. As discussed above, those previous approaches relied on substantial feature engineering and/or proprietary and time-sensitive features, compromising the use of those approaches in real-time information retrieval systems. In a recent study investigating an approach similar to ours, Marshall et al [29] developed CNN and support vector machine classifiers based on article title and abstract to identify reports of randomized controlled trials (RCTs). The best classifier obtained a recall of 98.5% and precision of 21% [29]. Although the authors also evaluated their classifiers against the Clinical Hedges database, the results cannot be directly compared with our study because their goal was to identify RCTs versus scientifically sound studies (not all RCTs are scientifically sound and not all scientifically sound studies are RCTs). Another difference was that Marshall et al [29] used a training set derived from RCTs identified in Cochrane systematic reviews while we used a dataset obtained using the Clinical Queries Treatment Narrow filter.

Error Analysis

We analyzed a random sample of 20 false-negatives and 20 false-positives identified by the deep learning model. The

majority of the false-negatives (16/20) were likely due to the lack of an explicit description of the study design in the article abstract, which led the deep learning model to miss these articles. Of the 20 false-negatives, the Clinical Query Broad filter was able to correctly identify 14 articles based on MeSH terms and publication type rather than words in the abstract or title. Two approaches can be investigated in future studies to address this problem. First, MeSH terms and publication type could be included as deep learning features. The caveat is that this approach would require feature engineering and would be limited by the time lag of MeSH terms and publication type described above. The second, and perhaps more promising approach, is to include the methods section from the article full-text as an input for deep learning. Since the methods section has many more details on the study methodology than the article abstract, it may lead to more accurate classification of scientifically sound studies.

False-positives were due to two main error categories. First, 7 of 20 cases were marginal articles that partially met quality criteria (eg, RCT without a clinical outcome) and therefore were more difficult to rate (7/20). Second, in 11 of 20 cases the abstract included terms related to high quality methodology but stated these outside the context of the study method (eg, abstract conclusion stating the need for future RCTs, editorial raising the need for RCTs on a specific topic). Approaches to mitigate both types of errors include using the full-text of the methods section as input for the deep learning model and developing separate subclassifiers to detect studies that meet partial quality criteria, and nonoriginal studies (eg, editorials, letters, reviews).

Limitations

Our study has four important limitations. First, although we focused on deep learning models and optimization strategies that were most likely to produce the best results, we have not exhausted all deep learning optimization possibilities. For instance, new work on RNNs may prove more accurate in document classification tasks [30,31]. We chose to focus our efforts on CNNs because they run more efficiently, given the large scale of our text data, but there is a valid investigation

into understanding the trade-offs between speed and accuracy by comparing these methods. We also did not exhaustively search the hyper-parameter space for our CNN. Many of our choices were empirical, as this is the first study, and further efforts might leverage more systematic approaches to hyper-parameter tuning [32]. Second, our approach is meant to be “end-to-end” (ie, text simply enters our pipeline and is classified). This approach is preferable because it does not require significant feature engineering or time-dependent features such as MeSH terms. However, further studies can explore adding richer features into our model to improve performance. For example, since the McMaster’s textword filter has equivalent recall as (but lower precision than) the Clinical Query filters, it is possible that MeSH-based features could improve the precision of our deep learning approach. Third, we have made comparisons with only one textword filter and no other machine learning approaches, since we did not have access to those machine learning classifiers. Comparisons with two of the three previous machine learning approaches are indirect, since those studies did not use Clinical Hedges as a gold standard. Last, we focused on identifying “treatment” studies; further work is needed to verify whether our approach generalizes to other areas, such as diagnosis, etiology, and prognosis.

Conclusion

We compared deep learning with state-of-the-art search filters to identify reports of scientifically sound studies in the biomedical literature. Overall, the resulting deep learning model compared well with other approaches, especially in scenarios involving recent citations prior to MeSH indexing. Advantages of the deep learning approach include low feature engineering requirements, no dependency on proprietary and time-sensitive features, and the use of a very large training set. Future work is needed to investigate further optimization opportunities and to adapt the deep learning approach to other clinical areas. Deep learning is a promising approach to identifying scientifically sound studies from the biomedical literature and warrants further investigation as a potential alternative for, or supplement to, current search filters.

Acknowledgments

GDF was funded by National Library of Medicine grant 1R01LM011416 and National Cancer Institute grant 1U24CA204800. Furthermore, this material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) Program Office under Contract No. W31P4Q-17-C-0103. The authors wish to acknowledge InferLink Corporation for their advice and support. The Clinical Hedges database was created with grants from the US National Library of Medicine and the Canadian Institutes of Health Research.

Conflicts of Interest

MM is the Chief Scientist of InferLink Corp. and CEO of Evid Science, Inc, both of which could benefit from using the above approach as a feature within existing or new medical literature analysis products. GDF, AI, CC, and RBH have no competing interests to declare.

References

1. Haynes RB. Where's the meat in clinical journals? ACP Journal club 1993;119(3):A22.
2. Ioannidis JPA. Why most published research findings are false. PLoS Med 2005 Aug;2(8):e124 [FREE Full text] [doi: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)] [Medline: [16060722](https://pubmed.ncbi.nlm.nih.gov/16060722/)]

3. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Inform Assoc* 1994;1(6):447-458 [[FREE Full text](#)] [Medline: [7850570](#)]
4. Wilczynski NL, McKibbon KA, Haynes RB. Enhancing retrieval of best evidence for health care from bibliographic databases: calibration of the hand search of the literature. *Stud Health Technol Inform* 2001;84(Pt 1):390-393. [Medline: [11604770](#)]
5. Wilczynski NL, Morgan D, Haynes RB, Hedges Team. An overview of the design and methods for retrieving high-quality studies for clinical care. *BMC Med Inform Decis Mak* 2005 Jun 21;5:20 [[FREE Full text](#)] [doi: [10.1186/1472-6947-5-20](#)] [Medline: [15969765](#)]
6. Wilczynski NL, McKibbon KA, Walter SD, Garg AX, Haynes RB. MEDLINE clinical queries are robust when searching in recent publishing years. *J Am Med Inform Assoc* 2013;20(2):363-368 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2012-001075](#)] [Medline: [23019242](#)]
7. Irwin A, Rackham D. Comparison of the time-to-indexing in PubMed between biomedical journals according to impact factor, discipline, and focus. *Res Social Adm Pharm* 2017;13(2):389-393 [[FREE Full text](#)] [doi: [10.1016/j.sapharm.2016.04.006](#)] [Medline: [27215603](#)]
8. Kilicoglu H, Demner-Fushman D, Rindfleisch TC, Wilczynski NL, Haynes RB. Towards automatic recognition of scientifically rigorous clinical research evidence. *J Am Med Inform Assoc* 2009;16(1):25-31 [[FREE Full text](#)] [doi: [10.1197/jamia.M2996](#)] [Medline: [18952929](#)]
9. Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis CF. Text categorization models for high-quality article retrieval in internal medicine. *J Am Med Inform Assoc* 2005;12(2):207-216 [[FREE Full text](#)] [doi: [10.1197/jamia.M1641](#)] [Medline: [15561789](#)]
10. Bernstam EV, Herskovic JR, Aphinyanaphongs Y, Aliferis CF, Sriram MG, Hersh WR. Using citation data to improve retrieval from MEDLINE. *J Am Med Inform Assoc* 2006 Jan;13(1):96-105 [[FREE Full text](#)] [doi: [10.1197/jamia.M1909](#)] [Medline: [16221938](#)]
11. Lee H, Grosse R, Ranganath R, Ng A. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. : ACM; 2009 Presented at: Proceedings of the 26th annual international conference on machine learning; 2009; Montreal, Quebec, Canada p. 609-616.
12. Lee H, Pham P, Largman Y, Ng AY. Unsupervised feature learning for audio classification using convolutional deep belief networks. 2009 Presented at: Advances in neural information processing systems; 2009; Vancouver, Canada p. 1096-1104.
13. Kim Y. arXiv. 2014. Convolutional neural networks for sentence classification URL: <https://arxiv.org/abs/1408.5882> [accessed 2018-05-30] [[WebCite Cache ID 6zoXJPr96](#)]
14. Mikolov T, Chen K, Corrado G, Dean J. arXiv. 2013. Efficient estimation of word representations in vector space URL: <https://arxiv.org/abs/1301.3781> [accessed 2018-05-30] [[WebCite Cache ID 6zoXQzoQh](#)]
15. Yin W, Kann K, Yu M, Schütze H. arXiv. 2017. Comparative study of CNN and RNN for natural language processing URL: <https://arxiv.org/abs/1702.01923> [accessed 2018-05-30] [[WebCite Cache ID 6zoXVRoAx](#)]
16. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700 [[FREE Full text](#)] [Medline: [19622552](#)]
17. Sampson M, Shojania KG, McGowan J, Daniel R, Rader T, Iansavichene AE, et al. Surveillance search techniques identified the need to update systematic reviews. *J Clin Epidemiol* 2008 Aug;61(8):755-762. [doi: [10.1016/j.jclinepi.2007.10.003](#)] [Medline: [18586179](#)]
18. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, Living Systematic Review Network. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol* 2017 Nov;91:23-30. [doi: [10.1016/j.jclinepi.2017.08.010](#)] [Medline: [28912002](#)]
19. Shojania K, Sampson M, Ansari M, Ji J, Garritty C, Rader T, et al. AHRQ Technical Reviews. Updating Systematic Reviews. Rockville, MD: Agency for Healthcare Research and Quality (US); 2007.
20. Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern Med* 2014 May;174(5):710-718. [doi: [10.1001/jamainternmed.2014.368](#)] [Medline: [24663331](#)]
21. Jansen B, Spink A, Saracevic T. Real life, real users, and real needs: a study and analysis of user queries on the web. *Inf Process Manag* 2000 Mar;36(2):207-227 [[FREE Full text](#)] [doi: [10.1016/S0306-4573\(99\)00056-4](#)]
22. Cohen AM, Ambert K, McDonagh M. Cross-topic learning for work prioritization in systematic review creation and update. *J Am Med Inform Assoc* 2009;16(5):690-704 [[FREE Full text](#)] [doi: [10.1197/jamia.M3162](#)] [Medline: [19567792](#)]
23. Krause J, Sapp B, Howard A, Zhou H, Toshev A, Duerig T, et al. The unreasonable effectiveness of noisy data for fine-grained recognition. 2016 Presented at: European Conference on Computer Vision; 2016; Amsterdam, The Netherlands p. 301-320.
24. Lee GE. A study of convolutional neural networks for clinical document classification in systematic reviews. In: SysReview at CLEF eHealth. 2017 Presented at: CEUR Workshop; 2017; Dublin, Ireland.
25. Hughes M, Li I, Kotoulas S, Suzumura T. Medical text classification using convolutional neural networks. *Stud Health Technol Inform* 2017;235:246-250. [Medline: [28423791](#)]

26. Wang Y, Liu S, Rastegar-Mojarad M, Wang L, Shen F, Liu F, et al. Dependency and AMR embeddings for drug-drug interaction extraction from biomedical literature. 2017 Presented at: 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics; 2017; Boston, MA, USA.
27. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. \$mathtt{DeepPr}\$: a convolutional net for medical records. IEEE J Biomed Health Inform 2017 Dec;21(1):22-30. [doi: [10.1109/JBHI.2016.2633963](https://doi.org/10.1109/JBHI.2016.2633963)] [Medline: [27913366](https://pubmed.ncbi.nlm.nih.gov/27913366/)]
28. Che Z, Cheng Y, Sun Z, Liu Y. arXiv. 2017. Exploiting convolutional neural network for risk prediction with medical feature embedding URL: <https://arxiv.org/abs/1701.07474> [accessed 2018-05-30] [WebCite Cache ID 6zoYNIfIE]
29. Marshall I, Noel-Storr A, Kuiper J, Thomas J, Wallace B. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Res Synth Methods 2018 Jan 04;1-13. [doi: [10.1002/jrsm.1287](https://doi.org/10.1002/jrsm.1287)] [Medline: [29314757](https://pubmed.ncbi.nlm.nih.gov/29314757/)]
30. Liu P, Qiu X, Huang X. Recurrent neural network for text classification with multi-task learning. : AAAI Press; 2016 Presented at: 25th International Joint Conference on Artificial Intelligence; 2016; New York, NY, USA p. AAA.
31. Lai S, Xu L, Liu K, Zhao J. Recurrent convolutional neural networks for text classification. : AAAI Press; 2015 Presented at: 29th AAAI Conference on Artificial Intelligence; 2015; Austin, Texas, USA p. 2267-2273.
32. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13:281-305.

Abbreviations

CNN: Convolutional Neural Network
DNN: Deep Neural Network
MeSH: Medical Subject Heading
RCT: randomized controlled trial
ReLU: Rectified Linear Unit
RNN: Recurrent Neural Network
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 05.03.18; peer-reviewed by M Banf, Y Wang; comments to author 12.04.18; revised version received 26.04.18; accepted 12.05.18; published 25.06.18.

Please cite as:

Del Fiol G, Michelson M, Iorio A, Cotoi C, Haynes RB

A Deep Learning Method to Automatically Identify Reports of Scientifically Rigorous Clinical Research from the Biomedical Literature: Comparative Analytic Study

J Med Internet Res 2018;20(6):e10281

URL: <http://www.jmir.org/2018/6/e10281/>

doi: [10.2196/10281](https://doi.org/10.2196/10281)

PMID: [29941415](https://pubmed.ncbi.nlm.nih.gov/29941415/)

©Guilherme Del Fiol, Matthew Michelson, Alfonso Iorio, Chris Cotoi, R Brian Haynes. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 25.06.2018. 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.

Original Paper

Content, Quality, and Assessment Tools of Physician-Rating Websites in 12 Countries: Quantitative Analysis

Fabia Rothenfluh¹, PhD; Peter J Schulz¹, PhD

Institute of Communication and Health, Università della Svizzera italiana, Lugano, Switzerland

Corresponding Author:

Fabia Rothenfluh, PhD

Institute of Communication and Health

Università della Svizzera italiana

Via G. Buffi 13

Lugano, 6900

Switzerland

Phone: 41 058 666 4485

Email: fabia.rothenfluh@usi.ch

Abstract

Background: Websites on which users can rate their physician are becoming increasingly popular, but little is known about the website quality, the information content, and the tools they offer users to assess physicians. This study assesses these aspects on physician-rating websites in German- and English-speaking countries.

Objective: The objective of this study was to collect information on websites with a physician rating or review tool in 12 countries in terms of metadata, website quality (transparency, privacy and freedom of speech of physicians and patients, check mechanisms for appropriateness and accuracy of reviews, and ease of page navigation), professional information about the physician, rating scales and tools, as well as traffic rank.

Methods: A systematic Web search based on a set of predefined keywords was conducted on Google, Bing, and Yahoo in August 2016. A final sample of 143 physician-rating websites was analyzed and coded for metadata, quality, information content, and the physician-rating tools.

Results: The majority of websites were registered in the United States (40/143) or Germany (25/143). The vast majority were commercially owned (120/143, 83.9%), and 69.9% (100/143) displayed some form of physician advertisement. Overall, information content (mean 9.95/25) as well as quality were low (mean 18.67/47). Websites registered in the United Kingdom obtained the highest quality scores (mean 26.50/47), followed by Australian websites (mean 21.50/47). In terms of rating tools, physician-rating websites were most frequently asking users to score overall performance, punctuality, or wait time in practice.

Conclusions: This study evidences that websites that provide physician rating should improve and communicate their quality standards, especially in terms of physician and user protection, as well as transparency. In addition, given that quality standards on physician-rating websites are low overall, the development of transparent guidelines is required. Furthermore, attention should be paid to the financial goals that the majority of physician-rating websites, especially the ones that are commercially owned, pursue.

(*J Med Internet Res* 2018;20(6):e212) doi:[10.2196/jmir.9105](https://doi.org/10.2196/jmir.9105)

KEYWORDS

physician rating websites; content analysis; website quality; patient Web portals; rating tools; health information; health care quality assessment; patient reviews

Introduction

Background

The internet has become an invaluable resource for any kind of question or query one may search an answer for. The search and selection of a physician via the internet is no exception,

especially if patient opinions can be easily obtained via the World Wide Web [1,2]. Physician-rating websites (PRWs) show numeric scores and textual appraisals about former patients' encounters and experiences with a physician. However, not only do specialized websites for physician assessments offer user reviews of doctors but general commercial webpages such as Yelp also ask users to review medical professionals [3]. Yet the

quality of PRWs and the rating tools they present to their users are largely unknown.

The content and quality of PRWs is a concern for both medical practitioners and website users. The former are afraid of unjustified reviews that do not reflect the true nature of their actual medical performance [4]. Biases in the user and the data, the risk of false allegations combined with website providers' negligence to systematically control PRW reviews, the anonymity of the ratings, as well as health care consumers' inability to judge certain aspects of care lead physicians to doubt the usefulness of PRW reviews [5]. Health care consumers on the other hand desire more quality of care information to improve their choices but have difficulties using such reports because of the complexity of the material [6,7]. To sum up, both physicians and health care consumers demand quality standards on PRWs that increase transparency while protecting both parties' freedom of speech and privacy. These insights call for an assessment of the availability and quality of PRWs to evaluate to what extent physicians' and health care consumers' worries are justified.

Study Objectives

A study assessing the quality, physician profile information, and rating tools present on PRWs across countries and languages has to our knowledge not yet been undertaken. This led us to the following research questions:

1. Of what website quality are PRWs? Which aspects of quality are most frequently met that are largely missing?
2. What information about the physicians and their practices is available on these websites?
3. How and based on which scales can users rate a doctor online?
4. How does quality and information content differ between countries?

Methods

Codebook

The website sample was collected through a Web search in August 2016 on the three largest search engines: Google, Bing, and Yahoo [8]. Web searches were conducted based on a list of keywords that were entered in English or German, based on the search country (Figure 1). Search engines were used with respective country codes (eg, in Germany we used search engine URLs ending in .de and in the United States URLs ending in .com) to mimic searches from residents looking for a doctor in their country. Websites were included in the sample if they fulfilled the following criteria: (1) accessibility in English or German, (2) retrievability via Web searches including one of the 12 preselected countries (United Kingdom, United States, Canada, Australia, New Zealand, India, Singapore, United Arab Emirates [Dubai], South Africa, Switzerland, Germany, and Austria), and (3) presence of ratings, evaluations, or written feedback sections to assess or rate physicians. We included PRWs from the United States and Germany in the study because the vast majority of publications up to date covered these two countries. To enlarge the sample, we included other countries where English or German is spoken. The first 100 webpages for each search term and engine were screened for the inclusion criteria, yielding a sample of N=208. The websites were coded from September 2016 to December 2016 and webcached. Examples of coded websites can be viewed in Figures 2 and 3 [9,10].

Metadata

The first part of our coding consisted of metadata, namely about the owner of the website, registration country, coverage area, or upgrade features. These indicators were coded for presence and absence and summarized in a table.

Figure 1. Search term strings entered in Google, Bing, and Yahoo to collect the website sample.

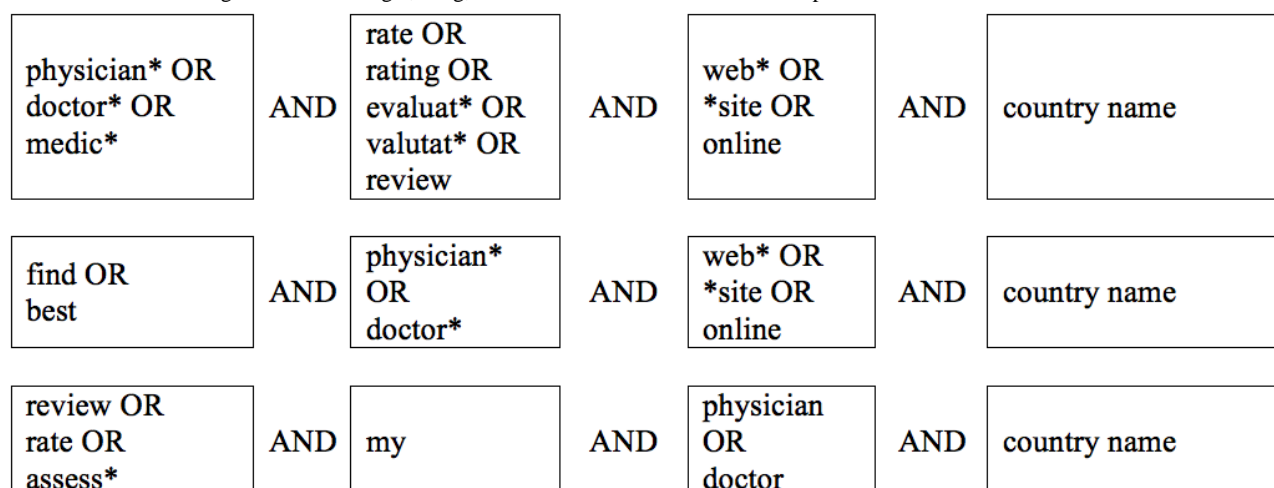
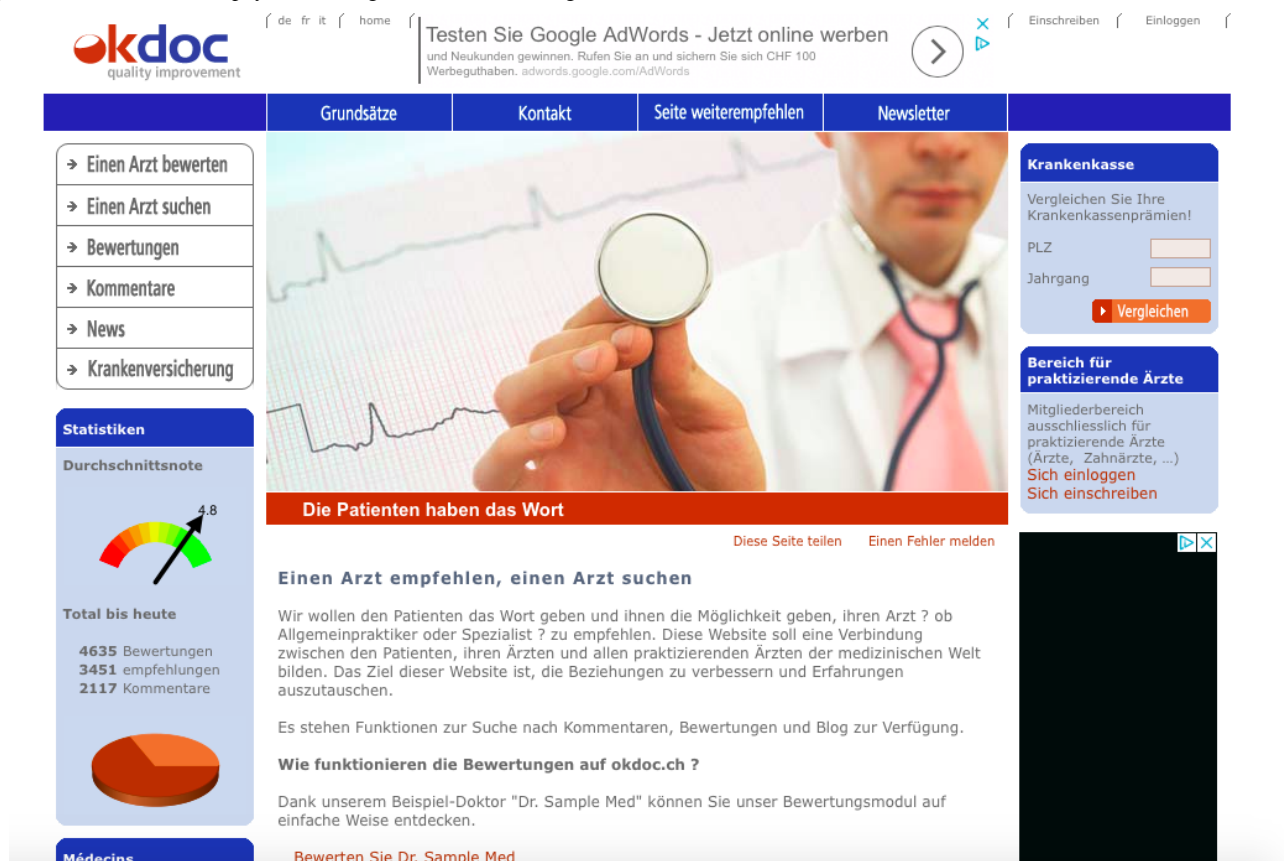
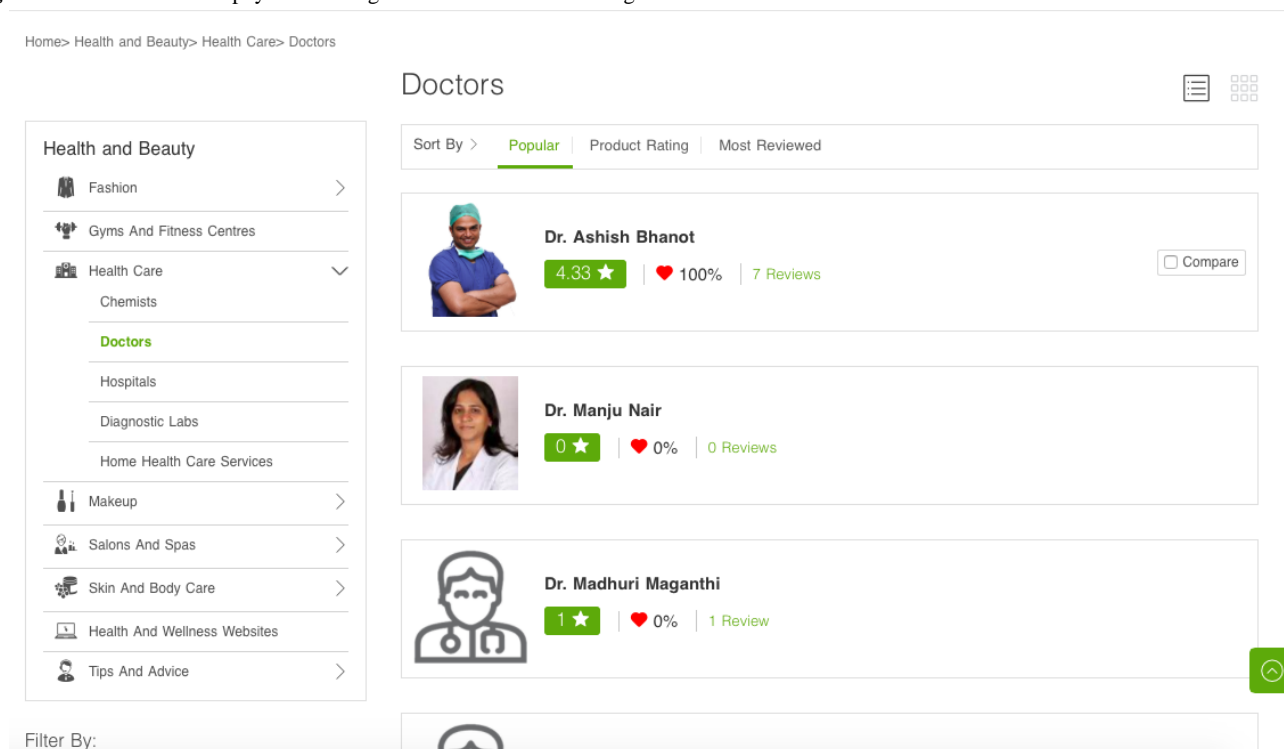


Figure 2. Screenshot of the physician-rating website okdoc.ch registered in Switzerland.**Figure 3.** Screenshot of the physician-rating website mouthshut.com registered in India.

Website Quality

The second section of the coding focused on website quality. On the basis of literature [4,5,11-13], five quality dimensions for PRWs quality were developed and indicators accordingly

derived. According to the definition applied in this study, a high quality PRW (1) publishes transparent, accurate, and neutral content from evident sources (eg, clear separation of advertisement and content); (2) respects the privacy and freedom of speech of both physician (eg, informs physicians about new

ratings and asks physicians to reply to reviews) and health care consumers (eg, publishes anonymous reviews and verifies the identity of reviewers); (3) has check mechanisms in place to ensure the accuracy and appropriateness of information content reviews (eg, number of reviews that a health care consumer can write is limited and all reviews are checked before publication); and (4) is easy to use and navigate (eg, has filters present and search masks available). The breakdown of these dimensions led to 47 indicators, which can be found in [Multimedia Appendix 1](#). The indicators were not weighted because the opposing views about the weight of individual indicators from physicians' and patients' perspectives can hardly be reconciled (eg, although physicians oppose anonymous reviews, patients highly value them as they fret about the impact of a negative review on future care).

Professional Information

The third section on information content consisted of 25 items covering information on the professional and educational background of the physician, practice access, and contact information, as well as personal data about the doctor. These indicators were developed iteratively; first, indicators were derived from a scoping review by Victoor and colleagues [14] and a study by Rothenfluh and Schulz [15] on aspects that were cited to be for patients when choosing a physician. In a second step, during the pretest, more information items were added until saturation was reached. All items that appeared during the actual coding and did not fit into the indicators collected during the development and pretest were ordered into separate categories labeled *other*.

Scales or Rating Tools

A fourth section of the codebook was dedicated to the scales or rating tools available on PRWs to assess a physician. An earlier study by Rothenfluh and Schulz [15] identified indicators that are, according to physicians and health care consumers, important to identify a good physician and assessable by patients. These indicators were developed based on Donabedian's quality of care model [16,17] that subdivides care quality into dimensions of structure (infrastructure, staff, equipment, organization, and accessibility), process (technical and interpersonal skill of the physician), and outcome (results of the treatment) of health care. All rating tools on the websites in this study's sample were coded based on this structure.

Website Traffic Rank

Furthermore, the website traffic rank, an indicator for website popularity, was recorded for each webpage on January 26, 2017, based on the Alexa Global Traffic Ranking [18]. For each website, both the traffic rank globally as well as the national ranking were recorded.

Coding Procedure

The codebook was pretested twice based on 10 websites each. Adjustments were made where necessary. To assure the

reliability of the data, a second coder was trained based on the codebook (three sessions of 3 hours). The level of agreement between the first and second coder was compared after each session and differences were discussed. This process was repeated three times until sufficient agreement was reached. Then, the second coder independently coded a randomly drawn subsample of 29 websites (20%), which is sufficient according to Riffe and colleagues [19].

Results

Coding Procedure and Inter-coder Reliability

On the basis of our Web search, the initial sample of websites consisted of 208 websites, which was consolidated to 143 during the coding process because of temporary inaccessibility of websites, deletion or cessation, or the disappearance of an online review function. The inter-coder reliability based on Krippendorff alpha [20] was satisfactory (average over all items $\alpha=.95$) after five items were excluded because of agreement below $\alpha=.667$, which is the lowest cutoff according to Krippendorff [20,21]. Among the excluded items were the completeness of the presented physician profiles and the format and source in which the profile information was presented (eg, open-ended text sections filled in by the physician, information provided by the provider, or not identifiable).

Metadata

In terms of metadata ([Table 1](#)), we found that the vast majority of websites were operated by commercial for-profit companies (120/143, 83.9%). Business models included various profile upgrade options for physicians that are often payable through monthly or annual fees (see detailed features in [Table 1](#)). Such benefits included that physicians could pay for commercial services (42/143, 29.4%) such as online appointment booking, or to enter the biddings to offer a client treatments (especially in dentistry). Furthermore, doctors could pay for their profiles to be further up or listed first on users' search results (51/143, 35.7%).

Website Quality

A quality index of 47 items was calculated (see [Multimedia Appendix 1](#)), awarding one point per quality criterion fulfilled, yielding scores between 0 and 47. The mean quality score was 18.67 (SD 4.13), ranging from 8 to 29 points, with 69.3% (99/143) of the sample scoring between 15 and 22 points. Overall, the three highest individual website quality scores of the 143 coded websites were attained by one website registered in Germany, one in Austria, and one in the United Kingdom. The individual websites with the lowest website quality were a website registered in Singapore and one in Canada reaching only 8 points or 9 points, respectively.

Table 1. Website metadata and features offered (N=143).

Metadata	n (%)
Owner of the website	
Unknown or not identifiable	11 (7.7)
Commercial for-profit company	120 (83.9)
Nonprofit organization	5 (3.5)
Medical professional organization	1 (0.7)
Political institution or governmental organization	2 (1.4)
Other (eg, hospital and health insurance)	4 (2.8)
Physician information updates and upgrades offered	
No upgrades offered	11 (7.7)
Profile update offered, no indication if at a cost	27 (18.9)
Profile update offered for free	4 (2.8)
Profile update offered at a cost	12 (8.4)
Profile update offered for free, upgrades available at a cost	84 (58.7)
No information available	5 (3.5)
Types of cost billing	
No information available	33 (23.1)
Absent	56 (39.2)
Fee (weekly, monthly, or annual)	41 (28.7)
Billing by case (per client served or gained)	5 (3.5)
Fee (monthly, annual) plus billing by case	7 (4.9)
Cost per information item the doctor adds	1 (0.7)
Upgrade benefits offered for free and at a cost	
Better or higher listing position of physician profile	
For free	7 (4.9)
At a cost	51 (35.7)
Seal that the doctor is excellent	
For free	3 (2.1)
At a cost	11 (7.7)
Google indexing for higher position in search results	
For free	3 (2.1)
At a cost	19 (13.3)
Customer service or profile maintenance	
For free	5 (3.5)
At a cost	22 (15.4)
Profile presentation enhancement by adding pictures, videos, or more information about the doctor	
For free	14 (9.8)
At a cost	61 (42.7)
Physicians can respond to patients' reviews	
For free	22 (15.4)
At a cost	10 (7.0)
Commercial benefits (eg, online appointment scheduling and bidding system for treatments)	
For free	11 (7.7)

Metadata	n (%)
At a cost	42 (29.4)

Table 2. Information available on doctors' profile pages about the physician and practice (N=143).

Available information on physicians' profiles	Present, n (%)
Address of the practice	136 (94.4)
Phone number of the practice	118 (81.9)
Directions to find address	111 (77.1)
Detailed degree or specializations of the doctor	93 (64.6)
List of medical conditions that the physician treats	85 (59.4)
List of medical procedures (treatments, etc) offered	86 (59.7)
Website of the physician and practice	78 (54.2)
Office hours of the practice	75 (52.1)
Doctor's years of work experience	60 (41.7)
Languages the physician speaks	58 (40.3)
Insurance plan restrictions information (eg, if a physician accepts only private insurance or self-payment)	58 (40.3)
Training and degrees of the doctor	70 (48.6)
Email address of the practice	67 (46.5)
List of medical continuous education courses the physician completed	56 (38.9)
Gender of the physician	49 (34.3)
Awards and honors that the physician received	34 (23.6)
Scientific publications by the physician	28 (19.4)
Insurance plans or health insurance companies the provider works with	26 (18.1)
Practice access information for handicapped people	21 (14.6)
Doctor's memberships in physician associations	15 (10.4)
Age of the physician	12 (8.3)
Physician's external quality assessment results	10 (6.9)
Number of surgical procedures that the physician performed in his career (surgeon volume)	4 (2.8)
Legal actions after errors, malpractice, or sanctions that were filed against the physician	2 (1.4)
Personal information (eg, doctor's marital status and family information)	2 (1.4)

Overall, indicators related to transparency, such as the type of website provider (132/143, 92.3%) and website background information (eg, website owner [122/143, 85.3%] or terms and conditions [128/143, 89.5%]) were available on the vast majority of websites. Furthermore, basic quality assurance criteria such as limiting the number of reviews by the same user were present in almost three-quarters of websites. However, the assurance of transparency proved to be less common when related to financial benefits for the website provider. For example, 75.5% (108/143) of websites did not clearly separate advertisement from content, and more than 69.9% (100/143) had some form of physician ad present. Furthermore, HON code certifications, a label that marks trustworthy health and medical information [22], were only displayed on 5.6% (8/143) of the cases. Statistical quality assurance indicators such as a minimum number of reviews online before reviews are displayed were largely absent (only in 9.3%, 13/143 present). Remarkably, only 4.2% (6/143) of the websites stated that they notify the

physicians whose profile goes online. This makes keeping track of their potentially numerous online profiles difficult for physicians. Furthermore, merely 11.9% (17/143) provided the physician's medical board registration number, which can be an indicator for users that may help him identify physicians who passed the country's requirements to practice medicine.

Professional Information

When Web users search for a doctor online, the amount and kind of information they find about the doctor and his or her practice may help users' decide whether to consult a certain doctor or not. Therefore, information content was added up to a score between 0 and 25 (see indicators in Table 2). With a mean of 9.95 (SD 4.01), the vast majority of websites had little information available about the doctors listed. The information most commonly available about doctors were the address of the practice (136/143, 94.4%), the practice phone number (118/143, 81.9%), and directions to find the practice (111/143, 77.1%),

which resembles the information one could also retrieve from a phone book.

Quality and Information Content on Physician-Rating Websites in Different Countries

To shed light on potential differences between countries' PRW information content and quality, the websites were split based on their registration country. According to our quality measure (0-47 points), websites registered in the United Kingdom had the best quality PRWs (mean 26.50, SD 1.00), followed by Australia (mean 21.50, SD 0.71) and Germany (mean 20.72, SD 4.12). It has to be noted though that both the United Kingdom and Australia had few websites registered in their countries and low variability compared with, for example, Germany, where the two highest quality websites were registered (see [Table 3](#)). In terms of information content (score between 0-25), websites registered in Australia scored the highest (mean 15.50, SD 6.64), followed by German ones (mean 10.00, SD 2.48).

Numerous websites (35/143) had multinational coverage areas, which may or may not overlap with the registration country according to which the countries are listed in [Table 3](#). For example, Canada had only 3 websites registered, but 10 pages covered physicians practicing in Canada. The United States on the other hand had 40 websites registered there, but only 28 exclusively displayed physician profiles from doctors practicing in the United States. Therefore, a comparison of websites across countries should be interpreted with caution. It should further be noted that the quality scores of websites not registered in one of our sample countries had the lowest quality, which points to potential legal issues that may emerge based on this gap between registration country and coverage area.

Scales or Rating Tools

PRWs may invite users to score doctors based on numeric scales, ask for written feedback, or a combination of both. In our sample, 15.3% (22/143) of the websites asked for numeric physician assessments only, 4.2% (6/143) for written reviews exclusively, and 76.9% (110/143) provided the option to give feedback based on both predefined rating scales as well as to type reviews or testimonials. Most frequently, PRW users were invited to rate the overall treatment encounter (75/143, 52.1%), punctuality and wait time in practice (51/143, 35.7%) or for the next appointment (27/143, 18.9%), and whether the user would recommend that specific doctor (44/143, 30.8%). Furthermore, users may be asked to rate the office environment (eg, practice comfort and cleanliness: 33/143, 23.1%), or the friendliness and courteousness of the staff (32/143, 22.4%). Looking at the assessment tools in terms of broader dimensions, one or several indicators on interpersonal aspects of care could be assessed on 65 (45.5%) of the coded websites. Specifically, information provision by the physician (comprehensiveness, clarity, questions answered, etc), bedside manner, helpfulness, and empathy (25/143, 17.4%), or if the doctor spent sufficient time with the patient (17/143, 11.8%) could be rated. Yet, scales on which users were asked to score one or several technical aspects of care were less frequently present (37/143, 25.9%). For example, few websites asked users to evaluate the physician's knowledge (16/143, 11.2%), competence (9/143, 6.3%), or the correctness of the diagnosis (9/143, 6.3%). Users were inquired to rate aspects concerning one or several rating items on the outcome of care on 24 (16.8%) of the websites. Assessment items included, for example, the presence and quality of the follow-up (9/143, 6.3%) or the efficiency of the treatment (8/143, 5.6%). Further rating scale items can be found in [Table 4](#).

Table 3. Website quality and information by registration country in descending order of quality.

Registration country ^a	Number of websites	Website quality (summative score 0-47)			Information content (summative score 0-25)		
		Mean (SD)	Minimum	Maximum	Mean (SD)	Minimum	Maximum
United Kingdom	4	26.50 (1.00)	26	28	8.00 (6.64)	2	16
Australia	2	21.50 (0.71)	21	22	15.50 (2.12)	14	17
Germany	25	20.72 (4.12)	15	29	10.00 (2.48)	5	16
Austria	13	19.85 (3.60)	14	28	12.85 (4.00)	8	20
Switzerland	11	18.91 (4.04)	13	25	9.27 (2.76)	3	13
United States	40	18.25 (3.03)	12	24	9.80 (4.24)	0	20
Singapore	5	18.20 (5.85)	8	22	10.40 (5.23)	6	17
South Africa	6	18.17 (3.43)	14	23	9.33 (3.93)	5	14
India	14	17.79 (1.81)	15	23	9.36 (4.52)	1	15
United Arab Emirates (Dubai)	3	17.00 (2.00)	15	19	13.33 (4.73)	8	17
Canada	3	15.67 (7.23)	11	24	8.33 (2.89)	5	10
Other or not identifiable registration country (eg, Spain and Romania)	17	15.29 (4.66)	9	22	8.53 (3.81)	3	16

^aThe registration country according to which websites are listed here is not always equivalent with the coverage areas of these websites.

Table 4. Rating scale items present on physician rating websites.

Dimension and indicators	Present, n (%)
Structure	
Infrastructure	
Office environment, cleanliness, comfort	33 (23.1)
Instruments in the practice to make the diagnosis or execute the treatment	9 (6.3)
Reachability of the practice by car or public transport	8 (5.6)
Organization	
Punctuality, wait time in practice	51 (35.7)
Scheduling or making appointments	23 (16.1)
Waiting time until the next appointment	27 (18.9)
Reachability of the practice via phone	9 (6.3)
Notification of patients in case of appointment delays or cancellations	3 (2.1)
Teamwork between physician and his team	2 (1.4)
Number of staff present in the practice to welcome and take care of patients	1 (0.7)
Staff	
Staff friendliness and courteousness	32 (22.4)
Staff experience and training	5 (3.5)
Process	
Interpersonal	
Comprehensiveness and completeness of information provision	31 (21.7)
Social skills of the doctor (attentiveness, helpfulness, empathy)	25 (17.5)
Amount of time spent with the patient	17 (11.9)
Friendliness of the physician	16 (11.2)
Physician's (active) listening skills	15 (10.5)
Conversation climate with the doctor	15 (10.5)
Trust in physician	13 (9.1)
Confidentiality, protection of privacy	6 (4.2)
Information provision about how to handle the illness or disease	10 (7.0)
Shared decision about the course of action together with the patient or shared decision making	6 (4.2)
Doctor's effort to engage the patient in shared decision making	6 (4.2)
Physician's skill to assess the patient's handicaps and presentation with appropriate information and treatment options	1 (0.7)
Communication and narration during the treatment execution	1 (0.7)
Technical or medical	
Physician's knowledge	16 (11.2)
Physician's competence	9 (6.3)
Correctness of the diagnosis, diagnostic ability of the physician	9 (6.3)
Improvement of the patient's health status	8 (5.6)
Timely referral to a specialist or the hospital if needed	5 (3.5)
Completeness and quality of anamnesis	4 (2.8)
Quality and variety of treatment suggestions	3 (2.1)
Cost consciousness of the physician when making tests or giving out medications	3 (2.1)
Physician's experience	2 (1.4)

Dimension and indicators	Present, n (%)
Responsible medication prescription	2 (1.4)
Systematic proceeding of physician to reach the correct diagnosis	2 (1.4)
Timeliness or promptness of the diagnosis and initiation of the treatment	2 (1.4)
Correctness of treatment execution by the physician and his team	1 (0.7)
Quality of the information provided to the patient	19 (0.7)
Physician's competence to execute the treatment competently	1 (0.7)
Outcome	
Likelihood of recommendation	44 (30.8)
Satisfaction with the doctor	12 (8.4)
Presence and quality of the follow-up care	9 (6.3)
Efficiency of the treatment or cost-benefit ratio	8 (5.6)
Price of the treatment	4 (2.8)
Cost coverage by the health insurance	2 (1.4)
Patient's increase in knowledge about his disease or injury	1 (0.7)
Number or kind of complications ^a	0 (0.0)
Patient loyalty or patient's intention to return for future or follow-up treatments ^a	0 (0.0)
Summative and other	
Summative or overall score	75 (52.4)
Other organization scores	14 (9.8)
Other interpersonal scores	12 (8.4)
Other overall scores	2 (1.4)
Other technical scores	2 (1.4)

^aThese indicators emerged in the literature as important to identify a good doctor but were not present on any physician-rating websites.

Website Traffic Rank

The website traffic rank on Alexa was recorded on January 26, 2017, serving as an indicator of the popularity of the PRWs in this study in their registration country. First, it should be noted that 8 websites did not have an Alexa global rank [18], while 44 were not ranked locally. The most frequently visited website was Yelp on a global scale, ranked on position 282, followed by Web MD ranked number 501 and Yellow Pages on position 1634 worldwide. Given that Yelp and Yellow Pages are primarily directories, they attract most likely the vast majority of their traffic through webpage visits unrelated to physician searches or review writing. The website most popular within a country was Herold in Austria on rank 86 nationally, followed by Just Dial in India positioned on rank 63, and by NHS Choices in Great Britain on national rank 143. Herold as well as Just Dial are also first and foremost directories, likely attracting most of their traffic through address searches, whereas users may not even be aware of its function to rate doctors. NHS Choices on the other hand is Britain's public health care system's webpage and therefore the first point of entry or first address about health issues in the United Kingdom.

Discussion

Principal Findings

This study assessed the quality, information content, and rating tools on websites providing physician rating in 12 countries of German or English language. Most websites were registered in the United States and Germany. Yet, one has to differentiate between registration country and coverage area of those websites as this has important legal implications for the physicians listed. On average, quality and information content of PRWs in various countries differed tremendously, whereas the quality of the majority did not even achieve half of the maximum quality points possible.

To our knowledge, this is the first study that analyzed the quality, physician information content, and rating tools of doctor rating websites in 12 countries on a broad basis. Prior content analyses that focused on the structure and content on PRWs were more restrictive in their inclusion criteria only coding websites that exclusively displayed physicians, leading to smaller sample sizes of 8 [23] or 28 websites, respectively [24]. Due to our broader inclusion criteria (see Methods section), we analyzed a more heterogeneous sample, including all websites that had some form of physician rating or review present, consequently providing a broader picture.

The majority of the websites in our sample was commercially owned (120/143, 83.9%). On such profit-focused websites, revenue is often generated via upgrades of physician profiles. This can be beneficial for both patients and physicians if features such as online appointment bookings are offered. However, other upgrades such as purchased top listing positions, which were present on 35.7% (51/143) of the websites, are problematic as they are often not evident as such to the user. As known from research in marketing, because of primacy or position effects, people tend to choose the first option on a list (eg, [25-27]). However, a purchased top physician listing may not reflect the actual quality of the doctor, thereby potentially misleading users.

Overall, the quality of the websites in our sample was mediocre. The vast majority attained less than half of the maximum score (between 15-22 points), whereas the highest quality website attained 29 out of 47 points. Additional information such as surgeon volume [28] or physician notification when the profile goes online were absent in almost 95% (6/143) of all websites. Furthermore, only a quarter of websites contained physicians' replies to reviews, even though health care consumers report physician feedback to be crucial [13].

New and transparent quality guidelines are called for. Such quality guidelines should strengthen the rights of both physicians and health care consumers. In Germany, such guidelines have been developed. However, as the results of this study show, this did not necessarily translate into higher quality PRWs overall [11]. Independent, nonprofit companies such as the HON code society [22] for the quality of health information have paved the way toward globally recognized labels. For the development of a PRW quality label, a mixed committee of patients and physicians should be involved, so that both patients' and doctors' wishes and concerns find their perspectives and needs represented in the development of such a label. Given the diversity of medical specializations (ie, general practitioners may require a different skill set than neurosurgeons) and the knowledge base of patient (ie, a very informed long-term diabetes patient may have a different skill to assess a doctor than a patient who visits a doctor for the first time in 20 years to find out that he is suffering from diabetes), PRWs should pay tribute to such differences.

In terms of what users are asked to rate about their doctors, we found that overall scores or aspects of the organization, such as waiting time, appeared most frequently. Outcome of care and technical aspects of the physician are less often listed rating tools. Furthermore, rating websites that asked users to assess structural or interpersonal aspects of their care ranked higher on website quality than websites on which those rating scales were absent. This is in line with previous studies that report that overall scores, communication, and structural factors are the most frequently available doctor assessment tools [23,29]. Emmert and colleagues [23] further found that only a minority of PRWs asked users to assess process quality or treatment outcomes [23], which our study confirms.

A study by Rao and colleagues [30] reported that health care consumers failed to correctly judge technical quality of care aspects, whereas other studies suggest a positive association between PRW ratings and objective care measures [31], or

Facebook recommendations and hospital readmissions [32]. Some studies suggest that there is a knowledge gap between doctors and patients related to medical knowledge [5], which makes it difficult for health care consumers to accurately evaluate a physician's medical performance. A study by Rothenfluh and Schulz [15] reports further that physicians and health care consumers are indeed reflective of their own capabilities to assess certain care aspects, especially if highly technical. The findings from this study evidence, however, that only a minority of websites present technical or medical criteria (37/143, 25.9%) to be assessed by patients, which may be a reassuring finding for doctors. Hence, these findings can debilitate some ethical concerns raised in the PRW literature (see [4,5]).

Beyond numeric reviews, written patient testimonials should also be focused on more in this context. In our sample, patient narratives have been present on 81.1% (116/143) of the PRWs we analyzed. These narratives have been promoted as fruitful tools to obtain patient quality of care feedback, yet, it is advocated that they should be collected based on strict standards, showing promising results [33]. Our study revealed that only 26.6% (38/143) of the pages we analyzed provided instructions on how to provide meaningful and appropriate written testimonials. Furthermore, not even a third (30.8%, 44/143) of the websites in our sample systematically screened all patient reviews before they went online. This calls for action and enforcement of stricter quality guidelines on PRWs. Furthermore, given that numeric physician ratings are often not in line with written reviews, they can also cause contradictions within reviews [34]. Hence, further research is needed on how ratings, as well as narratives, could be more effectively elicited to provide meaningful and valuable feedback for physicians and insightful information for patients.

The quality of the websites differed between the 12 countries, with websites registered in the United Kingdom and Australia scoring highest. Yet, the registration country was not necessarily equivalent to the website's coverage area. This could create difficulties, especially for doctors who want to take legal action against false or defamatory reviews. Even though there are the first publications on the legal grounds of PRWs, such as applying defamation law and medical nondisclosure agreements [12], the situation remains country-specific and complex. Newspaper articles on court cases in various countries, including Europe and North America, show that actions taken by physicians against defamatory comments are sometimes, but not always, successful [35-37]. A court decision in Germany shows though that the law increasingly recognizes the physicians' perspectives on PRWs, especially when business interests compromise the neutrality of the displayed information. A ruling by the German High Court forces a PRW to delete doctors' profiles when the doctors explicitly request for it. A PRW was sued because it displayed advertisements of upgraded doctor profiles on detail pages of physicians without such an upgrade, putting the doctors with a nonupgraded profile into a disadvantageous position [38]; Doctors could not request the removal of their profiles. The new ruling by the German High Court changes this situation. As a consequence, this ruling may change the landscape of physician profile upgrades in other

countries as well, favoring transparency, while punishing practices that may distort users' perception, such as profile upgrades.

Beyond a discussion and further research on legal issues on PRWs, country-specific differences in terms of the number, content, and quality of such websites may also be related to the health systems in which they operate. A health care system such as that of the United Kingdom is publicly financed and therefore leaves less choice to patients [39]. In countries such as the United States, on the contrary, individuals pay their health care expenses mostly out of their pocket [40]. Hence, PRWs in such systemically diverse countries may also lead to the emergence of dissimilar PRWs. Self-payers may therefore be more interested in comparisons of physicians, potentially explaining the large number of PRWs in the United States compared with, for example, the United Kingdom. However, these are just hypotheses, calling for further investigation.

The large number of websites that we found, for example, in Germany or the United States, point to a challenge doctors face; they have to stay on top of incoming reviews and to respond to them. Given the large number of review websites, this poses a daunting task, especially if websites do not inform doctors when their profiles or a review on them goes online. Only 14.0% (20/143) of the websites in our sample stated to inform doctors when a new review on them is posted, and 4.2% (6/143) notify physicians when their profile goes live. Given that physicians already have extraordinarily long working hours [41], replying to patient comments on such numerous websites in a timely manner can become burdensome, or even overwhelming [42]. It is therefore not surprising that only 34% of physicians surveyed in the United States reported to have made changes to their online profiles [43], even though the importance of responses to posted reviews is highly important for patients [13]. This situation has already opened up a new business opportunity: marketing firms offer physician profile maintenance at a cost [44-46]. This development is alarming because contrary to creating a physician-patient dialogue to improve quality of care [47] it outsources this potentially valuable feedback loop.

This study pointed to various shortcomings on websites that offer physician rating tools, reaching from low quality and limited information content to biased physician profile display. Nevertheless, online patient ratings of care quality can provide valuable and timely insights into shortcomings in care quality. Several studies have hinted at the association between objective care quality and patient feedback [32,48,49]. For example, a study by Glover and colleagues showed a significant association between Facebook ratings and hospital readmission rates, whereby lower ratings were associated with higher readmission rates [32]. Furthermore, patient ratings have the potential to change patients' choice of doctors or hospitals, thereby encouraging doctors to adjust their practicing based on negative reviews [48]. This has been evidenced in a German study that showed how doctors who read their negative reviews and also responded to them, make an effort to improve the aspects that were criticized in their work [47]. Hence, the need for the public to be involved in quality of care reporting is increasingly suggested [49,50]. Yet, the assurance of quality and content on

PRWs asks for further research and knowledge translation into practice.

Given that the vast majority of PRWs in our study was commercially owned, it remains questionable whether more neutral providers such as nonprofit organizations could outperform the existing websites in terms of traffic at the present state. As almost a third of websites sell profile upgrades or higher listing positions to physicians, a sign or label should clearly point out the absence of such business models. This can raise users' awareness and incentivize commercial websites to change their business strategy. Furthermore, a label on PRWs could serve as quality assurance certification that helps both physicians and patients to better navigate toward ethical and high-quality physician-rating webpages. In addition, the number of reviews would have to be increased; the more reviews are present, the higher the statistical representativeness and the less subject single reviews are to outliers. As an outcome, reviews would be created in an environment safe for both parties, inspiring health care improvements and constructive dialogue, thereby ultimately raising transparency and quality in health care.

Limitations

This study has to be considered under certain limitations. First, as the internet is a fast-changing environment, this content analysis of PRWs only provides a momentary picture of the situation in the 12 countries included in our sample. To respond to this issue, we provide readers with insight into the websites as they were when they were coded in the form of webcaching. Yet, webcached sites may not necessarily provide the same user experience.

Second, the Web search for PRWs only included pages that appeared based on the outlined search terms. The search terms we applied may not be complete though because of regional differences in search strings. Hence, some websites offering rating functions of physicians may not have been included in the sample.

Third, this content analysis was limited to information that was visible when users accessed the website (publicly available without user registration). However, it is possible that websites adhered to quality criteria in the version available to registered users, which we as unregistered visitors failed to notice. As a result, this study only attempts to draw conclusions based on publicly available data.

Fourth, the indicators comprising the website quality score were not weighted. Although it is undebated that some indicators are very important (eg, physicians can be high listing positions), other indicators are difficult to weight. For example, whether reviews should be published anonymously is debated among doctors and patients. Doctors value transparent publication of reviewers' names and demand that when they are reviewed and rated, the authors of such reviews should reveal their identities as well. For patients on the other hand, it is crucial that their opinions remain anonymous as they fret about the impact of negative reviews on their future care encounters with the same or other providers. Hence, we opted to not weight the single indicators but to calculate a simple mean and to point the reader

to the table where all indicators are listed separately. This way, the reader can draw his or her own conclusions.

Conclusions

This study evidences that websites that provide physician rating should improve and communicate their quality standards,

especially in terms of physician and user protection, as well as transparency. In addition, given that quality standards on PRWs are low overall, the development of transparent guidelines is required.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Quality criteria fulfillment levels.

[[PDF File \(Adobe PDF File\), 72KB - jmir_v20i6e212_app1.pdf](#)]

References

1. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. *J Am Med Assoc* 2014 Feb 19;311(7):734-735. [doi: [10.1001/jama.2013.283194](#)] [Medline: [24549555](#)]
2. 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](#)]
3. Merchant RM, Volpp KG, Asch DA. Learning by listening-improving health care in the era of yelp. *J Am Med Assoc* 2016 Dec 20;316(23):2483-2484. [doi: [10.1001/jama.2016.16754](#)] [Medline: [27997663](#)]
4. 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](#)]
5. Patel S, Cain R, Neailey K, Hooberman L. General practitioners' concerns about online patient feedback: findings from a descriptive exploratory qualitative study in England. *J Med Internet Res* 2015 Dec 8;17(12):e276 [FREE Full text] [doi: [10.2196/jmir.4989](#)] [Medline: [26681299](#)]
6. Zwijnenberg NC, Hendriks M, Bloemendal E, Damman OC, de Jong JD, Delnoij DM, et al. Patients' need for tailored comparative health care information: a qualitative study on choosing a hospital. *J Med Internet Res* 2016 Nov 28;18(11):e297 [FREE Full text] [doi: [10.2196/jmir.4436](#)] [Medline: [27895006](#)]
7. 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: [10.1177/10775587070640020301](#)] [Medline: [17406019](#)]
8. eBIZ MBA. 2017. Top 15 Most Popular Search Engines URL: <http://www.ebizmba.com/articles/search-engines> [WebCite Cache ID 6orPOA7LM250A]
9. Okdoc Quality Improvement. URL: <http://www.okdoc.ch/> [WebCite Cache ID 6kdUr6rDy]
10. Mouthshut. Doctors URL: <http://www.mouthshut.com/product/categories.php?cid=925756> [accessed 2016-11-30] [WebCite Cache ID 6mPGx027W]
11. Ärztliches Zentrum für Qualität in der Medizin. 2011. Gute Praxis Bewertungsportale: Qualitätsanforderungen für Arztbewertungsportale [in German] URL: <https://www.aezq.de/aezq/arztbewertungsportale> [WebCite Cache ID 6z9557HXL]
12. Lee SD. "I hate my doctor": reputation, defamation, and physician-review websites. *Health Matrix Clevel* 2013;23(2):573-606. [Medline: [24341082](#)]
13. Patel S, Cain R, Neailey K, Hooberman L. Exploring patients' views toward giving web-based feedback and ratings to general practitioners in England: a qualitative descriptive study. *J Med Internet Res* 2016 Aug 05;18(8):e217 [FREE Full text] [doi: [10.2196/jmir.5865](#)] [Medline: [27496366](#)]
14. Victoor A, Delnoij DM, Friele RD, Rademakers JJ. Determinants of patient choice of healthcare providers: a scoping review. *BMC Health Serv Res* 2012 Aug 22;12:272 [FREE Full text] [doi: [10.1186/1472-6963-12-272](#)] [Medline: [22913549](#)]
15. 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 May 1;19(5):e127 [FREE Full text] [doi: [10.2196/jmir.6875](#)] [Medline: [28461285](#)]
16. Donabedian A. The quality of care. How can it be assessed? *J Am Med Assoc* 1988;260(12):1743-1748. [Medline: [3045356](#)]
17. 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](#)]
18. Alexa. URL: <http://www.alexa.com> [accessed 2018-03-30] [WebCite Cache ID 6z8CIK4Ly]
19. Riffe D, Lacy S, Fico F. *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Second Edition. Mahwah, NJ; London: Lawrence Erlbaum Associates; 2005.

20. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Second Edition. California: SAGE Publications; 2004.
21. Krippendorff K. Reliability in content analysis. *Human Comm Res* 2004 Jul;30(3):411-433. [doi: [10.1111/j.1468-2958.2004.tb00738.x](https://doi.org/10.1111/j.1468-2958.2004.tb00738.x)]
22. Health on the Net Foundation. The HON Code of Conduct for medical and health Web sites (HONcode) URL: <http://www.hon.ch/HONcode/Conduct.html> [WebCite Cache ID 6z965mqFS]
23. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. *Methods Inf Med* 2012 Nov;51(2):112-120. [doi: [10.3414/ME11-01-0045](https://doi.org/10.3414/ME11-01-0045)] [Medline: [22101427](https://pubmed.ncbi.nlm.nih.gov/22101427/)]
24. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. *J Am Med Assoc* 2017 Feb 21;317(7):766-768. [doi: [10.1001/jama.2016.18553](https://doi.org/10.1001/jama.2016.18553)]
25. Bar-Hillel M. Position effects in choice from simultaneous displays: a conundrum solved. *Perspect Psychol Sci* 2015 Jul;10(4):419-433. [doi: [10.1177/1745691615588092](https://doi.org/10.1177/1745691615588092)] [Medline: [26177945](https://pubmed.ncbi.nlm.nih.gov/26177945/)]
26. Ert E, Fleischer A. Mere position effect in booking hotels online. *J Travel Res* 2014 Nov 20;55(3):311-321. [doi: [10.1177/0047287514559035](https://doi.org/10.1177/0047287514559035)]
27. Mantonakis A, Rodero P, Lesschaeve I, Hastie R. Order in choice: effects of serial position on preferences. *Psychol Sci* 2009 Nov;20(11):1309-1312. [doi: [10.1111/j.1467-9280.2009.02453.x](https://doi.org/10.1111/j.1467-9280.2009.02453.x)] [Medline: [19843263](https://pubmed.ncbi.nlm.nih.gov/19843263/)]
28. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? *J Med Internet Res* 2012 Apr 10;14(2):e50 [FREE Full text] [doi: [10.2196/jmir.2005](https://doi.org/10.2196/jmir.2005)] [Medline: [22491423](https://pubmed.ncbi.nlm.nih.gov/22491423/)]
29. Ellimoottil C, Leichtle SW, Wright CJ, Fakhro A, Arrington AK, Chirichella TJ, et al. Online physician reviews: The good, the bad, and the ugly. *Bull Am Coll Surg* 2013 Sep;98(9):34-39. [Medline: [24455818](https://pubmed.ncbi.nlm.nih.gov/24455818/)]
30. Rao M, Clarke A, Sanderson C, Hammersley R. Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: cross sectional study. *BMJ* 2006 Jul 1;333(7557):19 [FREE Full text] [doi: [10.1136/bmj.38874.499167.7C](https://doi.org/10.1136/bmj.38874.499167.7C)] [Medline: [16793783](https://pubmed.ncbi.nlm.nih.gov/16793783/)]
31. 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](https://doi.org/10.1136/bmjqs-2012-000906)] [Medline: [22523318](https://pubmed.ncbi.nlm.nih.gov/22523318/)]
32. Glover M, Khalilzadeh O, Choy G, Prabhakar AM, Pandharipande PV, Gazelle GS. Hospital evaluations by social media: a comparative analysis of Facebook ratings among performance outliers. *J Gen Intern Med* 2015 Oct;30(10):1440-1446 [FREE Full text] [doi: [10.1007/s11606-015-3236-3](https://doi.org/10.1007/s11606-015-3236-3)] [Medline: [25749881](https://pubmed.ncbi.nlm.nih.gov/25749881/)]
33. 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/NEJMs1502361](https://doi.org/10.1056/NEJMs1502361)] [Medline: [26267629](https://pubmed.ncbi.nlm.nih.gov/26267629/)]
34. Geierhos M, Bäumer FS, Schulze S, Stuß V. "I grade what I get but write what I think." Inconsistency Analysis in Patients' Reviews. 2015 Presented at: European Conference on Information Systems (ECIS); 2015; Münster, Germany. [doi: [10.18151/7217324](https://doi.org/10.18151/7217324)]
35. Pulkkinen L. Seattle Post-Intelligencer. 2015 Nov 4. Bellevue Doctor Sues Patient who Blasted him on Yelp: Patient Denies Defamation Claim URL: <https://www.seattlepi.com/local/article/Bellevue-doctor-sues-patient-who-blasted-him-on-6611490.php> [WebCite Cache ID 6z973qr0a]
36. Der Spiegel. 2017. BGH-Urteil zu Arztkritik: Bewertungsportale müssen Belege liefern URL: <http://www.spiegel.de/netzwelt/web/bundesgerichtshof-bewertungsportale-muessen-strenger-pruefen-a-1079993.html> [WebCite Cache ID 6z975puUt]
37. Maas B. CBS Denver. 2012 Sep 12. Doctor Suing Former Patient over Online Criticism URL: <http://denver.cbslocal.com/2012/07/19/doctor-suing-former-patient-over-online-criticism/> [WebCite Cache ID 6z97J7w0a]
38. Der Spiegel. 2018 Feb 20. BGH-Urteil zu Jameda: Bewertungsportal muss Ärzte-Profil löschen - und ändert sein Geschäftsmodell URL: <http://www.spiegel.de/netzwelt/web/jameda-aerztin-klagt-erfolgreich-auf-loeschung-aus-aerztebewertungsportal-a-1194390.html> [WebCite Cache ID 6z97ZYVFe]
39. Cylus J, Richardson E, Findley L, Longley M, O'Neil C, Steel D. United Kingdom: health system review. *Health Syst Transit* 2015;17(5):1-126 [FREE Full text] [Medline: [27049966](https://pubmed.ncbi.nlm.nih.gov/27049966/)]
40. Rice T, Rosenau P, Unruh LY, Barnes AJ, Saltman RB, van Ginneken E. United States of America: health system review. *Health Syst Transit* 2013;15(3):1-431 [FREE Full text] [Medline: [24025796](https://pubmed.ncbi.nlm.nih.gov/24025796/)]
41. Staiger DO, Auerbach DI, Buerhaus PI. Trends in the work hours of physicians in the United States. *J Am Med Assoc* 2010 Feb 24;303(8):747-753 [FREE Full text] [doi: [10.1001/jama.2010.168](https://doi.org/10.1001/jama.2010.168)] [Medline: [20179284](https://pubmed.ncbi.nlm.nih.gov/20179284/)]
42. 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](https://pubmed.ncbi.nlm.nih.gov/22546128/)]
43. Samora JB, Lifchez SD, Blazar PE, American Society for Surgery of the Hand Ethics and Professionalism Committee. Physician-rating web sites: ethical implications. *J Hand Surg Am* 2016 Jan;41(1):104-10.e1. [doi: [10.1016/j.jhsa.2015.05.034](https://doi.org/10.1016/j.jhsa.2015.05.034)] [Medline: [26304734](https://pubmed.ncbi.nlm.nih.gov/26304734/)]
44. Customer Flux. 2017. FAQs URL: <https://customerflux.com/faqs/> [accessed 2017-03-08] [WebCite Cache ID 6ooVHKd49]

45. Patinamarketing. 2012 Aug 12. Have You Claimed & Optimized Your Online Physician Profile? URL: <http://patinamarketing.com/2012/08/claimed-and-optimized-your-online-physician-profile/> [accessed 2017-03-08] [WebCite Cache ID 6ooVM6yIs]
46. Reputation Defender. Physician Online Reputation Management URL: <https://www.reputationdefender.com/blog/doctors/physician-online-reputation-management-online-review-management> [accessed 2017-03-08] [WebCite Cache ID 6ooVP44x9]
47. 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 [FREE Full text] [doi: [10.2196/jmir.5889](https://doi.org/10.2196/jmir.5889)] [Medline: [27644135](https://pubmed.ncbi.nlm.nih.gov/27644135/)]
48. Emmert M, Adelhardt T, Sander U, Wambach V, Lindenthal J. A cross-sectional study assessing the association between online ratings and structural and quality of care measures: results from two German physician rating websites. *BMC Health Serv Res* 2015 Sep 24;15:414 [FREE Full text] [doi: [10.1186/s12913-015-1051-5](https://doi.org/10.1186/s12913-015-1051-5)]
49. 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](https://doi.org/10.1136/bmjqs-2012-001527)] [Medline: [23349387](https://pubmed.ncbi.nlm.nih.gov/23349387/)]
50. Lagu T, Greaves F. From public to social reporting of hospital quality. *J Gen Intern Med* 2015;30(10):11397-11399 [FREE Full text] [doi: [10.2196/jmir.3758.8](https://doi.org/10.2196/jmir.3758.8)] [Medline: [25933624](https://pubmed.ncbi.nlm.nih.gov/25933624/)]

Abbreviations

PRW: physician-rating website

Edited by T Kool; submitted 03.10.17; peer-reviewed by S Kleefstra, S Grabner-Kräuter, C Schaefer; comments to author 09.11.17; revised version received 06.01.18; accepted 13.03.18; published 14.06.18.

Please cite as:

Rothenfluh F, Schulz PJ

Content, Quality, and Assessment Tools of Physician-Rating Websites in 12 Countries: Quantitative Analysis

J Med Internet Res 2018;20(6):e212

URL: <http://www.jmir.org/2018/6/e212/>

doi: [10.2196/jmir.9105](https://doi.org/10.2196/jmir.9105)

PMID: [29903704](https://pubmed.ncbi.nlm.nih.gov/29903704/)

©Fabia Rothenfluh, Peter J Schulz. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.06.2018. 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.

Original Paper

Cost-Effectiveness of Facilitated Access to a Self-Management Website, Compared to Usual Care, for Patients With Type 2 Diabetes (HeLP-Diabetes): Randomized Controlled Trial

Jinshuo Li¹, MPhil; Steve Parrott¹, MSc; Michael Sweeting², PhD; Andrew Farmer³, DM, FRCGP; Jamie Ross⁴, PhD; Charlotte Dack⁵, PhD; Kingshuk Pal⁴, PhD; Lucy Yardley^{3,6}, MSc, PhD, CPsychol; Maria Barnard⁷, MSc, FRCP; Mohammed Hudda⁸, MSc; Ghadah Alkhalidi⁹, MPH, PhD; Elizabeth Murray⁴, FRCGP, FRCP, PhD

¹Mental Health and Addiction Research Group, Department of Health sciences, University of York, York, United Kingdom

²Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

³Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

⁴Research Department of Primary Care and Population Health, University College London, London, United Kingdom

⁵Department of Psychology, University of Bath, Bath, United Kingdom

⁶Department of Psychology, University of Southampton, Southampton, United Kingdom

⁷Department of Diabetes & Endocrinology, Whittington Health NHS Trust, London, United Kingdom

⁸Population Health Research Institute, St. George's University of London, London, United Kingdom

⁹Community Health Sciences Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

Corresponding Author:

Elizabeth Murray, FRCGP, FRCP, PhD

Research Department of Primary Care and Population Health

University College London

London, NW3 2PF

United Kingdom

Phone: 44 2077940500 ext 38826

Fax: 44 2077941224

Email: elizabeth.murray@ucl.ac.uk

Abstract

Background: Type 2 diabetes mellitus is one of the most common long-term conditions, and costs health services approximately 10% of their total budget. Active self-management by patients improves outcomes and reduces health service costs. While the existing evidence suggested that uptake of self-management education was low, the development of internet-based technology might improve the situation.

Objective: To establish the cost-effectiveness of a Web-based self-management program for people with type 2 diabetes (HeLP-Diabetes) compared to usual care.

Methods: An incremental cost-effectiveness analysis was conducted, from a National Health Service and personal and social services perspective, based on data collected from a multi-center, two-arm individually randomized controlled trial over 12 months. Adults aged 18 or over with a diagnosis of type 2 diabetes and registered with the 21 participating general practices (primary care) in England, UK, were approached. People who were unable to provide informed consent or to use the intervention, terminally ill, or currently participating in a trial of an alternative self-management intervention, were excluded. The participants were then randomized to either usual care plus HeLP-Diabetes, an interactive, theoretically-informed Web-based self-management program, or to usual care plus access to a comparator website containing basic information only. The participants' intervention costs and wider health care resource use were collected as well as two health-related quality of life measures: the Problem Areas in Diabetes (PAID) Scale and EQ-5D-3L. EQ-5D-3L was then used to calculate quality-adjusted life years (QALYs). The primary analysis was based on intention-to-treat, using multiple imputation to handle the missing data.

Results: In total, 374 participants were randomized, with 185 in the intervention group and 189 in the control group. The primary analysis showed incremental cost-effectiveness ratios of £58 (95% CI –411 to 587) per unit improvement on PAID scale and £5550 (95% CI –21,077 to 52,356) per QALY gained by HeLP-Diabetes, compared to the control. The complete case analysis showed less cost-effectiveness and higher uncertainty with incremental cost-effectiveness ratios of £116 (95% CI –1299 to 1690)

per unit improvement on PAID scale and £18,500 (95% CI –203,949 to 190,267) per QALY. The cost-effectiveness acceptability curve showed an 87% probability of cost-effectiveness at £20,000 per QALY willingness-to-pay threshold. The one-way sensitivity analyses estimated 363 users would be needed to use the intervention for it to become less costly than usual care.

Conclusions: Facilitated access to HeLP-Diabetes is cost-effective, compared to usual care, under the recommended threshold of £20,000 to £30,000 per QALY by National Institute of Health and Care Excellence.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 02123133; <http://www.controlled-trials.com/ISRCTN02123133> (Archived by WebCite at <http://www.webcitation.org/6zqjhm00>)

(*J Med Internet Res* 2018;20(6):e201) doi:[10.2196/jmir.9256](https://doi.org/10.2196/jmir.9256)

KEYWORDS

cost-effectiveness; type 2 diabetes mellitus; self-management; internet

Introduction

Background

There is a global epidemic of diabetes mellitus, with an estimated 10% of the global population, or 422 million people, affected. Around 90% of these people have type 2 diabetes mellitus (T2DM) [1]. The personal and health care costs of T2DM are substantial. Recent estimates suggest that about 11% of the total global health expenditure is due to diabetes [2]. Most of these costs are due to preventable complications [3]. Also, diabetes also results in societal costs, including the cost of missed workdays [4], personal costs, including out of pocket costs [2], and the impact on employment and earnings potential [5,6]. Patient education and self-management support has been identified as a priority for global health in recent years [7] and has the potential to both improve outcomes and reduce costs [8]. However, internationally, uptake of self-management education remains low [9], partly due to logistical problems with attending courses [10].

Web-based self-management support has the potential to increase uptake by overcoming some of the logistical problems associated with other forms of delivery as it can be accessed at home, at the user's convenience. We have developed a comprehensive, evidence-based, theoretically informed, Web-based self-management program for adults with T2DM called Healthy Living for People with Type 2 Diabetes (HeLP-Diabetes). Overall content was guided by the Corbin and Strauss model, which hypothesizes that patients must undertake medical, emotional and role management in dealing with a long-term condition [11]. If effective, Web-based interventions have the potential to be highly cost-effective, as they can be delivered at scale across large populations, with relatively low additional costs per additional user [12], unlike telephone-based or face-to-face education where labor costs account for a substantial proportion of total cost [13].

We undertook an individually randomized controlled trial in primary care to determine the effectiveness and cost-effectiveness of HeLP-Diabetes compared to a simple, text-based website, all with the access to usual care for people with T2DM [14]. The aim of this paper is to present the health economic analysis of this comparison based on the data collected in the trial and to examine the cost-effectiveness of facilitated access to HeLP-Diabetes. The analysis was undertaken from the collective perspective of the National Health Services (NHS)

and personal social services (PSS), following the National Institute of Health and Care Excellence (NICE) guidance [15], as the two share the same resources. The results on clinical effectiveness are reported in a separate article previously published [16]. All costs are presented in pounds sterling (£) 2014 prices.

Methods

Approval and Ethical Considerations

Ethics approval was obtained from Camden and Islington National Research Ethics Service (NRES) committee, reference 12/LO/1571.

Design, Setting, and Participants

The HeLP-Diabetes trial was a multi-center, two-arm individually randomized controlled trial carried out in primary care settings in England, United Kingdom (UK). The detailed trial design was fully reported in the published protocol [14], and the clinical effectiveness article [16]. There were no changes to the methods after the protocol was agreed and the start of the trial. Twenty-one general practices from across England participated, with a mix of urban, suburban, and rural practices. Adults, aged 18 years or over, registered with participating practices, and diagnosed with T2DM were eligible for inclusion in the trial. People who were unable to provide informed consent (eg, due to psychosis or cognitive impairment), unable to use the intervention (eg, due to physical, sensory or intellectual impairment, or inability to understand basic spoken or written English), terminally ill, or currently participating in a trial of an alternative self-management intervention, were excluded. There were no exclusions based on the duration of diagnosis, level of diabetes control, previous experience of self-management education, computer and internet experience, or access to the internet at home.

Eligible participants were briefed on the trial by a practice or research nurse (See the Patient Information Sheet in [Multimedia Appendix 1](#)) and then individually randomized to either the intervention or control group using Web-based randomization independently of the trial team. Randomization was conducted in a 1:1 ratio using random permuted blocks of sizes 2, 4, and 6, stratified by recruitment center. Participants were informed the trial compared two forms of Web-based support but were blinded as to which was the intervention and which the comparator. Each participant had access to their allocated intervention for

12 months after randomization. Follow-up was undertaken at 3 and 12 months postrandomization.

Intervention

The intervention consisted of facilitated access to HeLP-Diabetes. Facilitation consisted of an introductory training session with practice nurses. In this session, patients were shown on a computer how to log on and set a username and password, and introduced to the structure, contents, and features of the website and how to navigate it. A booklet summarizing the information introduced in the session was given to the patients to take home.

HeLP-Diabetes was a theoretically informed, Web-based program, whose overall goals were to improve health outcomes and reduce diabetes-related distress [17]. It was developed using participatory design principles, with substantial input from patients with T2DM and health professionals caring for such patients. The content was designed to be accessible to people with a wide range of literacy and health literacy skills, with all essential content provided in both video and text. The content sections covered information on diabetes as a medical condition and its impact on people's life; behavior change components to support adoption of healthier lifestyles; and a third strand of components focusing on emotional well-being based on cognitive behavioral therapy and mindfulness.

The program also included an online forum where the participants could post and share their questions, concerns, and experiences. There was also an "Ask the Expert" facility, where questions were reviewed and responded to by a multi-disciplinary team including an information scientist, clinicians, and patient representatives (see [Multimedia Appendix 2](#)). The forum was monitored daily by both research staff and patient representatives.

Engagement with the program was promoted through regular newsletters, emails and mobile text messages containing updates on latest diabetes-related research or practice, seasonally-relevant advice, and links to specific relevant parts of the program. A medical information scientist reviewed the diabetes-related research published each month and provided a summary of the important, useful or relevant research. The summary was then discussed by a team of clinicians, psychologists, health service researchers and patient representatives before selected items of interest were written up for a patient audience. A more detailed description of the intervention is provided in a separate clinical article [16] and a National Institute for Health Research (NIHR) monograph [18].

Comparator

HeLP-Diabetes was designed to be provided as an addition to current practice. However, to improve acceptability to participants and to maintain blinding, all participants were assigned access to a website. Participants in the control group were given access to a simple information website, based on the information readily available in the public domain on the website of the main UK diabetes charity (Diabetes UK) or National Health Service patient information website (NHS Choices). Participants in the control group were also given an introductory facilitation meeting, in which they were shown

how to navigate the website, and an information booklet to take home.

Health Outcomes

The health outcomes for the health economic analysis were diabetes-related distress, measured by the Problem Areas in Diabetes (PAID) questionnaire, and quality of life, measured by EQ-5D-3L. The PAID questionnaire consisted of 20 items focusing on areas that cause difficulty for people living with diabetes, including social situations, food, friends and family, diabetes treatment, relationships with health care professionals and social support [19]. PAID scores range from 0-100, with lower scores indicating less stress, with a score of 40 or more indicating significant distress. EQ-5D-3L is a standardized instrument for measuring health-related quality of life, which has five domains (5D), each with three levels (3L) measuring daily difficulties in that domain [20]. Both these self-reported outcome measures were collected online at baseline, three months, and 12 months follow-up. The tariff for each combination of the EQ-5D-3L levels for the UK population was applied to calculate utility values [21]. The utility values range from -0.594 to 1, with higher values indicating better quality of life. Quality-adjusted life years (QALYs) were then calculated over the duration of the trial using the area under the curve of utility values from the three time points [22]. QALYs were not discounted because the assessing period was 12 months.

Costs of the Intervention

There were two types of costs related to the intervention: those incurred during the development and optimization of the intervention; and those related to ongoing delivery and maintenance of the intervention. Development costs were not taken into account for this analysis. As per NICE guidance [15], the evidence on costs should relate to the National Health Services (NHS) and personal social services (PSS) resources. The development costs, in this case, although considerable, related to research funding rather than NHS and PSS resources, and unlikely to be repeated if the intervention is adopted in practice.

The costs relating to ongoing maintenance and delivery of the intervention within the trial form the basis for the current analysis. These consisted of: the cost of delivery of the intervention; the cost of maintenance and updating of the intervention; and cost of facilitating activities undertaken to improve uptake and use. If the intervention were to be widely implemented into routine health care, all these activities would be required on an on-going basis. Therefore, the costs of these activities that occurred during the trial were used to estimate the real costs in practice.

Delivering, maintaining, and updating the intervention involved two types of costs: costs related to hardware and software; and staff costs. Staff costs related to activities for engagement, moderating the online forum, revising the content of the website, and responding to 'Ask the Expert' questions. Staff costs were also incurred by the third-party service provider responsible for hosting and maintaining the intervention.

Costs related to hardware, software, and work undertaken by the third-party service provider were recorded from actual

invoices. These included a weekly review of recent development in the field, domain names purchase, website hosting, quarterly maintenance of the website, and Security Socket Layer certificate purchase. Some of the third-party services were contracted for longer than 12 months, so costs were calculated for a one-year period based on the invoiced amount and their length of service.

Activities not undertaken by the third-party providers were carried out by either professional staff or by patient representatives. Costs related to work undertaken by patient representatives were recorded from the payments made to representatives, who were reimbursed for their time in line with INVOLVE guidance [23]. INVOLVE is a NIHR funded national advisory group to support active public involvement in NHS, public health and social care research. These included patient representatives' feedback on development review, forum monitoring, and their review of feedback from the clinical team. Costs related to activities undertaken by research staff and clinicians were estimated from workloads during the trial period, by recording the time taken for each activity, the frequency of that activity and the number and grade of staff involved. These included writing and sending emails, short message service (SMS), and newsletters, forum monitoring, interaction with patients on the website, and contents review and update. The costs were then calculated by multiplying the time spent by the average wage for each type of staff member. Hourly costs for research staff were taken from the academic pay scale [24], and hourly costs for General Practitioners (GPs) were taken from unit costs of health and social care edited by Personal Social Services Research Unit (PSSRU) [25].

Participants in both groups were provided with facilitation by practice or research nurses to encourage use and uptake of the intervention and comparator. Participants were also provided with a booklet, summarizing the training they had received (login details, how to use the website). All participants were provided with this introductory session and booklet. As this activity was not required for the comparator website and was only undertaken to maximize comparability between the intervention and the comparator, we considered it a research activity and assumed that no intervention costs were incurred in the control group.

Practice/research nurses required training in providing this introductory session. Each nurse attended an hour face-to-face training session provided by a member of the research team. They were also provided with printed training materials, reminding them how to register patients and how to introduce the website. The costs of the training and introductory session were calculated from the time spent on each activity by the

nurses and research team staff, plus travel time for the research team, multiplied by their respective hourly salary rates. In this case, we took the hourly rate of practice nurses for all nurses' time. The costs of the printed training materials and the booklets issued to participants were obtained from invoices from the printers.

All intervention costs were allocated to the participants in the intervention group of the trial to give a per participant cost.

Health Service Resource Use

Health care resource use, including primary, secondary and community services, was collected for both groups using bespoke service use questionnaires. The majority of information about service use and participants' prescriptions were extracted from participants' medical records by practice or research nurses. The remainder of the service use data were collected retrospectively from participants using a self-report questionnaire online. These data were collected at baseline for the 12 months period before the trial, at three months follow-up for the three months period after randomization and at 12 months follow-up for the nine months after three months follow-up. The quantities were then multiplied by a set of national average unit costs (Table 1). The total costs of health services were then summarised at an aggregated level, (ie, costs of health services use from data provided by nurses, medication costs from data provided by nurses, and costs of health and social services use from data provided by participants), for the corresponding periods respectively. Any missing data on individual services resulted in a missing cost for the entire section.

Where applicable, Value Added Tax at 20%, salary on-costs, and overheads were added. Unit costs for out-of-hours services were estimated based on a national audit [26], assuming the duration of consultations was the same as for in-hours services. Data on travel time for home visits were not available, so we adopted the assumed 12 minutes per visit estimate made by PSSRU 2014 [25]. No allowance was made for travel expenses.

Due to the large amount of medications taken by this particular population, only current prescriptions taken at the time of data collection were extracted. Prescribed items were matched with the Prescription Cost Analysis England 2014 [28] for a cost per item, using their generic name, form, and strength where available. In the absence of full information, a weighted average cost per item was calculated based on available information. Unless it was specified that no medication was prescribed, blank entries were considered as missing data. We also assumed that all prescriptions were for chronic conditions and issued for one month at a time over the corresponding period. Costs were not discounted as the assessing period was 12 months.

Table 1. National average unit cost used in the analysis.

Health Service Use	Unit cost (£) (per consultation or per episode)	Sources
GP^a consultation		
In surgery	38	PSSRU ^b 2014 [25]
Home visit	62	PSSRU 2014 [25]
Telephone	23	PSSRU 2014 [25]
Practice nurse consultation		
In surgery	11	PSSRU 2014 [25]
Home visit	18	PSSRU 2014 [25]
NHS ^c Walk-In Clinic	56	PSSRU 2014 [25]
Out-of-hour services		
Telephone advice	36	Out-of-hours GP services in England [26], PSSRU 2014 [25]
Home visit	117	Out-of-hours GP services in England [26], PSSRU 2014 [25]
In surgery	86	Out-of-hours GP services in England [26], PSSRU 2014 [25]
Accident and emergency service admission	167	Reference Costs 2013-14 [27]
Podiatrist	44	Reference Costs 2013-14 [27]
Optometry	97	Reference Costs 2013-14 [27]
Physiotherapy	46	Reference Costs 2013-14 [27]
Counselling		
Primary care	46	PSSRU 2014 [25]
Community	138	Reference Costs 2013-14 [27]
Clinical test	2	Reference Costs 2013-14 [27]
Outpatient appointment	111	Reference Costs 2013-14 [27]
Day case	698	Reference Costs 2013-14 [27]
Inpatient admission	1891	Reference Costs 2013-14 [27]
District nurse		
Home visit	46	Reference Costs 2013-14 [27], PSSRU 2014 [25]
In surgery or clinic	37	Reference Costs 2013-14 [27]
Social worker	55	PSSRU 2014 [25]
Occupational therapy	64	Reference Costs 2013-14 [27]
Dietician	80	Reference Costs 2013-14 [27]

^aGP: general practitioner.^bPSSRU: Personal Social Services Research Unit.^cNHS: National Health Service.

Analysis

Missing Data

Multiple imputation was used as the primary method to account for missing data at both baseline and follow-up. A chained equation model was developed, and predictive mean matching was used as the imputation method for continuous variables, using the five nearest neighbors to the prediction as a set to draw from. All missing data were imputed separately by trial group. The imputation was performed at the aggregated level

(ie, costs of health services use from data provided by nurses, prescription costs from data provided by nurses, costs of health and social services use from data provided by participants, PAID score and EQ-5D-3L utility value from data provided by participants). PAID and EQ-5D-3L data collected outside of the pre-specified “window” of 10-14 months following randomization were considered invalid as 12-month outcomes and only used as imputing factors along with baseline and other outcome data. The percentage of missing data served as the base of the number of imputations, as a rule of thumb [29].

Primary Analysis

The primary analysis followed a pre-specified analysis plan, comparing the groups as randomized (intention-to-treat). It focused on a within-trial analysis of costs and benefits, with no projected time horizon. A linear mixed effects model was fitted with the 12-month outcome as the dependent variable, adjusting for the baseline variables age, sex, presence of pre-existing cardiovascular disease, duration of diabetes, smoking status and corresponding baseline outcome (costs over 12 months before baseline, PAID score at baseline and EQ-5D-3L utility value at baseline, respectively) as fixed effect terms. Center effects were included as random-effects in the analysis. No time-dependent terms, interaction terms or effect modifiers were used. The difference in mean 12-month costs and outcomes were estimated based on the model. Incremental cost-effectiveness ratios (ICER) were calculated by dividing the difference in mean cost by the difference in mean health outcome. The resulting incremental cost per QALY gained was then compared against the recommended willingness-to-pay (WTP) threshold of £20,000-£30,000 per QALY by NICE [15].

A non-parametric bootstrap technique was employed to explore the uncertainty of point estimates of the difference in mean 12-month costs and outcomes from primary analyses. Five thousand bootstrapped datasets were created and the total costs and outcome estimated for each one. The results from bootstrap resampling were used to construct 95% CI for incremental costs, incremental PAID score, incremental QALYs, and to plot the cost-effectiveness plane and cost-effectiveness acceptability curve to show the uncertainty surrounding the primary results [30].

Sensitivity Analyses

Complete case analysis was undertaken to assess the performance of the imputation model compared to a complete case analysis that assumes data were missing completely at random.

The intervention cost per user estimated in the primary analysis was based on the number of trial participants, which would be unrealistic if the intervention were to be implemented more widely. We, therefore, undertook a one-way sensitivity analysis, exploring the cost per user as numbers of users increased.

The cost of nurse-led facilitation is per participant, so will be incurred by each registered user, regardless of the total number. Each practice referring patients to the intervention needed at least one nurse who is trained in the facilitating activities. In this trial, each practice or research nurse only undertook to facilitate activities for up to 10 patients, whereas practices are likely to have several hundred patients with T2DM who could be registered, which would reduce the per-user cost of training nurses to register patients. However, taking into consideration staff turnover and the possibility of multiple staff being trained, we made the conservative assumption that the costs of training practice staff would remain the same per user.

The costs of maintaining and delivering the intervention remain the same up to the total capacity of the current server (which is

for 10,000 active users). This means that the average cost per user reduces up to 10,000 active users. The one-way sensitivity analysis, therefore, explored the change in intervention cost per user about the ICER on the implicit assumption that the impact on health services use, and QALYs in the trial was generalizable. Because the national WTP threshold is only expressed regarding incremental cost per QALY, the one-way sensitivity analysis was not undertaken for the incremental PAID score.

All analyses were undertaken using STATA SE 14.2 software.

Results

Participants in the Study

Recruitment took place between September 2013 and December 2014. Of 374 participants randomized, 185 were allocated to the intervention and 189 to the control group. The average age at randomization was 64.9 (SD 9.5) years old in the intervention group and 64.7 (SD 9.1) years old in the control group. There were four participants missing time since diagnosis of diabetes, two in each group. Among the participants with this information, the mean duration was 7.8 (SD 5.7) years in the intervention group and 8.2 (SD 6.1) years in the control group. The proportion of male participants in the intervention group (127/185, 69%) was similar to that in the control group (131/189, 69%). Further description of the participants' characteristics can be found in the clinical effectiveness article [16].

Missing Data

In the intervention group, there were 143/185 (77%) participants who completed the three months self-report questionnaire and 129/189 (70%) participants who completed the 12 months one. In the control group, there were 152/189 (80%) participants who completed the three months questionnaire and 135/189 (71%) participants who completed the 12 months questionnaire. The difference in completion rate was not significant (Pearson's chi-squared test $P=.459$ at 3 months, $P=.718$ at 12 months). There were rare cases where participants completed the questionnaire with one or two items missing (see [Multimedia Appendix 3](#)). In general, the missing data on self-report were due to not responding. The percentage of missing data was low and did not exceed 30% for any variable. The number of imputations was therefore set to 30.

Intervention Costs

Maintenance and Delivery

Staff costs of maintenance and delivery of the intervention were estimated to be £18,783 a year, including patient representatives' feedback work (Table 2). The total infrastructure cost per year was estimated at £23,013, including £93 for 10 domain names, £3600 for website hosting, £19,200 for maintenance, and £120 for Security Socket Layer Certificate. The two parts gave a total operating cost of HeLP-Diabetes per year of £41,796. Allocating these costs to the 185 participants in the intervention group, the average cost per participant was £226.

Table 2. Staff costs of maintenance and delivery of HeLP-Diabetes.

Activities	Intensity	Payment scale or method	Unit cost (£) per hour	Cost (£) per year
Emails, newsletters, SMS ^a	1 day/2 weeks	Grade 6 staff	34	6675
Librarian review of recent development	1 hour/week	Fixed contract	30	1560
Patient representatives feedback on librarian's review	30 minutes/person, 2 persons/2 weeks	Cash payment	18	468
Forum monitoring by patient representatives	30 minutes/person, 2 persons/2 weeks	Cash payment	18	468
Forum monitoring by staff	1 hour/2 weeks	Grade 6 staff	34	890
Clinical team website interaction	1 hour/time, 5 times/year	GP	121	605
Patient representatives review feedback from clinical team	15 minutes/person, 2 persons/time, 5 times/year	Cash payment	18	45
Content checking, revising and updating by staff	2 hours/2 weeks	Grade 6 staff	34	1780
Content checking, revising and updating by clinical team	2 hours/2 weeks	GP	121	6292
Total				18,783

^aSMS: short message service.

Facilitating Activities

The costs of training the nurses to undertake the facilitation, including both the costs of the research staff providing the training and the costs of the nurses being trained, came to £3785 across all practices involved in the trial. The cost of the printed training materials was £78. Thus, the total training costs in the trial were £3863, or £21 per participant in the intervention group.

The time allocated to registering participants and introducing them to the intervention was 20 minutes per participant, hence cost one-third of a practice nurse's hourly consultation rate (£44) [25]. The cost of the booklet given to each participant in the intervention group was £0.95 per booklet.

The total intervention cost was, therefore, £263 per participant, made up of £226 costs of maintenance and delivery; £21 for initial training of practice/research nurses; and £16 for nurse-led facilitation. We assumed zero costs for the comparator, as although we incurred costs during the trial, these would not have been incurred outside of a trial.

Primary Analysis

The outcomes and the incremental cost and effectiveness of the primary analysis are presented in Table 3. The mean costs of health resources use in the 12 months trial period were higher in the control group than in the intervention group (for detailed health resources use, see Multimedia Appendix 3). The unadjusted difference in mean total costs was £12, with a lower

value in the control group. When adjusted for baseline health resource use 12 months before the trial, the mean total costs in the intervention group were £131 (SE £169) higher than the control group. After further adjusting for baseline variables (ie, age, sex, history of cardiovascular diseases, smoking status, time since diabetes diagnosis), the difference was reduced to £111 (95% CI –156 to 362).

The mean PAID score was higher in the control group than the intervention group throughout the trial. The unadjusted difference in PAID score at 12 months was 3.1, with a lower score in the intervention group. After adjusting for baseline PAID score, the mean PAID score was 1.9 (SE 1.3) lower in the intervention group than the control group. This remained similar when further adjusting for other baseline variables (ie, age, sex, history of cardiovascular diseases, smoking status, time since diabetes diagnosis).

Mean EQ-5D-3L utility in both groups increased at three months from baseline. At 12 months, the mean utility value fell in both groups, but the fall was greater in the control group. Regarding QALYs, during the 12-month trial period, the intervention group had a mean QALY of 0.802 (SE 0.016) compared with 0.764 (SE 0.023) in the control group, giving an unadjusted difference of 0.038. After adjusting for EQ-5D-3L utility values at baseline and other baseline variables (ie, age, sex, history of cardiovascular diseases, smoking status, time since diabetes diagnosis), the incremental QALY was 0.020 (95% CI –0.001 to 0.044), comparing the intervention group to the control group.

Table 3. Incremental cost-effectiveness analysis based on imputed data, by randomized group.

Outcomes	Intervention (n=185)	Control (n=189)
Costs (£)		
Intervention cost, mean	263	0
Health resources use in the 12 months trial period ^a , mean (SE)	1816 (125)	2067 (144)
Total cost during trial period, mean (SE)	2079 (125)	2067 (144)
Incremental cost for intervention group, adjusted for health resources use cost at baseline and other baseline variables, mean (95% CI)	111 (–156 to 362) ^b	N/A ^c
HRQoL^d		
Baseline PAID ^e , mean (SE)	18.1 (1.3)	19.9 (1.4)
Three months PAID, mean (SE)	15.7 (1.2)	17.3 (1.3)
Twelve months PAID, mean (SE)	14.5 (1.2)	17.6 (1.4)
Incremental PAID score at 12 months for intervention group, adjusted for baseline PAID score and other baseline variables, mean (95% CI)	–1.9 (–4.2 to 0.4) ^b	N/A
Baseline EQ-5D-3L utility ^f , mean (SE)	0.793 (0.018)	0.766 (0.021)
Three months EQ-5D-3L utility, mean (SE)	0.811 (0.016)	0.786 (0.024)
Twelve months EQ-5D-3L utility, mean (SE)	0.793 (0.023)	0.736 (0.037)
QALYs in the 12 months trial period ^g , mean (SE)	0.802 (0.016)	0.764 (0.023)
Incremental QALYs for intervention group, adjusted for baseline EQ-5D-3L utility value and other baseline variables, mean (95% CI)	0.020 (–0.001 to 0.044) ^b	N/A
ICER^h		
Incremental cost (£) per unit improvement on PAID scale, mean (95% CI)	58 (–411 to 587) ^b	N/A
Incremental cost (£) per QALY gained, mean (95% CI)	5550 (–21,077 to 52,356) ^b	N/A

^aThis was calculated based on the assumption that the medications were for chronic use and were prescribed monthly.

^bBaseline variables included age, sex, history of cardiovascular diseases, smoking status, time since diabetes diagnosis.

^cN/A: not applicable.

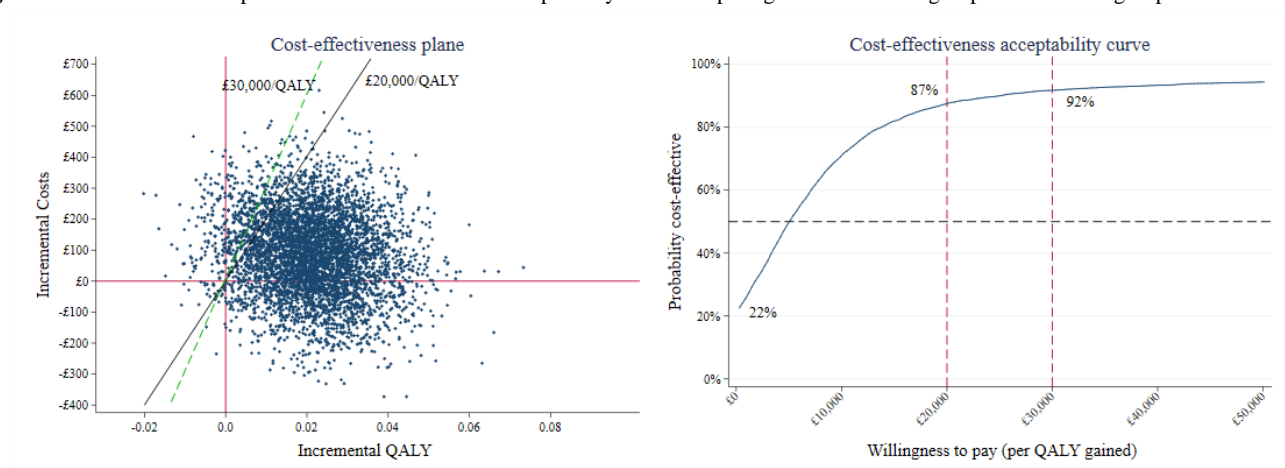
^dHRQoL: health-related quality of life.

^ePAID: Problem Areas in Diabetes.

^fEQ-5D-3L: A descriptive system of health-related quality of life state.

^gQALY: quality-adjusted life year.

^hICER: fully adjusted incremental cost-effectiveness ratio.

Figure 1. Cost-effectiveness plane and cost-effectiveness acceptability curve comparing the intervention group to the control group.

Incremental Cost-Effectiveness Analysis

The fully adjusted incremental cost-effectiveness ratio (ICER) was £58 (95% CI –411 to 587) per unit improvement on PAID scale, and £5550 (95% CI –21,077 to 52,356) per QALY gained, comparing the intervention with the control (Table 3). For both health outcomes, the ICERs indicated a costlier and more effective intervention. Comparing with the recommended WTP threshold for QALY, the intervention presented a much lower ratio. The majority of the bootstrapped replicates fell in the north-east corner of the cost-effectiveness plane, indicating a

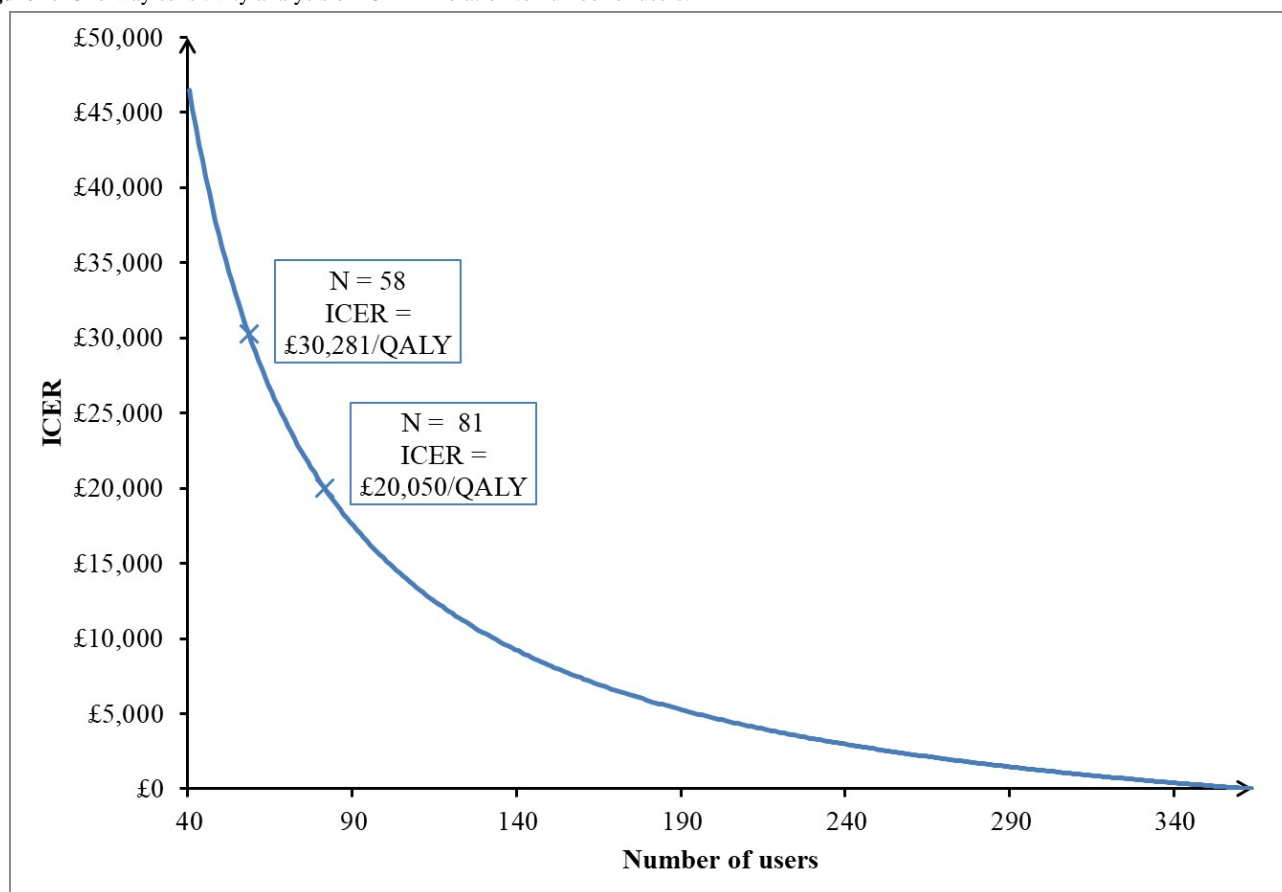
costlier and more effective intervention (Figure 1). A lesser proportion of the replicates fell in the south-east corner, indicating a less costly but more effective intervention. A small group (153/5,000, 3%) of the replicates fell on the left side of the Y axis, indicating a less effective intervention. Overall, the majority of bootstrapped replicates fell under the WTP thresholds. The CEAC further demonstrated the conclusion was likely to be robust with an 87% probability that the intervention was cost-effective at a WTP threshold of £20,000 per QALY and 92% at £30,000 per QALY (Figure 1).

Table 4. Comparison of outcomes between imputed data and complete cases.

Analysis	Intervention		Control	
	n	Mean	n	Mean
Costs of health services use (£)				
In the 12 months before recruitment				
Imputed (SE)	185	1792 (126)	189	2084 (164)
Complete cases (SD)	96	1793 (1,545)	101	1677 (1,418)
In the 12 months trial period				
Imputed (SE)	185	1816 (125)	189	2067 (144)
Complete cases (SD)	96	1695 (1,404)	101	1721 (1,539)
PAID^a scores				
Baseline				
Imputed (SE)	185	18.1 (1.3)	189	19.9 (1.4)
Complete cases (SD)	96	18.8 (16.8)	101	19.0 (16.5)
Twelve months				
Imputed (SE)	185	14.5 (1.2)	189	17.6 (1.4)
Complete cases (SD)	96	14.6 (15.5)	101	15.9 (15.2)
EQ-5D-3L^b utility				
Baseline				
Imputed (SE)	185	0.793 (0.018)	189	0.766 (0.021)
Complete cases (SD)	96	0.792 (0.232)	101	0.829 (0.207)
Three months				
Imputed (SE)	185	0.811 (0.016)	189	0.786 (0.024)
Complete cases (SD)	96	0.824 (0.186)	101	0.840 (0.229)
Twelve months				
Imputed (SE)	185	0.793 (0.023)	189	0.736 (0.037)
Complete cases (SD)	96	0.814 (0.218)	101	0.825 (0.250)

^aPAID: Problem Areas in Diabetes.

^bEQ-5D-3L: A descriptive system of health-related quality of life state.

Figure 2. One-way sensitivity analysis of ICER in relation to number of users.

Complete Case Analysis

A complete case analysis (CCA) was performed on 96/185 (52%) in the intervention group and 101/189 (53%) in the control group who had complete data on baseline variables required for adjustment, costs and the two health outcomes. Comparing the results from CCA to the primary analysis, the mean costs were lower in the CCA than in the primary analysis, except for the mean costs in the 12 months before recruitment in the intervention group (Table 4). Similarly, EQ-5D-3L utility scores at all time points were higher in the CCA than in the primary analysis, except for the score at baseline in the intervention group. The results of the CCA showed a smaller incremental outcomes (PAID -1.6, 95% CI -5.1 to 1.4 and QALY 0.010, 95% CI -0.018 to 0.044) and bigger incremental costs (£185, 95% CI -145 to 504), with higher variation surrounding each estimate, in comparison with the primary analysis. The ICER for a one unit improvement on PAID score was £116 (95% CI -1299 to 1690). The intervention was still cost-effective (ICER=£18,500/QALY, 95% CI -203,949 to 190,267), compared with the WTP threshold, but with a higher level of uncertainty (probability of being cost-effective was 45% at £20,000 and 52% at £30,000).

One-Way Sensitivity Analysis for the Cost of Intervention

Assuming the effectiveness and the impact on health resources use remained the same, the ICER declined rapidly with increasing numbers of users (Figure 2). Once there were 81 users, the ICER dropped to around £20,000/QALY. The ICER

became negative upon the intervention reaching 363 users, thereafter dominating the control (ie, the intervention became less costly and more effective than the control).

Figure 2. One-way sensitivity analysis of ICER in relation to the number of users

Discussion

Principal Results

In this within-trial economic evaluation, we found that HeLP-Diabetes plus usual care was highly likely to be cost-effective, compared with free information-only websites, such as that provided by NHS Choices or Diabetes UK, plus usual care. The ICER in the primary analysis was estimated at £5550 per QALY gained with 87% and 92% probability of being cost-effective for WTP thresholds of £20,000 and £30,000 per QALY, respectively. Once there were over 363 users, HeLP-Diabetes became dominant (ie, less costly and more effective) compared to free information websites, on top of usual care.

Strengths and Limitations

The strength of this study comes from the reliance on actual data from a trial. We collected a wide range of information on health and social service resource use, including community, primary and secondary care. Most of the service use data came from the GP electronic health care records, supplemented by self-report data from participants for services not recorded in the electronic notes. This approach reduced the potential bias

coming from retrospective recall by participants. The completeness of self-report sections was mostly secured by the mandatory questionnaire procedure which prevented skipping individual questions, therefore reduced missing data. Due to the volume of medication taken in this population, however, only the current medication at the time of data collection was extracted, and the assumption was made that all medications were for chronic use. This could overestimate medication costs, but it should not introduce a bias towards either group. The comparison between CCA and the primary analysis showed that the imputation had a greater impact on the control group than on the intervention group. The results suggested that the participants who completed all tests and questionnaires were likely to have a lower level of health services use and better health.

We decided at the outset not to include the investment costs incurred in the development of the intervention in our analysis. This was because, being funded by research grants, these costs were unlikely to re-occur if the intervention was widely implemented by the NHS, therefore being irrelevant to the decision maker for future planning. We used the running costs occurred in the trial to estimate the operating costs in practice. However, there may be more maintenance costs over a longer period as technology changes, and there may be a need for software updates. One example would be that adapting this program to optimize user experience on a mobile phone screen would incur additional programming costs. Neither PAID nor QALYs are clinical outcomes, and their association with outcomes such as glycated hemoglobin (HbA_{1c}) is not always straightforward. Although necessary for comparability with other interventions, their clinical relevance is limited.

According to National Diabetes Audit 2016-17 [31], 56% of registered T2DM patients are male, and 47% are under age 65 in England. While the mean age is similar, the proportion of male patients in our study participants is higher. What impact this might have on the overall effects of HeLP-Diabetes is unknown at this point, but it is worth noting that many face-to-face self-management interventions appear to appeal more to women than men, so the gender imbalance in favour of men seen in our trial may suggest that providing both face-to-face and online self-management interventions may be one way to reach both men and women [32]. Furthermore, the participants in our study were well-managed in their condition from the beginning, with mean HbA_{1c} of 7.3% (56 mmol/mol) at baseline. In a related point, the participants in our study had a duration of diagnosis ranging from 0-34 years. However, the effectiveness data from the trial showed that neither duration of diabetes nor baseline HbA_{1c} impacted on the overall change in HbA_{1c} [16].

Population impact of an intervention is a product of effect size, reach and uptake—a highly efficacious intervention that is only used by a very small proportion of the population may have less impact than a less efficacious intervention which is widely used. These trial data cannot give us an estimate of the eventual reach and uptake of this intervention, as trial participants are known to differ from the total population of patients who may be targeted by an intervention [33]. In parallel with the trial

reported here, we undertook an implementation study to explore issues around reach and uptake, along with factors which impacted on these. These data will be reported separately.

A potential disadvantage of a rigorous evaluation is the time and resource required. There may be questions about the value of such investment in a rapidly evolving landscape where the intervention in question may be rendered obsolete by the time the evaluation is completed. However, while the digital technology changes rapidly, the underlying principles do not. Another potential limitation to consider is the importance of taking into account the target population of the intervention. It is possible that benefits will not scale across the population equally. For instance, in 2017, 37% of the adults aged 65+ in Great Britain read online newspapers or magazines, in contrast to over 70% of the adults aged under 55, whilst 20% of adults aged 65+ shared self-created content online, compared with over 50% in the age groups under 55 [34]. Interventions that have a wide target population will need to reflect the heterogeneity in preferences of their target populations, and the use and impact of the interventions may vary, affecting cost-effectiveness.

Comparison with Prior Work

Although one of the major drivers for research into digital health interventions, such as HeLP-Diabetes, is their expected cost-effectiveness [12,35], there has been relatively little published evidence. A Cochrane review of digital interventions for alcohol consumption published in 2017 identified only 7/42 (16.7%) qualified studies reported economic evaluations [36]. Several systematic reviews of economic evaluations on mental health-related digital interventions conducted in recent years identified 5 studies for anxiety disorder [37], 16 articles for mental health in general [38], and 12 studies for depression [39], respectively. Other economic evaluations of digital health interventions cover a wide range of target conditions, including irritable bowel syndrome [40], substance misuse [41], weight management [42], insomnia [43,44], eating disorders [45,46] and postoperative recovery [47]. Results of these studies were favourable regarding costs, especially when wider health care or societal costs were taken into account but did not show a significant impact on quality of life during follow-up periods. Our results are therefore broadly in line with previous findings and thus contribute to the growing but still insufficient evidence pool of economic evaluation of digital interventions.

Conclusions

As there are 3.5 million people diagnosed with T2DM in Great Britain in 2016 [48-50], and over 90% of the households have internet access in 2017 [34], HeLP-Diabetes has the potential of delivering an effective intervention on a wide scale with negligible marginal costs. Although we do not expect internet-based interventions to be suitable for everyone at the moment, with the internet further permeating our daily life and people adapting to the internet era, there is potential for digital health interventions to help alleviate the burden of chronic conditions on health care systems in the long run. Our findings supported the cost-effectiveness of the intervention once taken up by patients. Further research is needed on how digital health interventions such as HeLP-Diabetes can be delivered and

maintained in a sustainable and cost-effective manner, with the focus on user experience outside of a study setting. The successful realization of this effect might lie in identifying the more susceptible user groups and engaging them at an optimal

time. More empirical studies are needed to help plan the systematic incorporation of digital interventions in medical practice.

Acknowledgments

This project was funded by the NIHR under its Program Grants for Applied Research Program (RP-PG-0609-10135) from 1 March 2011 to 29 August 2016, a total of £2,015,521.00.

Work conducted at the Cardiovascular Epidemiology Unit, University of Cambridge by MS and MH was additionally funded by the UK Medical Research Council (MR/L003120/1), British Heart Foundation (RG/13/13/30194) and UK National Institute for Health Research Cambridge Biomedical Research Centre.

The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report and the manuscript.

Conflicts of Interest

AF is an NIHR Senior Investigator and receives funding from NIHR Oxford Biomedical Research Centre. Elizabeth Murray is the Managing Director of HeLP Digital Community Interest Company, a not-for-profit Social Enterprise, set up to disseminate HeLP-Diabetes across the NHS. She has not, does not, and will not, take any financial remuneration for this role.

Multimedia Appendix 1

Patient information sheet used in the trial.

[[PDF File \(Adobe PDF File\), 101KB](#) - [jmir_v20i6e201_app1.pdf](#)]

Multimedia Appendix 2

Screenshot of HeLP-Diabetes Forum and help page.

[[PNG File, 167KB](#) - [jmir_v20i6e201_app2.png](#)]

Multimedia Appendix 3

The summary of health resources use by group, at baseline, three months and twelve months follow-ups.

[[PDF File \(Adobe PDF File\), 37KB](#) - [jmir_v20i6e201_app3.pdf](#)]

Multimedia Appendix 4

CONSORT-EHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 522KB](#) - [jmir_v20i6e201_app4.pdf](#)]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *The Lancet* 2016 Apr 09;387(10027):1513-1530 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(16\)00618-8](#)] [Medline: [27061677](#)]
2. da Rocha Fernandes J, Ogurtsova K, Linnenkamp U, Guariguata L, Seuring T, Zhang P, et al. IDF Diabetes Atlas estimates of 2014 global health expenditures on diabetes. *Diabetes Res Clin Pract* 2016 Jul;117:48-54. [doi: [10.1016/j.diabres.2016.04.016](#)] [Medline: [27329022](#)]
3. Alva ML, Gray A, Mihaylova B, Leal J, Holman RR. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). *Diabet Med* 2015 Apr;32(4):459-466. [doi: [10.1111/dme.12647](#)] [Medline: [25439048](#)]
4. Bishu KG, Gebregziabher M, Dismuke CE, Egede LE. Quantifying the Incremental and Aggregate Cost of Missed Workdays in Adults with Diabetes. *J Gen Intern Med* 2015 Dec;30(12):1773-1779 [[FREE Full text](#)] [doi: [10.1007/s11606-015-3338-y](#)] [Medline: [25986134](#)]

5. Schofield D, Cunich M, Kelly S, Passey ME, Shrestha R, Callander E, et al. The impact of diabetes on the labour force participation, savings and retirement income of workers aged 45-64 years in Australia. *PLoS One* 2015;10(2):e0116860 [FREE Full text] [doi: [10.1371/journal.pone.0116860](https://doi.org/10.1371/journal.pone.0116860)] [Medline: [25706941](https://pubmed.ncbi.nlm.nih.gov/25706941/)]
6. Schofield DJ, Cunich M, Shrestha RN, Callander EJ, Passey ME, Kelly SJ, et al. The impact of diabetes on the labour force participation and income poverty of workers aged 45-64 years in Australia. *PLoS One* 2014;9(2):e89360 [FREE Full text] [doi: [10.1371/journal.pone.0089360](https://doi.org/10.1371/journal.pone.0089360)] [Medline: [24586716](https://pubmed.ncbi.nlm.nih.gov/24586716/)]
7. Chan JCN, Gregg EW, Sargent J, Horton R. Reducing global diabetes burden by implementing solutions and identifying gaps: a Lancet Commission. *Lancet* 2016 Apr 09;387(10027):1494-1495. [doi: [10.1016/S0140-6736\(16\)30165-9](https://doi.org/10.1016/S0140-6736(16)30165-9)] [Medline: [27061676](https://pubmed.ncbi.nlm.nih.gov/27061676/)]
8. Panagioti M, Richardson G, Small N, Murray E, Rogers A, Kennedy A, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. *BMC Health Serv Res* 2014;14:356 [FREE Full text] [doi: [10.1186/1472-6963-14-356](https://doi.org/10.1186/1472-6963-14-356)] [Medline: [25164529](https://pubmed.ncbi.nlm.nih.gov/25164529/)]
9. Nicolucci A, Kovacs BK, Holt RIG, Comaschi M, Hermanns N, Ishii H, DAWN2 Study Group. Diabetes Attitudes, Wishes and Needs second study (DAWN2™): cross-national benchmarking of diabetes-related psychosocial outcomes for people with diabetes. *Diabet Med* 2013 Jul;30(7):767-777. [doi: [10.1111/dme.12245](https://doi.org/10.1111/dme.12245)] [Medline: [23711019](https://pubmed.ncbi.nlm.nih.gov/23711019/)]
10. Horgan G, Davies M, Findlay-White F, Chaney D, Coates V. Reasons why patients referred to diabetes education programmes choose not to attend: a systematic review. *Diabet Med* 2017 Dec;34(1):14-26. [doi: [10.1111/dme.13120](https://doi.org/10.1111/dme.13120)] [Medline: [26996982](https://pubmed.ncbi.nlm.nih.gov/26996982/)]
11. Corbin JM, Strauss AL. *Unending Work and Care: Managing Chronic Illness at Home*. San Francisco: Jossey-Bass Publishers; 1988:1-358.
12. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and Undertaking a Health Economics Study of Digital Health Interventions. *Am J Prev Med* 2016 Nov;51(5):852-860. [doi: [10.1016/j.amepre.2016.05.007](https://doi.org/10.1016/j.amepre.2016.05.007)] [Medline: [27745685](https://pubmed.ncbi.nlm.nih.gov/27745685/)]
13. Schechter CB, Walker EA, Ortega FM, Chamany S, Silver LD. Costs and effects of a telephonic diabetes self-management support intervention using health educators. *J Diabetes Complications* 2016 Mar;30(2):300-305 [FREE Full text] [doi: [10.1016/j.jdiacomp.2015.11.017](https://doi.org/10.1016/j.jdiacomp.2015.11.017)] [Medline: [26750743](https://pubmed.ncbi.nlm.nih.gov/26750743/)]
14. Murray E, Dack C, Barnard M, Farmer A, Li J, Michie S, et al. HeLP-Diabetes: randomised controlled trial protocol. *BMC Health Serv Res* 2015 Dec 29;15:578 [FREE Full text] [doi: [10.1186/s12913-015-1246-9](https://doi.org/10.1186/s12913-015-1246-9)] [Medline: [26715038](https://pubmed.ncbi.nlm.nih.gov/26715038/)]
15. National Institute for HealthCare Excellence (NICE): Guide to the methods of technology appraisal. 2013. 2013 URL: <https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781> [accessed 2017-10-23] [WebCite Cache ID 6uQWXTaES]
16. Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. *BMJ Open* 2017 Sep 27;7(9):e016009 [FREE Full text] [doi: [10.1136/bmjopen-2017-016009](https://doi.org/10.1136/bmjopen-2017-016009)] [Medline: [28954789](https://pubmed.ncbi.nlm.nih.gov/28954789/)]
17. Pal K, Dack C, Ross J, Murray E. Integrating theory, qualitative data and participatory design to develop HeLP-Diabetes: an internet self-management intervention for people with type 2 diabetes. 2013 May 18 Presented at: International Society for Research on Internet Interventions 6th Scientific Meeting; 2013 May 16-18; Chicago, Illinois, USA.
18. Murray E, Ross J, Pal K, Li J, Dack C, Stevenson F, et al. Development, evaluation (including randomised controlled trial and economic evaluation) and implementation of a web-based self-management programme for people with type 2 diabetes (HeLP-Diabetes). In: NIHR Journals Library. London, UK: NIHR Journals Library; 2017.
19. Polonsky WH, Anderson BJ, Lohrer PA, Welch G, Jacobson AM, Aponte JE, et al. Assessment of diabetes-related distress. *Diabetes Care* 1995 Jun;18(6):754-760. [Medline: [7555499](https://pubmed.ncbi.nlm.nih.gov/7555499/)]
20. Knapp M, Beecham J. Costing mental health services. *Psychol Med* 1990 Nov;20(4):893-908. [Medline: [2126630](https://pubmed.ncbi.nlm.nih.gov/2126630/)]
21. Kind P, Hardman G, Macran S. York: Centre for Health Economics, University of York. 1999. UK Population Norms for EQ-5D URL: <https://www.york.ac.uk/che/pdf/DPI72.pdf> [accessed 2018-03-06] [WebCite Cache ID 6xiREZ6gC]
22. Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. *Health Econ* 2004 Dec;13(12):1203-1210. [doi: [10.1002/hec.901](https://doi.org/10.1002/hec.901)] [Medline: [15386669](https://pubmed.ncbi.nlm.nih.gov/15386669/)]
23. National Institute for Health Rearch. INVOLVE URL: <http://www.invo.org.uk/> [accessed 2017-10-04] [WebCite Cache ID 6txmZsGhI]
24. University College of London. UCL Non-clinical grade structure with spinal points - 2013/14 2014 URL: http://www.ucl.ac.uk/hr/salary_scales/final_grades_2013.php [accessed 2017-10-23] [WebCite Cache ID 6uQWymGIX]
25. Curtis L. Unit Costs of Health and Social Care. Kent: The Univeristy of Kent, Personal Social Services Research Unit (PSSRU) UoK; 2014. URL: <https://www.pssru.ac.uk/pub/uc/uc2014/full-with-covers.pdf> [accessed 2018-03-06] [WebCite Cache ID 6xiRhG7IJ]
26. National Audit Office. Out-of-hours GP services in England. London, UK: Press Office, National Audit Office, England; 2014. URL: <https://www.nao.org.uk/wp-content/uploads/2014/09/Out-of-hours-GP-services-in-England1.pdf> [accessed 2018-03-06] [WebCite Cache ID 6xiS0kZ7t]
27. Department of Health and Social Care. Reference costs 2013-14. 2014. URL: <https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014> [accessed 2018-03-06] [WebCite Cache ID 6xiSDgIDV]

28. Prescribing & Medicines Team, Health and Social Care Information Centre. Prescription Cost Analysis England 2014.: Health and Social Care Information Centre; 2015 Apr 18. URL: <https://digital.nhs.uk/catalogue/PUB17274> [accessed 2017-10-23] [WebCite Cache ID 6uQisKrAw]
29. White IR, Kalaitzaki E, Thompson SG. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. *Stat Med* 2011 Nov 30;30(27):3192-3207 [FREE Full text] [doi: [10.1002/sim.4360](https://doi.org/10.1002/sim.4360)] [Medline: [21948462](https://pubmed.ncbi.nlm.nih.gov/21948462/)]
30. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001 Dec;10(8):779-787. [Medline: [11747057](https://pubmed.ncbi.nlm.nih.gov/11747057/)]
31. Health and Social Care Information Centre. National diabetes audit report 1 Care processes and treatment targets 2016-17. 2017 Nov 10. URL: <https://digital.nhs.uk/catalogue/PUB30142> [accessed 2018-02-15] [WebCite Cache ID 6xFmEWJzp]
32. Kennedy A, Reeves D, Bower P, Lee V, Middleton E, Richardson G, et al. The effectiveness and cost effectiveness of a national lay-led self care support programme for patients with long-term conditions: a pragmatic randomised controlled trial. *J Epidemiol Community Health* 2007 Mar;61(3):254-261 [FREE Full text] [doi: [10.1136/jech.2006.053538](https://doi.org/10.1136/jech.2006.053538)] [Medline: [17325405](https://pubmed.ncbi.nlm.nih.gov/17325405/)]
33. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 2015 Nov 03;16:495 [FREE Full text] [doi: [10.1186/s13063-015-1023-4](https://doi.org/10.1186/s13063-015-1023-4)] [Medline: [26530985](https://pubmed.ncbi.nlm.nih.gov/26530985/)]
34. Office for National Statistics. Internet access - households and individuals. 2017. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017> [accessed 2018-02-16] [WebCite Cache ID 6xH6QD82m]
35. Bergmo TS. How to Measure Costs and Benefits of eHealth Interventions: An Overview of Methods and Frameworks. *J Med Internet Res* 2015;17(11):e254 [FREE Full text] [doi: [10.2196/jmir.4521](https://doi.org/10.2196/jmir.4521)] [Medline: [26552360](https://pubmed.ncbi.nlm.nih.gov/26552360/)]
36. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. *Cochrane Database Syst Rev* 2017 Dec 25;9:CD011479. [doi: [10.1002/14651858.CD011479.pub2](https://doi.org/10.1002/14651858.CD011479.pub2)] [Medline: [28944453](https://pubmed.ncbi.nlm.nih.gov/28944453/)]
37. Ophuis RH, Lokkerbol J, Heemskerk SCM, van Balkom AJLM, Hiligsmann M, Evers SMAA. Cost-effectiveness of interventions for treating anxiety disorders: A systematic review. *J Affect Disord* 2017 Mar 01;210:1-13. [doi: [10.1016/j.jad.2016.12.005](https://doi.org/10.1016/j.jad.2016.12.005)] [Medline: [27988373](https://pubmed.ncbi.nlm.nih.gov/27988373/)]
38. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. *Psychol Med* 2015 Dec;45(16):3357-3376. [doi: [10.1017/S0033291715001427](https://doi.org/10.1017/S0033291715001427)] [Medline: [26235445](https://pubmed.ncbi.nlm.nih.gov/26235445/)]
39. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: A systematic review. *J Affect Disord* 2018 Jan 01;225:733-755. [doi: [10.1016/j.jad.2017.07.018](https://doi.org/10.1016/j.jad.2017.07.018)] [Medline: [28922737](https://pubmed.ncbi.nlm.nih.gov/28922737/)]
40. Andersson E, Ljótsson B, Smit F, Paxling B, Hedman E, Lindefors N, et al. Cost-effectiveness of internet-based cognitive behavior therapy for irritable bowel syndrome: results from a randomized controlled trial. *BMC Public Health* 2011 Apr 07;11:215 [FREE Full text] [doi: [10.1186/1471-2458-11-215](https://doi.org/10.1186/1471-2458-11-215)] [Medline: [21473754](https://pubmed.ncbi.nlm.nih.gov/21473754/)]
41. Murphy SM, Campbell ANC, Ghitza UE, Kyle TL, Bailey GL, Nunes EV, et al. Cost-effectiveness of an internet-delivered treatment for substance abuse: Data from a multisite randomized controlled trial. *Drug Alcohol Depend* 2016 Apr 01;161:119-126 [FREE Full text] [doi: [10.1016/j.drugalcdep.2016.01.021](https://doi.org/10.1016/j.drugalcdep.2016.01.021)] [Medline: [26880594](https://pubmed.ncbi.nlm.nih.gov/26880594/)]
42. Little P, Stuart B, Hobbs FR, Kelly J, Smith ER, Bradbury KJ, et al. Randomised controlled trial and economic analysis of an internet-based weight management programme: POWeR+ (Positive Online Weight Reduction). *Health Technol Assess* 2017 Jan;21(4):1-62 [FREE Full text] [doi: [10.3310/hta21040](https://doi.org/10.3310/hta21040)] [Medline: [28122658](https://pubmed.ncbi.nlm.nih.gov/28122658/)]
43. Thiarth H, Ebert DD, Lehr D, Nobis S, Buntrock C, Berking M, et al. Internet-Based Cognitive Behavioral Therapy for Insomnia: A Health Economic Evaluation. *Sleep* 2016 Oct 01;39(10):1769-1778 [FREE Full text] [doi: [10.5665/sleep.6152](https://doi.org/10.5665/sleep.6152)] [Medline: [27450686](https://pubmed.ncbi.nlm.nih.gov/27450686/)]
44. De Bruin EJ, van Steensel FJA, Meijer AM. Cost-Effectiveness of Group and Internet Cognitive Behavioral Therapy for Insomnia in Adolescents: Results from a Randomized Controlled Trial. *Sleep* 2016 Dec 01;39(8):1571-1581 [FREE Full text] [doi: [10.5665/sleep.6024](https://doi.org/10.5665/sleep.6024)] [Medline: [27306272](https://pubmed.ncbi.nlm.nih.gov/27306272/)]
45. König H, Bleibler F, Friederich H, Herpertz S, Lam T, Mayr A, et al. Economic evaluation of cognitive behavioral therapy and Internet-based guided self-help for binge-eating disorder. *Int J Eat Disord* 2018 Feb;51(2):155-164. [doi: [10.1002/eat.22822](https://doi.org/10.1002/eat.22822)] [Medline: [29345848](https://pubmed.ncbi.nlm.nih.gov/29345848/)]
46. Aardoom JJ, Dingemans AE, van Ginkel JR, Spinhoven P, Van Furth EF, Van den Akker-van Marle ME. Cost-utility of an internet-based intervention with or without therapist support in comparison with a waiting list for individuals with eating disorder symptoms: a randomized controlled trial. *Int J Eat Disord* 2016 Dec;49(12):1068-1076. [doi: [10.1002/eat.22587](https://doi.org/10.1002/eat.22587)] [Medline: [27441418](https://pubmed.ncbi.nlm.nih.gov/27441418/)]
47. Bouwsma EVA, Bosmans JE, van Dongen JM, Brölmann HAM, Anema JR, Huirne JAF. Cost-effectiveness of an internet-based perioperative care programme to enhance postoperative recovery in gynaecological patients: economic

- evaluation alongside a stepped-wedge cluster-randomised trial. *BMJ Open* 2018 Dec 21;8(1):e017782 [FREE Full text] [doi: [10.1136/bmjopen-2017-017782](https://doi.org/10.1136/bmjopen-2017-017782)] [Medline: [29358423](https://pubmed.ncbi.nlm.nih.gov/29358423/)]
48. Information Services Division Scotland. Prevalence reported from QOF registers (practices with any contract type). 2016 Jun 28. URL: https://www.isdscotland.org/Health-Topics/General-Practice/Publications/2016-10-11/QOF_Scot_201516_Boards_all_prevalence.xls [accessed 2018-05-03] [WebCite Cache ID 6z8WRFNCs]
 49. National Health Service Digital. Recorded disease prevalence, achievements and exceptions. 2016 Oct 27. URL: <http://webarchive.nationalarchives.gov.uk/20180328130852tf/http://content.digital.nhs.uk/catalogue/PUB22266/> [accessed 2018-05-03] [WebCite Cache ID 6z8Wd7Ccl]
 50. Welsh Government. Results for Wales and for each local health board, by year, disease register, and domain level achievement, 2015-16. 2016 Oct 19. URL: <https://gov.wales/docs/statistics/2016/161019-general-medical-services-contract-quality-outcomes-framework-2015-16-en.pdf> [accessed 2018-05-03] [WebCite Cache ID 6z8WjQBdG]

Abbreviations

CCA: complete case analysis

EQ-5D-3L: a standardized instrument developed by the EuroQoL Group that consisted of 5 domains with 3 levels each

GP: General Practitioner

HeLP-Diabetes: healthy living for people with type 2 diabetes

ICER: incremental cost-effectiveness ratio

NHS: National Health Services

NICE: National Institute of Health and Care Excellence

NIHR: National Institute of Health Research

PAID: problem areas in diabetes

PSS: Personal Social Services

PSSRU: Personal and Social Services Research Unit

QALY: quality-adjusted life year

SMS: short message service

T2DM: type 2 diabetes mellitus

UK: United Kingdom

WTP: willingness-to-pay

Edited by G Eysenbach; submitted 25.11.17; peer-reviewed by C Buntrock, I Swinkels; comments to author 01.02.18; revised version received 06.03.18; accepted 03.04.18; published 08.06.18.

Please cite as:

Li J, Parrott S, Sweeting M, Farmer A, Ross J, Dack C, Pal K, Yardley L, Barnard M, Hudda M, Alkhalidi G, Murray E. Cost-Effectiveness of Facilitated Access to a Self-Management Website, Compared to Usual Care, for Patients With Type 2 Diabetes (HeLP-Diabetes): Randomized Controlled Trial

J Med Internet Res 2018;20(6):e201

URL: <http://www.jmir.org/2018/6/e201/>

doi: [10.2196/jmir.9256](https://doi.org/10.2196/jmir.9256)

PMID: [29884608](https://pubmed.ncbi.nlm.nih.gov/29884608/)

©Jinshuo Li, Steve Parrott, Michael Sweeting, Andrew Farmer, Jamie Ross, Charlotte Dack, Kingshuk Pal, Lucy Yardley, Maria Barnard, Mohammed Hudda, Ghadah Alkhalidi, Elizabeth Murray. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 08.06.2018. 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.

Original Paper

Evaluation of a Diabetes Self-Management Program: Claims Analysis on Comorbid Illnesses, Health Care Utilization, and Cost

Ralph M Turner¹, PhD; Qinli Ma¹, PhD; Kate Lorig², DrPH; Jay Greenberg³, ScD; Andrea R DeVries¹, PhD

¹HealthCore, Inc, Wilmington, DE, United States

²Stanford Patient Education Research Center, Palo Alto, CA, United States

³National Council on Aging Services, Arlington, VA, United States

Corresponding Author:

Qinli Ma, PhD

HealthCore, Inc

123 Justison Street

Wilmington, DE, 19801

United States

Phone: 1 302 230 2000

Fax: 1 302 230 2020

Email: qma@healthcore.com

Abstract

Background: An estimated 30.3 million Americans have diabetes mellitus. The US Department of Health and Human Services created national objectives via its Healthy People 2020 initiative to improve the quality of life for people who either have or are at risk for diabetes mellitus, and hence, lower the personal and national economic burden of this debilitating chronic disease. Diabetes self-management education interventions are a primary focus of this initiative.

Objective: The aim of this study was to evaluate the impact of the Better Choices Better Health Diabetes (BCBH-D) self-management program on comorbid illness related to diabetes mellitus, health care utilization, and cost.

Methods: A propensity score matched two-group, pre-post design was used for this study. Retrospective administrative medical and pharmacy claims data from the HealthCore Integrated Research Environment were used for outcome variables. The intervention cohort included diabetes mellitus patients who were recruited to a diabetes self-management program. Control cohort subjects were identified from the HealthCore Integrated Research Environment by at least two diabetes-associated claims (International Classification of Diseases-Ninth Revision, ICD-9 250.xx) within 2 years before the program launch date (October 1, 2011-September 30, 2013) but did not participate in BCBH-D. Controls were matched to cases in a 3:1 propensity score match. Outcome measures included pre- and postintervention all-cause and diabetes-related utilization and costs. Cost outcomes are reported as least squares means. Repeated measures analyses (generalized estimating equation approach) were conducted for utilization, comorbid conditions, and costs.

Results: The program participants who were identified in HealthCore Integrated Research Environment claims (N=558) were matched to a control cohort of 1669 patients. Following the intervention, the self-management cohort experienced significant reductions for diabetes mellitus-associated comorbid conditions, with the postintervention disease burden being significantly lower (mean 1.6 [SD 1.6]) compared with the control cohort (mean 2.1 [SD 1.7]; $P=.001$). Postintervention all-cause utilization was decreased in the intervention cohort compared with controls with $-40/1000$ emergency department visits vs $+70/1000$; $P=.004$ and -5780 outpatient visits per 1000 vs $-290/1000$; $P=.001$. Unadjusted total all-cause medical cost was decreased by US \$2207 in the intervention cohort compared with a US \$338 decrease in the controls; $P=.001$. After adjustment for other variables through structural equation analysis, the direct effect of the BCBH-D was $-\text{US } \$815$ ($P=.049$).

Conclusions: Patients in the BCBH-D program experienced reduced all-cause health care utilization and costs. Direct cost savings were US \$815. Although encouraging, given the complexity of the patient population, further study is needed to cross-validate the results.

(*J Med Internet Res* 2018;20(6):e207) doi:[10.2196/jmir.9225](https://doi.org/10.2196/jmir.9225)

KEYWORDS

diabetes mellitus; patient education; health care utilization; cost

Introduction

Background

Current national estimates indicate that 9.4% of Americans (30.3 million people) have diabetes mellitus (DM) [1-3]. Annual (2012) DM-related expenditures are estimated at US \$245 billion, with US \$176 billion in direct medical costs and US \$69 billion in indirect costs attributable to disabilities, work absences, and premature death, among others [1-3]. Looking beyond the economic issues, having DM as well as common comorbid conditions such as hypertension and depression negatively impact quality of life (QoL), decrease functioning and ability to manage self-care, and raise mortality risk [4-11]. In response, the US Department of Health and Human Services created national objectives via its Healthy People 2020 initiative to help improve the QoL of people who either have or at risk for DM, and hence, lower the personal and national economic burden [12].

Diabetes Self-Management Education (DSME) interventions, which focus on guiding DM management and promoting healthy behaviors and lifestyle changes, are a primary part of this initiative. Many such interventions have been reported. Pal and colleagues completed a meta-analysis of 16 computer-based interventions (delivered through clinics, the internet, and mobile phones) and found small effects on glycemic control and no effects on depression or QoL [13]. Sherifali et al conducted an internet-based diabetes lifestyle management program and found improvements in body mass index, body fat percentage, and activity levels in participants [14]. A systematic review of nine internet-based diabetes lifestyle modifications demonstrated improvements in diet or activity level (2 studies) and improvements in glycemic control (2 studies) [15]. A meta-analysis of 13 studies using community health workers as interventionist found a modest reduction in glycated hemoglobin (HbA_{1c}) compared with usual care [16]. A more recent meta-analysis for group-based intervention involving 47 studies concluded that group-based programs showed better outcomes when taught by one or more professionals with or without a peer when compared with peer-led education [17]. It should be noted that these studies illustrate different aspects of the current intervention—computer-based, internet-based, peer-led and theory-based, all of which are aspects of the intervention discussed in this paper.

There have also been several diabetes education cost-effectiveness studies. A recently completed meta-analysis of 8 studies found that 4 were based on reduction of clinical measures and 4 on quality-adjusted life years (QALYs) [18]. A program in South Africa that lowered blood pressure was reported as cost-effective based on QALYs gained [19]. Similar findings were found with a nurse-delivered telecoaching program in Belgium [20]. Evaluations of diabetes counseling and education programs have produced mixed economic results. Although most have shown a reduction in overall health care costs, not all programs have produced positive changes [21,22]. Telecoaching programs, including Web-based and phone-based programs, show particular promise in improving diabetes care and costs [23]. In a systematic review of interactive computer

technology interventions to improve diabetes care, Jackson et al found an overall positive impact on diabetes care measures and health care utilization [24]. More recently, a quasi-experimental study by Nundy et al within an employer health plan found that mobile phone-based diabetes education was able to produce positive behavioral changes, improve glycemic control, and lower overall medical costs [25].

The intervention reported on in this study, Better Choices Better Health Diabetes (BCBH-D) program, originally developed at Stanford University, is an intensive DSME series of 2.5-hour sessions over 6 consecutive weeks that has demonstrated effectiveness [26]. The program was first evaluated as small groups in community settings near patients' homes. More recently, online workshops have been developed and offered to persons with diabetes who preferred online communication because they lived in rural areas, wanted anonymity, were homebound, or had busy schedules [27]. The online program contains all the elements of the small group method, except participants log in and do the work from their personal computers. Participants complete exercises, read posted materials, and interact virtually with others in their group. Two related studies evaluating the small group and internet-based BCBH-D program, led by a consortium of the Stanford Patient Education Research Center, National Council on Aging (NCOA), and Anthem Health Plans, reported modest benefits among the program participants based on 6- and 12-month follow-up periods [28,29]. The 6-month follow-up study reported statistically significant improvements in 13 of 14 outcome measures, including HbA_{1c} and health behaviors (eg, communication with physicians, stretching or strengthening exercise, medication adherence, and frequency of eye, foot, cholesterol, and kidney exams) [28]. In the 12-month follow-up program, more than two-thirds (69.7%, 597/857) of the baseline study population experienced statistically significant improvements in 13 of the 15 prespecified outcome measures. The researchers noted that the improvements at 6 months were maintained and augmented during the 12-month period [29]. However, to date, there has been no research that has evaluated the impact of the BCBH-D self-management program on health care utilization and costs.

Objective

This study was designed as an administrative claims-based observational evaluation of the BCBH-D self-management program first offered in October 2013 to Anthem health plan enrollees in a real-world setting [28,29]. The objective of the present research was to evaluate (1) The program's impact on comorbid illness related to DM, (2) Health care utilization, and (3) Health care costs within 12 months after program enrollment, compared with a propensity score matched control cohort of DM patients who received usual care but did not participate in the BCBH-D program. In addition, the study sought to better understand how baseline comorbid illness burden, age, gender, and prior health care costs influenced or mediated the 12-month changes in health care costs and if these differed by intervention and control groups.

Methods

Design

A propensity score matched two-group, pre-post observational design was used for this study. Retrospective administrative medical and pharmacy claims data from the HealthCore Integrated Research Environment (HIRE) were used for outcome variables. Administrative claims data comprehensively contains patient's use of medical and pharmacy services, including hospitalization; emergency room; services occurring in an outpatient setting, such as office visits and laboratory test; and prescription fills. Researchers accessed a limited dataset in a Health Insurance Portability and Accountability Act compliant manner. Central Institutional Review Board (IRB) approval was obtained before the initiation of the study.

Participants

The intervention cohort included DM patients who were recruited from October 2013 to October 2014 to participate in the program (reported on by Lorig et al), attended at least one session, and were identified in the HIRE claims database [29]. For inclusion in the BCBH-D, the participants were required to speak English and provide IRB-approved informed consent and were not permitted to have previously participated in a self-management program developed at Stanford. Methods and results of the active intervention have been previously reported [27-29]. For inclusion in this study, patients had to be found in the health care claims data with enrollment in the health plan for 1 year before and 1 year following the intervention. The index date for intervention cohort members was defined as the program enrollment date. Control cohort subjects were identified from the HIRE by at least two claims associated with diabetes (International Classification of Diseases-Ninth Revision, ICD-9 250.xx) within 2 years before the program launch date (October 1, 2011-September 30, 2013). Controls were 3:1 propensity score matched based on age, gender, health plan type, residence region, Metropolitan Statistical Areas (MSA or non-MSA), Deyo-Charlson Comorbidity Index (DCI) score (range: 0-25), comorbid illnesses, health care utilization, and total medical cost at 1-year baseline [30]. The index date was defined for controls as a randomly selected date during the same period as the BCBH-D program recruitment period (October 2013-October 2014). As the data for the control cohort was a limited dataset, a waiver was sought and obtained from the IRB, and informed consent was not required.

Patients in both cohorts were required to be members of an Anthem-affiliated health plan, aged 18 years or older, with type II diabetes, and to have continuous medical eligibility for 1 year before and after the index date. Exclusion criteria included major treatment for cancer (radiation, chemotherapy, or surgery) or pregnancy.

Measures

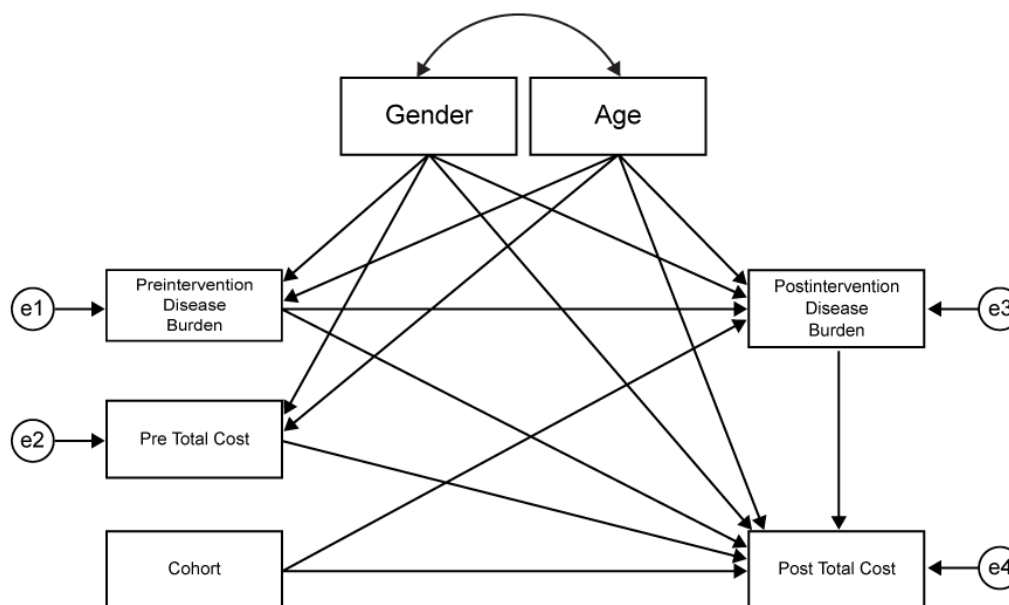
Demographic measures included the preintervention values of gender, age, and geographical region of residence. DCI score was calculated to measure overall preintervention illness burden.

Other clinical measures included a series of dichotomous variables documenting the absence or presence of 15 DM-associated comorbid illnesses, including hypertension, hyperlipidemia, obesity, chronic obstructive pulmonary disease, renal disease, depression, metabolic syndrome, ischemic heart disease, coronary heart disease, osteoporosis, osteoarthritis, lower back pain, peripheral vascular disease, musculoskeletal disorder, and sleep apnea, based on ≥ 2 ICD-9 codes for each (recorded in the outpatient; inpatient; or emergency department, ED setting) [30]. The 15 comorbid conditions were also measured during the follow-up year. In addition, an overall composite index that aggregated these 15 diseases at both pre- and postintervention periods was calculated for each patient to capture DM-associated comorbid illnesses burden.

Outcome measures included pre- and postintervention all-cause and diabetes-specific utilization and costs recorded in the HIRE claims database. Diabetes specific utilization was defined as hospitalizations and ED visits with primary diagnosis of diabetes and outpatient services (including but not limited to office visits, imaging, laboratory tests, and procedures) with any diabetes diagnosis on the claim. All-cause utilization was defined as any claims-based health care utilization inclusive of diabetes and any other diagnosis on the claim. All-cause and diabetes specific utilization measures for the 12-month pre- and postintervention periods encompassed hospitalizations, ED visits, and outpatient services, reported as visits per 1000 patients. All-cause and diabetes-specific costs were measured for overall medical services and for each service type, including hospitalizations, ED visits, and outpatient services. All-cause pharmacy costs were assessed for the 12-month pre- and postintervention periods.

Statistical Analysis

The demographic and clinical characteristics at baseline were reported by percentage or mean for the Intervention and matched control cohorts. Chi-square tests and *t* tests were conducted to examine the differences between two cohorts. Repeated-measures analyses were conducted using the generalized estimating equation (GEE) approach, using a binomial distribution and a logit link for presence of comorbid conditions for objective 1, a negative binomial distribution and a logit link for utilization for objective 2, and a gamma distribution with log link for costs for objective 3. All costs are reported as least squares values. The GEE analysis allowed for two main independent comparisons. First, a statistical test assessed the overall cohort \times time interaction that determined if the cohorts had different slopes (ie, different changes in outcome measures) between pre- and postintervention periods. Second, the test for a main effect among cohorts was divided into a comparison between the two cohorts at preintervention period and a comparison between the cohorts for differences at postintervention period. To better understand the cohort difference in the 12-month changes of health care costs controlling for the influence of baseline comorbid illness burden, age, gender, and prior health care costs groups, a maximum likelihood structural equation model (SEM) analysis was conducted.

Figure 1. Starting model for structural equation model (SEM) analysis.

Starting Model
Model for BCBH-D vs Matched Controls for Total Health Care Costs

SEM was used because it permits a mediation analysis to identify and explain the process that accounts for the relationship between the treatment effect and total medical cost via the inclusion of a third variable, a mediator variable, which in this study is the composite index of DM-associated comorbid illnesses. This permits a test that changes in the composite index of DM-associated comorbid illnesses mediates a portion of the total postintervention medical costs.

Although SEM allows for the use of latent or unobserved variables, this study used only manifest measures. The starting model is shown in Figure 1. It was assumed that in addition to a cohort effect, patient gender, age, comorbid illness burden, and preintervention, total medical cost might account for postintervention total cost. As propensity score matching was used in this study, it was assumed that the cohort independent variable was uncorrelated with the sex, preintervention total costs, and preintervention comorbid illness burden. However, as treatment might impact on postintervention comorbid illness burden, this variable was assessed for its mediation effects.

Because the chi-square test of absolute model fit is sensitive to sample size and non-normality in the underlying distribution of the input variables, we used several descriptive fit statistics to assess the overall fit a model to the data, including the Comparative Fit Index (CFI), the Normed Fit Index (NFI), the Akaike Information Criterion (AIC), and the root mean square error (RMSE), indices to aid in judging the quality of model fit [31-33]. Statistical significance was defined as $P < .05$. Analyses were conducted using SAS Institute's SAS Software, Version 9.3, except for the SEM, which was conducted in Stata, Version Release 14 by StataCorp [34,35].

Results

Identification of Clinical Study Patients in Health Care Claims Data

There were 1229 small-group and online participants in the Lorig et al (2016) study [29]. Within this population, 8 (0.65%) had no health insurance, 183 (14.89%) had Medicare, 2 (0.16%) were enrolled in Medicaid, 1 (0.08%) had Supplemental Security Income, 11 (0.90%) had Veterans benefits, 1014 (82.51%) had private insurance, and the insurance status of 10 study participants was unknown. We were able to identify 558 (55.03%, 558/1014) of privately insured) in the HIRE claims database who met the study criteria and also were covered for a full year before and after the participation date. The inability to identify the rest was probably because while participants were covered by an Anthem plan when they entered the study, they were not covered for a full year before and after the index date. In other cases, participants were enrolled in a Medicare supplemental plan where full health care utilization data were not available.

We compared the characteristics of the patients identified in claims ($N=558$) with the study participants we could not identify ($N=671$) to assess the potential for bias in our sample. The results are presented in Table 1. The majority of claims-identified patients (528/558, 94.6%) participated in the online workshop format and 28 (5.0%, 28/558) participated in the small-group format, whereas, 229 (18.6%, 229/1229) of all BCBH-D study participants chose the small-group program in the original study [29]. Claims-identified patients had a larger percentage of females, fell more heavily into the 41 to 64 years age categories, had a higher white than black racial composition, and self-reported a smaller arthritis comorbid illness burden than the patients that were not identified in claims.

Table 1. Characteristics of claims-identified study participants vs nonidentified study participants.

Characteristic	BCBH-D ^a found in claims	BCBH-D not found in claims	<i>P</i> value	<i>R</i> ²
Total patients, n	558	671		
Female, n (%)	359 (64.5)	205 (30.5)	.001	.10
Age categories (years), n (%)			.001	.04
21-30	3 (0.5)	35 (0.6)		
31-40	25 (4.5)	95 (5.6)		
41-50	133 (23.8)	336 (15.2)	<.05 ^b	
51-64	335 (60.4)	154 (53.8)	<.05 ^b	
≥65	61 (10.9)	624 (24.7)	<.05 ^b	
Race, n (%)			.002	.02
American Indian or Alaska Native	4 (0.7)	7 (1.0)		
Asian (includes Indian)	22 (0.04)	22 (3.3)		
Black or African-American	51 (9.1)	115 (17.1)	<.05 ^b	
Native Hawaiian or Pacific Islander	1 (0.2)	1 (0.1)		
White (includes Hispanic or Latino)	466 (83.5)	504 (75.0)	<.05 ^b	
Two or more races	11 (2.0)	20 (3.0)		
Declined	3 (0.5)	3 (0.4)		
Hispanic, n (%)	43 (7.7)	52 (7.7)	.99	.000
Married, n (%)	400 (71.7)	470 (69.9)	.42	.000
Years of education, mean (SD)	15.4 (2.7)	15.3 (2.8)	.52	.001
Study status, n (%)				
Completed post measures	545 (97.7)	655 (97.5)	.29	.05
Died	0 (0.0)	3 (0.4)		
Lost to follow-up	12 (2.2)	14 (2.1)		
Study treatment mode, n (%)			.001	.09
Online	528 (94.6)	473 (70.4)	<.05 ^b	
Small group	30 (5.4)	199 (29.6)	<.05 ^b	
Self-reported comorbidities, n (%)				
None	104 (18.6)	117 (17.4)	.66	.000
Hypertension	340 (60.9)	418 (62.2)	.68	.000
Asthma	51 (9.2)	54 (8.0)	.48	.004
Arthritis	96 (17.2)	164 (24.4)	.002	.008
Coronary heart disease	56 (10.1)	73 (10.9)	.65	.000
Chronic obstructive pulmonary disease	19 (3.4)	18 (2.7)	.45	.000
Cancer	27 (4.8)	32 (4.8)	.94	.000
Renal disease	16 (2.9)	28 (4.2)	.22	.001
Depression	103 (18.5)	127 (18.9)	.86	.000
Mental health problems	68 (12.2)	77 (11.5)	.69	.000
HbA_{1c}^c	N=432 (with valid test result)	N=523 (with valid test result)		
Pre, mean (SD)	8.29 (1.62)	8.12 (1.48)	.10	.00
Post, mean (SD)	7.67 (1.32)	7.54 (1.41)	.28	.00

Characteristic	BCBH-D ^a found in claims	BCBH-D not found in claims	<i>P</i> value	<i>R</i> ²
Patient Health Questionnaire depression score	N=555	N=670		
Pre, mean (SD)	5.85 (4.75)	6.19 (5.34)	.24	.001
Post, mean (SD)	4.83 (4.55)	4.95 (5.15)	.72	.000

^aBCBH-D: Better Choices Better Health Diabetes.

^bMultiple comparison post-hoc tests following a statistically significant omnibus chi-square test indicate that there are statistically significant differences between the cohorts at the $P<.05$ level (one-tailed).

^cHbA_{1c}: glycated hemoglobin.

However, claims-identified patients did not differ from the rest of the study population regarding depression, HbA_{1c} levels, other comorbid disease burden, years of education, marital status, completion of all study measures, or loss to follow-up. Although there is potential for some biases between the full Lorig et al (2016) study sample and the claims sample based on sex, age, and format of treatment, on the whole, the claims-identified cohort represents the original study population [29].

Propensity Score Matching Results

The 558 claims-identified patients were 1:3 propensity score matched to a pool of 685,412 patients diagnosed with type II DM and who met the eligibility criteria. The final matched control cohort consisted of 1669 patients. Baseline demographics and health care utilization data for postmatch patients are presented in Table 2. The intervention and matched control cohort had similar demographic characteristics with average age of 55 years, around 65.0% (360/558 and 1097/1669) female, and mostly residing in South (31.0%, 176/558 and 514/1669) and Midwest (42.0%, 231/558 and 697/1669) regions. Both cohorts had similar DCI scores (mean 1.6 [SD 1.2] vs mean 1.6 [SD 1.2]; $P=.47$). As expected, given the propensity score matching process, no significant differences were seen in the prevalence of comorbid illnesses, annual health care utilization, or annual health care costs across the two cohorts at baseline.

Impact on Comorbid Illness

As can be seen in Table 3, following the intervention, the intervention cohort was associated with significant reductions in medical claims associated with the following comorbid illnesses: hypertension (−50/558, −9.0% vs +10/1669, +0.54%; $P=.001$), hyperlipidemia (−39/558, −7.0% vs −23/1669, −1.44%; $P=.04$), and depression (−16/558, −2.9% vs +12/1669, +0.72%; $P=.01$) compared with the matched control cohort. In addition, the matched control cohort showed significant increases in claims for health care services associated with renal disease (+32/1669, +1.98% vs +1/558, +0.2%; $P=.006$), rheumatoid arthritis or osteoarthritis (+52/1669, +3.12% vs −12/558, −2.2%; $P=.001$), and musculoskeletal disorders (3.72% vs −0.5%, $P=.04$). Thus, there was a decrease in medical claims associated with key comorbid conditions within the intervention cohort from pre- to postintervention period assessment and an increase in claims for conditions among the matched control cohort. An overall composite index that aggregated these 15 diseases for both the pre- and postintervention periods was calculated to capture these shifts in DM-associated comorbid illnesses (eg, an index of 2 means a patient had 2 out of these 15 diseases).

During the preintervention period, the matched control's disease burden (mean 2.0 [SD 1.6]) did not differ statistically from the intervention cohort's value (mean 2.0 [SD 1.6]; $P=.86$). However, at postintervention period, the matched control's disease burden (mean 2.1 [SD 1.7]) was significantly higher than the intervention cohort's (mean 1.6 [SD 1.6]; $P=.001$). In addition, the slope of the post- minus preintervention change in disease burden for the intervention cohort was −0.4 (SD 1.5) vs 0.1 (SD 1.4) for the matched control cohort ($P=.001$).

Health Care Utilization

Change of health care utilization results following the intervention are presented in Table 4. The intervention cohort had reduced utilization of all-cause ED visits and outpatient services compared with the matched control cohort. Although the cohorts showed no significant differences in utilization during the preintervention period, the intervention cohort experienced a significant reduction in all-cause ED visits (−40 per thousand) compared with an increase of 70/1000 for the matched control cohort. Likewise, the intervention cohort had a decrease of 30/1000 inpatient hospitalizations during the follow-up period, whereas the matched control cohort experienced an increase of 10/1000, although this difference did not reach statistical significance ($P=.10$). Finally, the intervention cohort's use of outpatient services decreased by −5780/1000 contacts during the follow-up compared with a decrease of −290/1000 for the matched control cohort ($P=.001$). For diabetes-related utilization, a statistically significant finding was obtained on outpatient services only.

Health Care Costs

Health care cost data are presented in Table 5. On the basis of data screening before the main analyses of the cost centers, we noted significant cost outliers in the inpatient and outpatient categories above four SDs of the mean during both the pre- and postintervention assessment periods. For inpatient services, there was one matched control and one intervention patient, each of whom were outliers at preintervention period. At postintervention period, there was one matched control, with inpatient costs above four SDs above the mean. For outpatient costs during the preintervention period, there were two matched control and one intervention cohort members who were above four SDs; at postintervention, there were two matched control and one intervention cohort members who fell into this category. We examined the claims associated with all patients, with inpatient and outpatient costs greater than three SDs above average and found that the most frequently associated primary diagnoses were chronic liver disease and cirrhosis, septicemia,

acute kidney failure, acute endocarditis, venous embolism and thrombosis, transient cerebral ischemia, care involving use of rehabilitation procedures, pneumonia, intervertebral disc disorders, disorders of intestine, and venipuncture. On the basis

of our review, we capped inpatient costs per year at US \$200,000 and outpatient costs at US \$150,000. On the whole, this reduced bias against the intervention cohort.

Table 2. Baseline demographic and utilization data.

Characteristics	Intervention cohort (N=558)	Matched control cohort (N=1669)	P value
Female, n (%)	360 (64.5)	1097 (65.73)	.60
Age (years), mean (SD)	55.29 (8.89)	54.97 (11.19)	.48
Residence region, n (%)			
Northeast	71 9 (12.7)	222 (13.30)	.98
South	176 (31.5)	514 (30.80)	
Midwest	231 (41.4)	697 (41.76)	
West	80 (14.3)	236 (14.14)	
Medicare advantage, n (%)	17 (3.0)	67 (4.01)	.30
Comorbidities			
DCI ^a , (mean, SD)	1.59 (1.19)	1.55 (1.16)	.47
Hypertension, n (%)	315 (56.5)	945 (56.62)	.94
Hyperlipidemia, n (%)	315 (56.5)	945 (56.62)	.94
Obesity, n (%)	62 (11.1)	173 (10.37)	.62
Chronic obstructive pulmonary disease, n (%)	26 (4.7)	80 (4.79)	.90
Renal disease, n (%)	13 (2.3)	26 (1.56)	.23
Depression, n (%)	58 (10.4)	162 (9.71)	.64
Metabolic syndrome, n (%)	1 (0.2)	4 (0.24)	>.99
Ischemic heart disease, n (%)	30 (5.4)	82 (4.91)	.66
Coronary heart disease, n (%)	3 (0.5)	8 (0.48)	>.99
Peripheral vascular disease, n (%)	11 (2.0)	36 (2.16)	.79
Osteoporosis, n (%)	7 (1.3)	22 (1.32)	.91
Rheumatoid arthritis or osteoarthritis, n (%)	50 (9.0)	152 (9.11)	.92
Low back pain, n (%)	69 (12.4)	223 (13.36)	.55
Musculoskeletal disorders (low back pain excluded), n (%)	88 (15.8)	269 (16.12)	.85
Sleep apnea, n (%)	88 (15.8)	248 (14.86)	.60
Health care utilization, mean (SD)			
Hospitalization	0.11 (0.42)	0.1 (0.45)	.60
ED ^b visit	0.19 (0.65)	0.19 (0.57)	.83
Office visit	7.85 (6.41)	7.9 (6.62)	.88
Lab test	6.14 (4.99)	6.19 (7.03)	.85
Health care cost (US \$), mean (SD)			
Total medical cost	7997 (18,045)	8551 (18,235)	.75

^aDCI: Deyo-Charlson Comorbidity Index.

^bED: emergency department.

Table 3. Analysis of common comorbid disorders.

Disease	Cohort ^a						Difference in slopes, %	P value ^b		
	BCBH-D ^c intervention ^a (N=558)			Matched controls ^a (N=1669)				Difference in slopes	Cohorts time 1	Cohorts time 2
	Pre	Post	Slope, %	Pre	Post	Slope, %				
Hypertension, n (%)	316 (56.6)	266 (47.6)	−9.0	945 (56.62)	955 (57.16)	0.54	−9.5	.001	.98	<.001
Hyperlipidemia, n (%)	316 (56.6)	277 (49.7)	−7.0	945 (56.62)	921 (55.18)	−1.44	−5.5	.04	.98	.03
Obesity, n (%)	62 (11.1)	67 (12.0)	0.9	174 (10.37)	164 (9.77)	−0.60	1.5	.40	.15	.62
Chronic obstructive pulmonary disease, n (%)	26 (4.7)	22 (3.9)	−0.7	80 (4.79)	73 (4.43)	−0.36	−0.4	.71	.90	.62
Renal disease, n (%)	13 (2.3)	14 (2.5)	0.2	27 (1.56)	58 (3.54)	1.98	−1.8	.01	.27	.20
Depression, n (%)	58 (10.4)	42 (7.5)	−2.9	162 (9.71)	174 (10.43)	0.72	−3.6	.01	.63	.03
Metabolic syndrome, n (%)	1 (0.2)	1 (0.2)	0.0	3 (0.24)	8 (0.84)	0.24	−0.2	.64	.78	.22
Ischemic heart disease, n (%)	30 (5.4)	27 (4.8)	0.6	82 (4.91)	103 (6.17)	1.00	−1.5	.25	.67	.22
Coronary heart disease, n (%)	26 (4.7)	22 (3.9)	−0.8	80 (4.79)	78 (4.67)	−0.36	−0.4	.42	.87	.28
Osteoporosis, n (%)	7 (1.3)	6 (1.1)	−0.2	22 (1.32)	42 (2.46)	1.00	−1.2	.17	.91	.02
Rheumatoid arthritis or osteoarthritis, n (%)	50 (9.0)	38 (6.8)	−2.2	152 (9.11)	204 (12.22)	3.12	−5.3	.001	.93	<.001
Lower back pain ^d , n (%)	69 (12.4)	67 (12.0)	−0.4	224 (13.36)	247 (14.80)	1.44	−1.8	.27	.55	.09
Peripheral vascular disease, n (%)	11 (2.0)	12 (2.2)	0.2	37 (2.16)	45 (2.70)	0.54	−0.4	.70	.79	.46
Musculoskeletal disorders ^d , n (%)	88 (15.8)	85 (15.3)	−0.5	269 (16.12)	330 (19.83)	3.72	−4.3	.04	.86	.01
Sleep apnea, n (%)	88 (15.8)	62 (11.1)	−4.7	249 (14.86)	219 (13.12)	1.74	−2.9%	.07	.60	.20
Total disease burden score, mean (SD)	2.0 (1.6)	1.6 (1.6)	−0.4	2.0 (1.6)	2.1 (1.7)	0.1	−0.3	.001	.86	.001

^aAll subjects have 364 days of pretreatment and follow-up time.

^bAnalysis was conducted using generalized estimating equation repeated measures model with a binomial distribution and a logit link. The matched control cohort is the reference category in all analyses.

^cBCBH-D: Better Choices Better Health Diabetes.

^dLow back pain is separated out from musculoskeletal disorders.

Table 4. Analysis of health care utilization.

Utilization	Cohort ^a						Difference in slopes	<i>P</i> value ^b		
	BCBH-D intervention (N=558)			Matched controls (N=1669)				Difference in slopes	Cohorts time 1	Cohorts time 2
	Pre	Post	Slope	Pre	Post	Slope				
All-cause medical utilization (visits per 1000 members per year)										
Inpatient	110	80	−30	100	110	10	−40	.10	.55	.15
Emergency de- partment	190	150	−40	190	260	70	−110	.004	.80	<.001
Outpatient	22,880	17,100	−5780	22,880	22,590	−290	−5490	<.001	.99	<.001
Diabetes-related medical utilization (visits per 1000 members per year)										
Inpatient	10	0	−10	10	0	−10	0	.57	.75	.71
Emergency de- partment	10	0	−10	10	10	0	−10	.27	.91	.30
Outpatient	6630	5200	−1430	6760	6530	−230	−1200	<.001	.68	<.001

^aAll subjects have 1 year of pretreatment time and 1 year of follow-up time.

^bAll analysis were conducted using generalized estimating equation repeated measures model with a logit link and negative binomial distribution for inpatient and emergency department visits and zero-inflated negative binomial distribution for outpatient services. The matched control cohort is the reference category in all analyses.

As Table 5 shows, there were no statistically significant differences in any of the all-cause cost categories during the preintervention period. However, during the postintervention period, the intervention cohort showed significantly lower costs than the matched control cohort for inpatient ($P=.01$), ED services ($P=.003$), and outpatient services ($P=.01$), as well as total all-cause medical costs ($P=.001$). Results for diabetes-specific medical costs presented a different picture from all-cause costs. Diabetes-specific Inpatient and ED cost were low, and no statistically significant findings were obtained between the two cohorts. The cohorts did not have significantly different total all-cause pharmacy costs during the preintervention period ($P=.34$). However, at postintervention period, the matched controls (US \$5675) showed higher pharmacy cost than the intervention cohort (US \$4264), $P=.001$. There were significant differences in the trajectories of pharmacy costs of the intervention cohort (−US \$141) and the matched controls (US \$936) over time; $P=.001$.

Adjusted Cost Savings

Other than the intervention program, the observed costs (in 2016-adjusted US currency) in Table 5 were influenced by patient's demographic and clinical factors. The maximum likelihood SEM analysis was used to estimate the adjusted cost saving of the program. The starting model for the SEM analysis is shown in Figure 1. This model provided excellent fit to the data ($\chi^2_5=2.5$, $P=.77$, CFI=0.997, NFI=0.996, AIC=62.52, RMSE=0.0001 (95% CI 0.00002-0.02). However, none of the regression coefficients from gender to other model variables were statistically significant: preintervention illness burden ($P=.19$), postintervention illness burden ($P=.69$), preintervention total cost ($P=.14$), and postintervention total cost ($P=.22$). Likewise, the regression coefficients from age to preintervention total cost ($P=.26$) and postintervention total cost ($P=.64$) were not significant. Age did show significant effects on

preintervention disease burden (0.03 [$P<.001$]) and postintervention disease burden (0.03 [$P<.001$]).

Therefore, gender was removed from the structural model, and age was limited to having effects on pre- and postintervention burden in the trimmed model shown in Figure 2. This final model continued to show excellent fit to the data ($\chi^2_6=3.9$, $P=.69$, CFI=0.995, NFI=0.996, AIC=45.91, RMSE=0.001 (95% CI 0.0001-0.02). The final model added one degree of freedom, which allowed us to test if there was a difference in the quality of model fit. A test of the change in the chi-square statistic showed the final model did not show inferior fit compared with the starting model ($\chi^2_1=1.3$, $P>.10$), and the comparison of the AIC values indicated the final model to be preferred over the starting model (starting model AIC=62.52 vs final model AIC=45.91).

As Figure 2 illustrates, each additional year of age increased preintervention burden by 0.03 and postintervention burden by 0.02 ($P=.001$ for both). Although age had no direct effects upon postintervention total health care costs, it did have total indirect effects through preintervention burden, postintervention disease burden and preintervention total costs, and postintervention disease burden of US \$180. Preintervention disease burden had a direct effect on postintervention disease burden of an increased 0.62 illness ($P=.001$), preintervention total costs of US \$3617, and postintervention total costs of US \$2092. The indirect effect of preintervention disease burden on postintervention total cost through preintervention total cost was US \$1197 and US \$2521 through postintervention burden. Thus, the total effects for preintervention disease burden were US \$5810. Preintervention total health care costs increased postintervention health care costs by US \$0.33 per every preintervention dollar spent during the preintervention period. There were no indirect effects of preintervention total health care costs, so the direct effect equaled the total effect, as illustrated in Figure 2.

Table 5. Analysis of health care costs.

Cost (US \$)	Cohort ^a						Difference in slopes	<i>P</i> value ^e	Cohorts time 1	Cohorts time 2
	BCBH-D ^b intervention, medical (N=558) ^c , pharmacy (N=330) ^d			Matched controls, medical (N=1669) ^c , pharmacy (N=990) ^d						
	Pre	Post	Slope (change in cost)	Pre	Post	Slope (change in cost)				
All-cause medical cost										
Inpatient	2274	1406	−868	2446	2490	44	−912	.03	.78	.01
Emergency de- partment	381	316	−64	434	502	69	−133	.16	.54	.003
Outpatient	5327	4040	−1288	5651	5201	−450	−838	.001	.50	.01
Total	7997	5789	−2207	8551	8213	−338	−1869	.001	.53	.001
Diabetes-related medical cost										
Inpatient	97	124	26	247	74	−172	198	.16	.19	.64
Emergency de- partment	13	6	−8	18	14	−4	−4	.75	.58	.20
Outpatient	1437	1249	−189	1943	1711	−232	44	.92	.007	.007
Total	1558	1383	−175	2223	1808	−415	240	.58	.005	.04
All-cause pharmacy cost										
Total	4405	4264	−141	4739	5675	936	−1078	.001	.34	.001

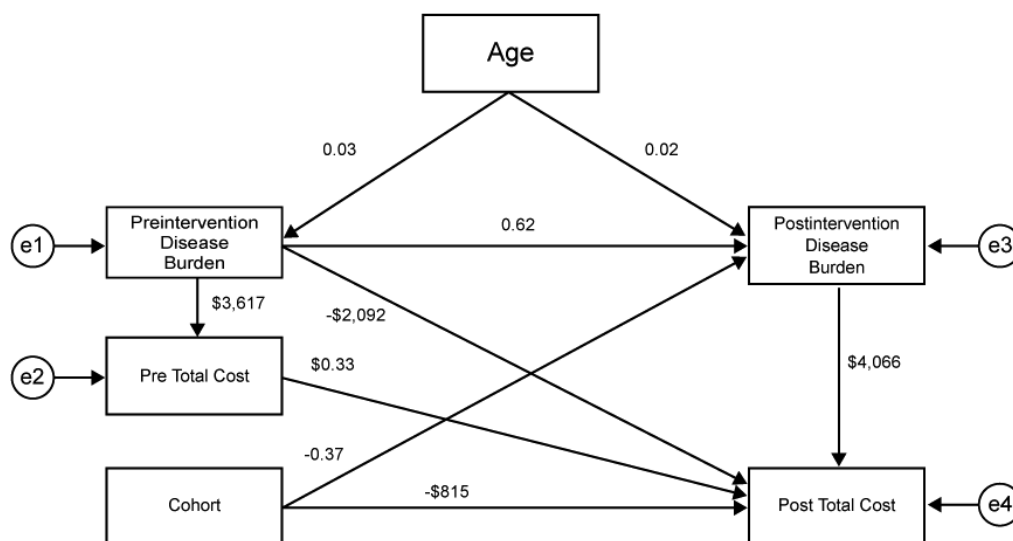
^aAll subjects have 1 year of pretreatment time and 1 year of follow-up observation time.

^bBCBH-D: Better Choices Better Health Diabetes.

^cN of subjects with medical coverage.

^dN of subjects with pharmacy coverage.

^eAll analyses were conducted using generalized estimating equation repeated measures model with a gamma distribution and a logit link. The matched control cohort is the reference category in all analyses.

Figure 2. Structural equation model (SEM) final model.**Model for BCBH-D vs Matched Controls for Total Health Care Costs**

Chi-square₆=3.9, P=.69

CFI=.995, NFI=.996, AIC=45.91, RMSE=.001 (95% CI 0.0001-0.021)

All path estimates are significant at P=.001, except for the Cohort effect where P=.049

The intervention had an indirect association of -0.37 with postintervention disease burden, which had a direct effect of increasing postintervention total health care costs by US \$4066. This translated into an indirect effect of $-\text{US } \$1504$ for the intervention cohort on postintervention total health care cost through postintervention disease burden. Consequently, the total effects associated with the intervention were $-\text{US } \$2220$. After all model adjustments, intervention had a direct effect of reducing postintervention total health care costs by $-\text{US } \$815$ ($P=.049$) compared with the matched control cohort as shown in Figure 2.

Discussion

Principal Findings

This study shows that beyond the clinical benefits seen in the previous Lorig et al study, participants in the peer-facilitated BCBH-D program experienced reduced all-cause health care utilization and medical costs relative to the matched control group during the year the intervention occurred [29]. The utilization and costs related directly to diabetes care were low to begin with and remained stable over the year of intervention. For diabetes specific care, as documented in the health care utilization and cost results, there was stability, and not worsening, across time. On the other hand, we observed large and significant reductions in all-cause utilization and costs for the intervention cohort for inpatient, ED, and outpatient services, as well as total all-cause medical and pharmacy costs. Intriguingly, there was reduced utilization in the intervention cohort for chronic conditions that often co-occur with DM. Specifically, there were fewer claims for hypertension, hyperlipidemia, and depression among participants in the intervention cohort during the follow-up period, whereas the matched control cohort showed significant increases in claims for health care services associated with renal disease, ischemic heart disease, osteoporosis, and musculoskeletal disorders compared with the intervention cohort. This strengthened the conclusion that a mix of comorbid health issues, many associated with diabetes, were primarily responsible for the all-cause findings. The SEM analysis reinforced this perspective and allowed us to develop significant insight into the systemic effects. Here, we found that postintervention disease burden had a direct effect of increasing postintervention total health care costs by US \$4066. The total effects associated with the intervention cohort on postintervention total health care costs were $-\text{US } \$2220$ after controlling for all the other variables in the structural model. SEM allowed us to break the total program effects into a $-\text{US } \$815$ direct effect component and $-\text{US } \$1504$ indirect effect through a postintervention disease burden component. This provides a more accurate estimate of the direct impact of the program's health care savings.

Even though the study participants were matched on the amount of care they required for comorbid illness in the preintervention period, during the postintervention period, the matched controls required more care for both diabetes-related comorbidities (eg, renal disease and ischemic heart disease) and other conditions (eg, osteoporosis and musculoskeletal disorders). In comparison, the intervention cohort experienced a period in which utilization

for hypertension, hyperlipidemia, and depression decreased and care for their other comorbid disorders remained stable. The decreased utilization for depression correlates with one of the main findings of the Lorig et al study that indicated that patients' self-reported depression was reduced [29]. This may be evidence that the structure or content of the intervention was more successful in decreasing distress and depression associated with diabetes management than the mobile health app studied by Quinn et al [36]. Furthermore, intervention cohort reported increased aerobic activity in the original clinical study, which may correlate with improvements seen in claims for depression and musculoskeletal disorders [29]. It is also plausible that the increased activity may have resulted in weight loss, leading to decreased utilization for hypertension and hyperlipidemia. The successful improvement in physical activity in the intervention cohort adds to the evidence reported by Gibson et al: technology-based strategies can positively impact behavioral measures in patients with diabetes [37].

The generally positive findings reported here need to be put into context with the mixed economic results from DSME reported in other studies. Wertz et al, with a pharmacist-led intervention, found that overall, diabetes-related costs increased to a greater extent in the intervention cohort compared with the control cohort, primarily driven by increased office visits, outpatient visits, and pharmacy claims, whereas costs associated with cardiovascular-related emergency room visits and inpatient visits were significantly higher in the control cohort [38]. This is similar to the findings of Sullivan et al, in which intervention subjects had improved HbA_{1c} values, greater experience of hypoglycemic events, and higher utilization and cost measures after receiving nonspecified diabetes education compared with the control group [21]. The Asheville project, which monitored the 5-year impact of a pharmacist-led diabetes education program, found that total direct medical costs declined while diabetes prescription costs increased; however, there was no control group for comparison [39]. A systematic review of pharmacist-led diabetes education studies by Wang et al showed positive economic cost savings results in the intervention groups compared with usual care [22]. Finally, a study by Burton et al found improvements in clinical measures but no impact on health care utilization or cost [40]. None of these studies are directly analogous to this study. The Wertz study and the Asheville project involved financial incentives that were lacking in this study [38,39]. Burton et al used an eight-session educational program with many possible educational add-ons for a much underserved population [40]. Sullivan et al observed differences between controls and a cohort that had claims codes for any diabetes education but not a specific program [21]. Nevertheless, these studies suggest that not all educational programs are equal, and we have a great deal to learn about intensity of intervention, mode of delivery, use of financial incentives, and population mix before we can make generalized statement about the cost-effectiveness of diabetes self-management education. This is especially evident in the failure of a mobile app, designed for diabetes self-management but lacking a structured educational program, that was noted by Thies et al [41].

The studies that most closely approximate the present observational research are those that utilize computer technology, mobile phone technology or both to disseminate educational and motivational messages. The intervention by Nundy et al was similar to this study in that it had a self-paced education program that lasted for 10 weeks and differed in that it also used a system of reminders and alerts to impact behavior [25]. The control group for Nundy consisted of those who had not responded to invitations to participate in the program, and the possibility of introduced bias was noted by the authors [25]. Similar to our results, positive outcomes were achieved in behavioral, clinical, and economic arenas [25]. The systematic review of computer-based diabetes education programs by Jackson et al, which in general showed positive economic results, had no data regarding hospitalization utilization for internet-based studies but did show a decrease in hospitalization in the computer-based studies, mirroring the success of the intervention by Lorig et al [24,29]. Although several studies are underway that aim to evaluate the impact of technology-mediated diabetes education on cardiovascular comorbidities [42] and cardiovascular risk reduction ([43], results have not yet been published [42,43]. It is important to distinguish that our study is novel in that it reports evidence that comorbidities can be impacted positively.

Limitations

This study is subject to some specific limitations regarding the sample and more general limitations associated with retrospective analyses of health care claims data. First, we could only identify a portion of the original Lorig et al (2016) study sample because only 82.5% had private insurance [29]. Of those who did have private insurance, we could only find 55% based on name, gender, age, and residence. In addition, some study patients did not meet the 12-month pre- and 12-month postinsurance coverage requirement. Third, it is possible that

unobserved patient-specific factors might not be balanced between the Intervention and matched control group, such as health literacy, health coping strategies, life style behaviors, and socioeconomic status, which can impact patients' access to health promoting resources. As is well known, compliance with interventions, whether physician-directed or self-guided, is always a limitation on effectiveness. Some participants did not attend all the sessions and consequently did not complete the full course of the program. To compensate for this, we used intention-to-treat analysis, which may have conservatively measured the program impact. Finally, our population, from a national sample of commercially insured persons, may not be generalizable to other populations such as a Medicaid population. This potential to create a different result was reflected in Burton et al's work in an underserved population [40]. Further study in patient groups with different demographic and economic characteristics would be beneficial.

Conclusions

This study expanded upon the Lorig et al (2016) study by assessing the program's impact on diabetes-specific and all-cause health care utilization and the cost associated with utilization, as well as shifts in care for common comorbid chronic conditions often observed among patients diagnosed with DM. We found that beyond the clinical benefits seen in the previous Lorig et al report, participants in the peer-facilitated BCBH-D program experienced reduced all-cause health care utilization and medical costs relative to the matched control group [29]. Therefore, we conclude that BCBH-D and other intensive, theory-driven diabetes, self-management programs can produce important clinical changes along with related health care cost savings. The results will require replication in other commercial and noncommercial databases—in particular, the apparent indirect impact of intervention on other disease states will require prospective controlled trial research to verify them.

Acknowledgments

Funding for the BCBH-D program was provided by the Bristol-Meyers-Squibb Foundation and the Anthem Foundation to the NCOA. The observational and economic portion of the research was funded by Anthem Inc through HealthCore Inc, a wholly owned subsidiary of Anthem Inc. Reviewers at Anthem approved of the manuscript before publication.

The authors wish to thank the National Council of Young Men's Christian Association of the United States and the Oasis Institute for conducting the small group sessions. The authors also wish to thank Philip Ritter for management of the Stanford Patient Education Center clinical data and Elizabeth Marks for providing manuscript preparation support.

Conflicts of Interest

KL has the potential to receive royalties from the BCBH-D program and from the book used in the intervention. RMT, QM, and ARD are employees of Anthem. JG has no potential conflict of interest to disclose.

References

1. Centers for Disease Control and Prevention. CDC. Atlanta, GA: US Department of Health and Human Services; 2017. National Diabetes Statistics Report URL: <https://www.cdc.gov/diabetes/data/statistics/statistics-report.html> [accessed 2017-03-28] [WebCite Cache ID 6uLy2ayyY]
2. American Diabetes Association. Economic costs of diabetes in the US in 2012. *Diabetes Care* 2013 Apr;36(4):1033-1046 [FREE Full text] [doi: [10.2337/dc12-2625](https://doi.org/10.2337/dc12-2625)] [Medline: [23468086](https://pubmed.ncbi.nlm.nih.gov/23468086/)]
3. Zhuo X, Zhang P, Barker L, Albright A, Thompson TJ, Gregg E. The lifetime cost of diabetes and its implications for diabetes prevention. *Diabetes Care* 2014 Sep;37(9):2557-2564. [doi: [10.2337/dc13-2484](https://doi.org/10.2337/dc13-2484)] [Medline: [25147254](https://pubmed.ncbi.nlm.nih.gov/25147254/)]

4. Piette JD, Kerr EA. The impact of comorbid chronic conditions on diabetes care. *Diabetes Care* 2006 Mar;29(3):725-731. [Medline: [16505540](#)]
5. Druss BG, Marcus SC, Olfson M, Tanielian T, Elinson L, Pincus HA. Comparing the national economic burden of five chronic conditions. *Health Aff (Millwood)* 2001;20(6):233-241 [FREE Full text] [Medline: [11816664](#)]
6. Maddigan SL, Feeny DH, Johnson JA. Health-related quality of life deficits associated with diabetes and comorbidities in a Canadian National Population Health Survey. *Qual Life Res* 2005 Jun;14(5):1311-1320. [Medline: [16047506](#)]
7. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002 Nov 11;162(20):2269-2276. [Medline: [12418941](#)]
8. Ciechanowski PS, Katon WJ, Russo JE. Depression and diabetes: impact of depressive symptoms on adherence, function, and costs. *Arch Intern Med* 2000 Nov 27;160(21):3278-3285. [Medline: [11088090](#)]
9. Glasgow RE, Ruggiero L, Eakin EG, Dryfoos J, Chobanian L. Quality of life and associated characteristics in a large national sample of adults with diabetes. *Diabetes Care* 1997 Apr;20(4):562-567. [Medline: [9096981](#)]
10. Li R, Shrestha SS, Lipman R, Burrows NR, Kolb LE, Rutledge S. Diabetes self-management education and training among privately insured persons with newly diagnosed diabetes--United States, 2011-2012. *MMWR Morb Mortal Wkly Rep* 2014 Nov 21;63(46):1045-1049 [FREE Full text] [Medline: [25412060](#)]
11. Pantalone KM, Hobbs TM, Wells BJ, Kong SX, Kattan MW, Bouchard J, et al. Clinical characteristics, complications, comorbidities and treatment patterns among patients with type 2 diabetes mellitus in a large integrated health system. *Br Med J Open Diabetes Res Care* 2015;3(1):e000093 [FREE Full text] [doi: [10.1136/bmjdr-2015-000093](#)] [Medline: [26217493](#)]
12. U.S. Department of Health and Human Services. Healthypeople.gov. Washington, DC: U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion Healthy People 2020 URL: <https://www.healthypeople.gov/2020/topics-objectives/topic/diabetes> [accessed 2017-10-20] [WebCite Cache ID 6uLztP684]
13. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* 2014 Jun;37(6):1759-1766. [doi: [10.2337/dc13-1386](#)] [Medline: [24855158](#)]
14. Sherifali D, Hess R, McTigue KM, Brozic A, Ng K, Gerstein H. Evaluating the feasibility and impact of an internet-based lifestyle management program in a diabetes care setting. *Diabetes Technol Ther* 2014 Jun;16(6):358-362. [doi: [10.1089/dia.2013.0278](#)] [Medline: [24580377](#)]
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 [FREE Full text] [doi: [10.1016/j.jdiacomp.2013.07.003](#)] [Medline: [24332469](#)]
16. Palmas W, March D, Darakjy S, Findley SE, Teresi J, Carrasquillo O, et al. Community health worker interventions to improve glycemic control in people with diabetes: a systematic review and meta-analysis. *J Gen Intern Med* 2015 Jul;30(7):1004-1012 [FREE Full text] [doi: [10.1007/s11606-015-3247-0](#)] [Medline: [25735938](#)]
17. Odgers-Jewell K, Ball LE, Kelly JT, Isenring EA, Reidlinger DP, Thomas R. Effectiveness of group-based self-management education for individuals with type 2 diabetes: a systematic review with meta-analyses and meta-regression. *Diabet Med* 2017 Aug;34(8):1027-1039. [doi: [10.1111/dme.13340](#)] [Medline: [28226200](#)]
18. Lian JX, McGhee SM, Chau J, Wong CK, Lam CL, Wong WC. Systematic review on the cost-effectiveness of self-management education programme for type 2 diabetes mellitus. *Diabetes Res Clin Pract* 2017 May;127:21-34. [doi: [10.1016/j.diabres.2017.02.021](#)] [Medline: [28315575](#)]
19. Mash R, Kroukamp R, Gaziano T, Levitt N. Cost-effectiveness of a diabetes group education program delivered by health promoters with a guiding style in underserved communities in Cape Town, South Africa. *Patient Educ Couns* 2015 May;98(5):622-626 [FREE Full text] [doi: [10.1016/j.pec.2015.01.005](#)] [Medline: [25641665](#)]
20. Odnoletkova I, Ramaekers D, Nobels F, Goderis G, Aertgeerts B, Annemans L. Delivering diabetes education through nurse-led telecoaching. Cost-effectiveness analysis. *PLoS One* 2016;11(10):e0163997 [FREE Full text] [doi: [10.1371/journal.pone.0163997](#)] [Medline: [27727281](#)]
21. Sullivan SD, Dalal MR, Burke JP. The impact of diabetes counseling and education: clinical and cost outcomes from a large population of US managed care patients with type 2 diabetes. *Diabetes Educ* 2013;39(4):523-531. [doi: [10.1177/0145721713486525](#)] [Medline: [23640302](#)]
22. Wang Y, Yeo QQ, Ko Y. Economic evaluations of pharmacist-managed services in people with diabetes mellitus: a systematic review. *Diabet Med* 2016 Apr;33(4):421-427. [doi: [10.1111/dme.12976](#)] [Medline: [26433008](#)]
23. Dalton JE. Web-based care for adults with type 2 diabetes. *Can J Diet Pract Res* 2008;69(4):185-191. [doi: [10.3148/69.4.2008.185](#)] [Medline: [19063808](#)]
24. Jackson CL, Bolen S, Brancati FL, Batts-Turner ML, Gary TL. A systematic review of interactive computer-assisted technology in diabetes care. Interactive information technology in diabetes care. *J Gen Intern Med* 2006 Feb;21(2):105-110 [FREE Full text] [doi: [10.1111/j.1525-1497.2005.00310.x](#)] [Medline: [16390512](#)]
25. Nundy S, Dick JJ, Chou C, Nocon RS, Chin MH, Peek ME. Mobile phone diabetes project led to improved glycemic control and net savings for Chicago plan participants. *Health Aff (Millwood)* 2014 Feb;33(2):265-272 [FREE Full text] [doi: [10.1377/hlthaff.2013.0589](#)] [Medline: [24493770](#)]

26. Stanford Self-Management Programs. Selfmanagementresource. Palo Alto, CA: Stanford University; 2016. URL: <http://www.selfmanagementresource.com/> [accessed 2017-10-20] [WebCite Cache ID 6uM0FIgPt]
27. Lorig K, Ritter PL, Laurent DD, Plant K, Green M, Jernigan VB, et al. Online diabetes self-management program: a randomized study. *Diabetes Care* 2010 Jun;33(6):1275-1281 [FREE Full text] [doi: [10.2337/dc09-2153](https://doi.org/10.2337/dc09-2153)] [Medline: [20299481](https://pubmed.ncbi.nlm.nih.gov/20299481/)]
28. Lorig K, Ritter PL, Turner RM, English K, Laurent DD, Greenberg J. Benefits of diabetes self-management for health plan members: a 6-month translation study. *J Med Internet Res* 2016;18(6):e164 [FREE Full text] [doi: [10.2196/jmir.5568](https://doi.org/10.2196/jmir.5568)] [Medline: [27342265](https://pubmed.ncbi.nlm.nih.gov/27342265/)]
29. Lorig K, Ritter PL, Turner RM, English K, Laurent DD, Greenberg J. A diabetes self-management program: 12-month outcome sustainability from a nonreinforced pragmatic trial. *J Med Internet Res* 2016 Dec 15;18(12):e322 [FREE Full text] [doi: [10.2196/jmir.6484](https://doi.org/10.2196/jmir.6484)] [Medline: [27979790](https://pubmed.ncbi.nlm.nih.gov/27979790/)]
30. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992 Jun;45(6):613-619. [Medline: [1607900](https://pubmed.ncbi.nlm.nih.gov/1607900/)]
31. Bentler PM, Chou C. Practical issues in structural modeling. *Sociol Methods Res* 1987 Aug 01;16(1):78-117. [doi: [10.1177/0049124187016001004](https://doi.org/10.1177/0049124187016001004)]
32. Hoyle RH. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications; 1995.
33. Kline RB. *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York, NY: Guilford Press; 2005.
34. SAS Institute. *SAS Software, Version 9.3*. Cary, NC: SAS Institute Inc; 2015.
35. StataCorp. *Stata Statistical Software, Version Release 14*. College Station, TX: StataCorp LP; 2015.
36. Quinn CC, Swasey KK, Crabbe JC, Shardell MD, Terrin ML, Barr EA, et al. The impact of a mobile diabetes health intervention on diabetes distress and depression among adults: secondary analysis of a cluster randomized controlled trial. *JMIR Mhealth Uhealth* 2017 Dec 07;5(12):e183 [FREE Full text] [doi: [10.2196/mhealth.8910](https://doi.org/10.2196/mhealth.8910)] [Medline: [29217502](https://pubmed.ncbi.nlm.nih.gov/29217502/)]
37. Gibson B, Marcus RL, Staggers N, Jones J, Samore M, Weir C. Efficacy of a computerized simulation in promoting walking in individuals with diabetes. *J Med Internet Res* 2012;14(3):e71 [FREE Full text] [doi: [10.2196/jmir.1965](https://doi.org/10.2196/jmir.1965)] [Medline: [22576226](https://pubmed.ncbi.nlm.nih.gov/22576226/)]
38. Wertz D, Hou L, DeVries A, Dupclay Jr L, McGowan F, Malinowski B, et al. Clinical and economic outcomes of the Cincinnati Pharmacy Coaching Program for diabetes and hypertension. *Manag Care* 2012 Mar;21(3):44-54 [FREE Full text] [Medline: [22471165](https://pubmed.ncbi.nlm.nih.gov/22471165/)]
39. Cranor CW, Bunting BA, Christensen DB. The Asheville Project: long-term clinical and economic outcomes of a community pharmacy diabetes care program. *J Am Pharm Assoc (Wash)* 2003;43(2):173-184. [Medline: [12688435](https://pubmed.ncbi.nlm.nih.gov/12688435/)]
40. Burton J, Eggleston B, Brenner J, Truchil A, Zulkiewicz BA, Lewis MA. Community-based health education programs designed to improve clinical measures are unlikely to reduce short-term costs or utilization without additional features targeting these outcomes. *Popul Health Manag* 2017 Apr;20(2):93-98 [FREE Full text] [doi: [10.1089/pop.2015.0185](https://doi.org/10.1089/pop.2015.0185)] [Medline: [27268018](https://pubmed.ncbi.nlm.nih.gov/27268018/)]
41. Thies K, Anderson D, Cramer B. Lack of adoption of a mobile app to support patient self-management of diabetes and hypertension in a federally qualified health center: interview analysis of staff and patients in a failed randomized trial. *JMIR Hum Factors* 2017 Oct 03;4(4):e24 [FREE Full text] [doi: [10.2196/humanfactors.7709](https://doi.org/10.2196/humanfactors.7709)] [Medline: [28974481](https://pubmed.ncbi.nlm.nih.gov/28974481/)]
42. Seto E, Ware P, Logan AG, Cafazzo JA, Chapman KR, Segal P, et al. Self-management and clinical decision support for patients with complex chronic conditions through the use of smartphone-based telemonitoring: randomized controlled trial protocol. *JMIR Res Protoc* 2017 Nov 21;6(11):e229 [FREE Full text] [doi: [10.2196/resprot.8367](https://doi.org/10.2196/resprot.8367)] [Medline: [29162557](https://pubmed.ncbi.nlm.nih.gov/29162557/)]
43. Lynch CP, Williams JS, Ruggeiro KJ, Knapp RG, Egede LE. Tablet-aided behavioral intervention effect on self-management skills (TABLETS) for diabetes. *Trials* 2016 Mar 22;17:157 [FREE Full text] [doi: [10.1186/s13063-016-1243-2](https://doi.org/10.1186/s13063-016-1243-2)] [Medline: [27005766](https://pubmed.ncbi.nlm.nih.gov/27005766/)]

Abbreviations

AIC: Akaike Information Criterion
BCBH-D: Better Choices Better Health Diabetes
CFI: Comparative Fit Index
DCI: Deyo-Charlson Comorbidity Index
DM: diabetes mellitus
DSME: Diabetes Self-Management Education
ED: emergency department
GEE: generalized estimating equation
HbA_{1c}: glycated hemoglobin
HIRE: HealthCore Integrated Research Environment
ICD-9: International Classification of Diseases-Ninth Revision
IRB: institutional review board
MSA: Metropolitan Statistical Area

NCOA: National Council on Aging

NFI: Normed Fit Index

QALY: quality-adjusted life years

QoL: quality of life

RMSE: root mean square error

SEM: structural equation model

Edited by G Eysenbach; submitted 31.10.17; peer-reviewed by N Bragazzi, M Lozano-Lozano; comments to author 01.02.18; revised version received 22.03.18; accepted 06.04.18; published 22.06.18.

Please cite as:

Turner RM, Ma Q, Lorig K, Greenberg J, DeVries AR

Evaluation of a Diabetes Self-Management Program: Claims Analysis on Comorbid Illnesses, Health Care Utilization, and Cost
J Med Internet Res 2018;20(6):e207

URL: <http://www.jmir.org/2018/6/e207/>

doi: [10.2196/jmir.9225](https://doi.org/10.2196/jmir.9225)

PMID: [29934284](https://pubmed.ncbi.nlm.nih.gov/29934284/)

©Ralph M Turner, Qinli Ma, Kate Lorig, Jay Greenberg, Andrea R DeVries. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.06.2018. 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.

Original Paper

Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions

Robert R Morris¹, PhD; Kareem Kouddous¹, MSc; Rohan Kshirsagar^{1,2}, BS; Stephen M Schueller³, PhD

¹Koko, New York, NY, United States

²Department of Computer Science, Columbia University, New York, NY, United States

³Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:

Robert R Morris, PhD

Koko

155 Rivington St

New York, NY, 10002

United States

Phone: 1 617 851 4967

Email: rob@koko.ai

Abstract

Background: Conversational agents cannot yet express empathy in nuanced ways that account for the unique circumstances of the user. Agents that possess this faculty could be used to enhance digital mental health interventions.

Objective: We sought to design a conversational agent that could express empathic support in ways that might approach, or even match, human capabilities. Another aim was to assess how users might appraise such a system.

Methods: Our system used a corpus-based approach to simulate expressed empathy. Responses from an existing pool of online peer support data were repurposed by the agent and presented to the user. Information retrieval techniques and word embeddings were used to select historical responses that best matched a user's concerns. We collected ratings from 37,169 users to evaluate the system. Additionally, we conducted a controlled experiment (N=1284) to test whether the alleged source of a response (human or machine) might change user perceptions.

Results: The majority of responses created by the agent (2986/3770, 79.20%) were deemed acceptable by users. However, users significantly preferred the efforts of their peers ($P<.001$). This effect was maintained in a controlled study ($P=.02$), even when the only difference in responses was whether they were framed as coming from a human or a machine.

Conclusions: Our system illustrates a novel way for machines to construct nuanced and personalized empathic utterances. However, the design had significant limitations and further research is needed to make this approach viable. Our controlled study suggests that even in ideal conditions, nonhuman agents may struggle to express empathy as well as humans. The ethical implications of empathic agents, as well as their potential iatrogenic effects, are also discussed.

(*J Med Internet Res* 2018;20(6):e10148) doi:[10.2196/10148](https://doi.org/10.2196/10148)

KEYWORDS

conversational agents; mental health; empathy; crowdsourcing; peer support

Introduction

Background

Conversational agents are software applications that respond to users with natural language, often with the goal of helping a user complete a task [1]. These agents reduce the need for visual interfaces or input devices and allow for more seamless interaction between humans and machines. Such tools have evolved rapidly in recent years, from smartphone integrations

(such as Apple's Siri, Android's Google Now, Samsung's S Voice, and Microsoft's Cortana) to stand-alone devices that are placed within the home (Amazon's Alexa and Google Home).

While ostensibly designed for utilitarian purposes—such as booking an airline ticket, ordering food, or playing music—some of these agents are also programmed to react to the user's emotional state. For instance, if you say you're feeling sad, Siri might reply, "I'm sorry to hear that" or, "from our deepest sadness springs our deepest joy." Unfortunately, as of today,

Siri is quick to repeat herself and she can only draw from a limited set of simple, generic phrases and platitudes. Furthermore, some of her remarks could be interpreted as flippant and insensitive (eg, “I would give you a shoulder to cry on, if I had one.”). If you offer a more detailed account of a stressful situation (eg, “I’m really nervous about an upcoming exam, I think I’m going to fail.”), she’ll simply say that she doesn’t understand. Alexa and Google Home exhibit similar characteristics; they occasionally acknowledge the distress of the user and react in kind, but only in a very general sense. In short, their ability to empathize remains inchoate.

This issue might be forgivable for some agents, such as those that are designed mostly for transactional interactions, but it becomes problematic when automated systems are increasingly being called upon to support health care applications, especially in the domain of mental health. Indeed, many new mental health applications in the market are making use of conversational agents and text-based dialogue systems [2]. For example, commercially available products like Woebot, 7Cups, and Koko have used chatbots for various tasks, such as providing psychological assessments and psychoeducational materials. These agents can be deployed on messenger systems (eg, Facebook Messenger, Kik, Twitter) and are designed to present mental health materials in an interactive and conversational style.

Woebot, a mental health app that relies exclusively on a bot for user interactions, was evaluated in a two-week trial among college students who self-identified as experiencing depression and anxiety [3]. Participants who used the service experienced a significant reduction in symptoms of depression whereas those who received an information control did not [3]. Similarly, “Shim” is a mental health chatbot that was designed for a nonclinical sample. In a controlled trial, users who engaged with the Shim bot experienced increased well-being and reduced perceived stress [4]. Thus, it appears that it is feasible to build such automated conversational agents, they can be engaging enough that some people may continue to use them over a relatively short period of time (ie, up to two weeks), and that those who use them report benefits on a variety of metrics related to mental health, including well-being, stress, and depression.

Although these mental health agents frequently ask users to disclose personal anxieties and vulnerabilities, their reactions remain limited to simple, short remarks. As of today, when a user discloses negative thoughts to Woebot, the agent says, “I see, you certainly have a lot going on at the moment.” The bot on Koko does not provide any sort of empathic reaction and simply acknowledges that the message has been received (“Thanks for sharing... sending this off to the Koko community”). As with Alexa and Siri, these agents are currently unable to reflect any deep understanding of the user’s particular situation.

This simple approach may be sensible today, given the current state of these technologies. An attempt to show empathy that misses the mark, however slightly, could offend the user. However, agents with sophisticated empathic abilities (ie, agents that seem to truly understand the user’s emotional experience)

could have a profound effect on the user. Previous research suggests that this ability can lead to increased user satisfaction [5] and affinity for the agent [6].

Regarding agents in mental health applications specifically, strong empathic abilities could be especially useful. Empathy is a commonly cited “nonspecific” factor in psychotherapy that has been proposed to be a major driver of the benefits that clients derive from treatment [7]. Empathy is a strong predictor of therapeutic alliance, which has itself been found to predict outcomes in various trials of psychotherapy [8]. In client-centered therapy, therapists might convey empathy through techniques such as active listening, reflection, and adopting a nonjudgmental stance and warm tone. Conversational agents might be able to model similar behaviors, at least through speech.

Researchers have taken strides in this direction, exploring ways for software agents to mimic emotions [9,10], adjust personality style [11], and programmatically communicate expressions of concern [12]. Yet, more work is needed before these agents can pass anything resembling an *empathic* Turing test—that is, the ability to engage in empathic dialogue in ways that are indistinguishable from a real human. To achieve this goal, the agent may need to express shared understanding, offer new perspectives, and generally take into account the unique situation and feelings experienced by the user. In the words of Carl Rogers, a pioneer of person-centered psychotherapy:

Being empathic is to perceive the internal frame of reference of another with accuracy and with the emotional components and meanings which pertain thereto... it means to sense the hurt or the pleasure of another as he senses it and to perceive the causes thereof as he perceives them. [13]

Considerable advances in commonsense reasoning and natural language processing would be needed to generate this kind of empathy from scratch. In this paper, we explored an alternative approach that sidestepped many of the remaining, unsolved challenges for artificial intelligence. Specifically, we examined whether a conversational agent could express rich, empathic understanding simply by repurposing preexisting emotional support data.

Goal of This Work

The goal of this work was to take initial steps towards building a conversational agent that can respond immediately, and convincingly, with empathic verisimilitude. To do this, we used a corpus-based approach; preexisting emotional support statements were drawn from a large corpus of online interactions and were presented as if they were authored by the agent, creating the semblance of personalized, empathic expression. We used information retrieval techniques and word embeddings to automate this process in real-time.

We first conducted a preliminary test of this system, assessing performance metrics and user perceptions. In a separate controlled study, we examined the upper limit of such a system; that is, if we are able to simulate an empathic agent that performs at the level of a human peer, how is it perceived by users? Lastly, we discuss some of the ethical implications of these

types of systems and some of the unintended consequences that might attend them.

System Design

Koko Platform

Emotional support statements were drawn from a corpus of Koko data. Koko is a mobile, peer-to-peer platform that aims to promote emotional resilience [14] that was derived from *Panoply*, a Web-based platform that was previously shown to reduce symptoms of depression [15]. Koko employs a text-based user-interface and is available on various messaging platforms (eg, Kik, Facebook Messenger, Twitter), as well as mobile and desktop browsers. A chatbot (KokoBot) is used to introduce users to the platform, teach cognitive reappraisal skills, and facilitate peer-to-peer interactions.

Unlike some other peer support platforms, Koko does not support repeated interactions between users (such as extended, private messages or multithreaded comments). Instead, all interactions follow a simple *post-response* format, with the chatbot passing messages between users seeking help and those who have opted to give help. When posting on Koko, users are prompted to describe a stressful situation and record any associated negative thoughts they may have. Users are taught to respond to each other's posts with positive reappraisals and messages of acceptance.

Character count limits of 950 and 600 are imposed on posts and responses, respectively. This convention further distinguishes Koko from many online peer support forums, such as those found on Reddit or Facebook. Most support forums do not impose character count limits, allowing users to write as much as they like. Long-form posts would be very difficult to repurpose for the short, query-response interactions one typically has with a conversational agent. By contrast, posts and responses on Koko are generally quite short, containing an average of 223.05 (SD 159.42) and 222.65 characters (SD 136.35), respectively.

Users rate the quality of each response they receive on a single-item, three-point Likert scale (*good, ok, bad*). For pragmatic reasons, we were not able to conduct multi-item assessments, such as those intended to capture various facets of *perceived empathy*. Lengthier assessments such as these are difficult to employ in consumer applications, since they can lead to high levels of user attrition. Additional outcome measures would need to be collected to specifically examine whether users found the responses empathic in the way this construct is traditionally operationalized. That being said, it is important to note that users on Koko are explicitly told to help each other feel supported and understood. As such, it is likely that many users considered various empathic factors when making their ratings (eg, *Did this response make me feel understood? Did it make me feel better?*). Future studies should deploy longer assessments to help test this issue more systematically.

All peer interactions on Koko were supervised by a hybrid human-machine moderation system. A suite of deep neural nets

was used to detect abusive behavior, inappropriate contributions, users in crisis, and various other infractions (see Kshirsagar et al [16] for a specific description of our crisis model and Calvo et al [17] for a general discussion of how natural language processing can be used to glean mental health information from text). Users at risk of harming themselves were referred to crisis resources, while those who behaved maliciously were banned from the platform.

Corpus

The corpus of peer interactions consisted of 72,785 posts and 339,983 responses. Content that was quarantined by our moderation system was not included, nor were responses that received "bad" ratings by users. Additionally, we removed any responses that included solicitations to chat on other platforms, such as Kik or Instagram. All of the aforementioned properties of the Koko platform (the style of interaction, the character limits, the strict moderation procedures, and the user labels) made for a highly structured corpus of data. Without this structure, it would have been very difficult to undertake the information retrieval approach we pursued. The anonymous nature of the dataset is also important, as it excludes cases where interaction partners addressed each other by name or drew upon shared experiences.

Back-End System

The two components of this system were: (1) a back-end system to automatically pair previously archived responses with incoming posts, and (2) a front-end system to display responses and solicit user feedback. We used an information retrieval approach to automatically return responses to the user. For any incoming post on our network, we searched for similar posts within the existing corpus of peer support interactions on Koko. Once we found a post, we looked at the associated responses. If we found one that was rated favorably, we returned it to the user (see Table 1 for examples).

The success of this method depends on the similarity between the incoming post and any that were archived previously. In an idealized case, when an incoming post matches a historical post verbatim, there is no need to solicit new responses from the network; responses to the historical post should be very well-matched to the new, incoming post. Conversely, if two posts are highly dissimilar, responses to one would not necessarily apply to the other.

To compute similarity between posts, we first used Elasticsearch [18], an open source full-text search engine that is designed to be faster and more scalable than typical relational databases. Specifically, we used Elasticsearch's *more like this* query, which computes the terms in a document with the highest term frequency-inverse document frequency (*tf-idf*) weights. The *tf-idf* ratio identifies sets of terms that are most discriminative for a document (in our case a "post") [19]. The top such terms are then used in a disjunctive query to find documents containing similar terms. This approach was used to return a set of potentially relevant posts as quickly as possible, so the system could respond in near real-time.

Table 1. Examples from the full end-to-end system. Spelling and grammatical errors, as well as profanities, are preserved to illustrate the nature of the peer interactions on the network. Here, we illustrate examples of responses that were deemed *good*, *ok*, and *bad*. The response that was rated *bad* (depicted in row 3) exhibited a mismatch in gender pronouns.

Incoming Post	Retrieved Post	Retrieved Response	User Rating
I have really bad social anxiety and im really afraid of public places, i cant go to the grocery store alone and im 17. i'll never get over it	I had really bad anxiety and I'm always scared to go places or to get out the house. And my mom doesn't do anything about it I'll never get over anxiety	I deal with anxiety too but it will get better. I know it seems hard right now but it will get better, please never forget that. I know it seems really hard but life gets better. Never forget that no matter what happens you can always get through it. It'll be okay. It will.	Good
I'm anxious about a class I'm taking because I don't know anyone in it. The class is going to be tricky because I skipped two levels to take this third level of the class, and it feels like i don't have anyone that can help me. :-(My anxiousness is going to get the best of me and cause me to fail the class by not participating.	It's the second week of school, and I'm very stressed about the amount of homework that I have and the level of classes I chose to take. I don't know how to handle it well that I'm going to fail all my classes and I'm so lost I don't think it will get better	I definitely understand how stressful that can be, and I'm sorry you're feeling this way An important thing to remember is that everyone is different. Sadly too many often place too much pressure on students, but just remember YOU and your health come FIRST.	OK
This Girl i like likes me too but i think ill fuck things up cause i am not really confident. Im really not confident	I really like this guy and he used to be really awkward around me and I was with him too but all of the sudden he's gotten really smooth and confident but I still can't even look him in the eye without my fave turning so red I just don't think I can be confident I really want to but I don't know how	hey, this is sweet. and I myself am a guy and its best to be yourself around us guys, that's what we love. a girl who is herself! good luck!	Bad

Among the set of posts that were returned by the *more like this* query, we then computed a word mover's distance (WMD) score [20] to further refine our measure of similarity between the incoming post and those returned by Elasticsearch. The WMD approach uses *word2vec*, a word-embedding procedure that learns a vector representation of words [21]. The word vectors encode the semantic relationships between words, positioning related words in nearby vector space. We used Google's pretrained *word2vec* model which includes word vectors for 3 million words and phrases, trained on approximately 100 billion words from the Google News dataset. WMD uses the *word2vec* embeddings to compute the distance between two documents, finding semantic similarities even when the documents have few words in common.

The WMD score was used to not only rank documents, but also to set a threshold to help decide whether the similarity between posts was sufficient to recycle a previous reply. We set a threshold that would return matches for approximately 10% of posts. Having no data to start with, we were not sure how well the documents would match in practice and what level of document similarity would lead to favorable outcomes. We reasoned that this threshold would at least give us a reasonably large dataset of user ratings to help evaluate the system and refine it further. Extremes on either end would be impractical; responding to 100% of posts would yield an intolerable false positive rate, whereas responding to 1% would likely improve precision, but not yield as much user feedback.

Front-End User Experience

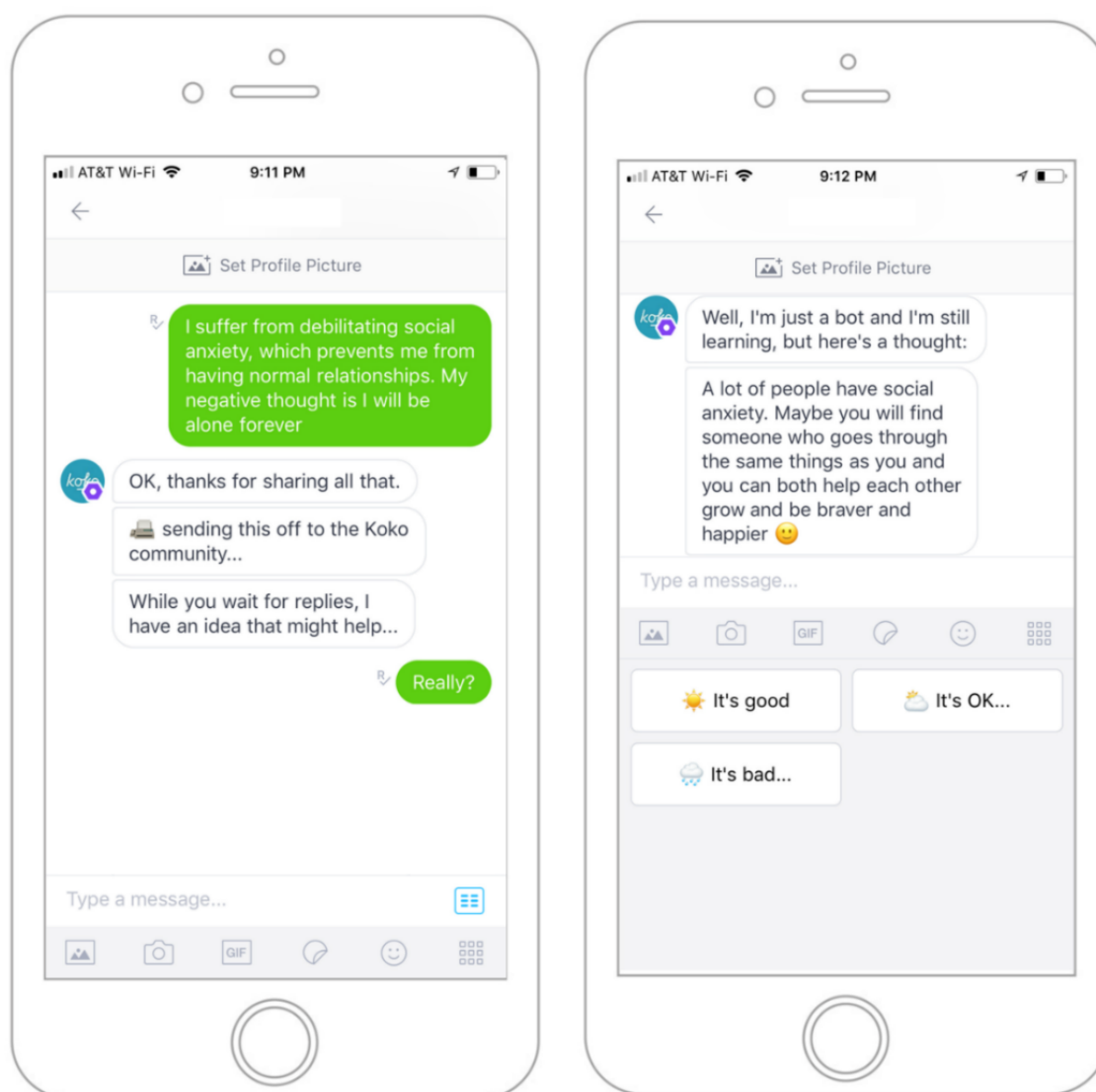
Once a preexisting response was retrieved, it was presented to the user as though it was algorithmically generated by the robot.

Users were not told that the agent was passing off other people's words as its own. Immediately after posting, the chatbot informed users that it might have a response of its own. Specifically, the bot said, "While you wait for responses, I may have an idea that might help..." After reading the response, users were asked to rate it on a three-point Likert scale (*good*, *ok*, *bad*), as shown in Figure 1. Users were encouraged to be honest and were told that their ratings would be hidden.

Participants

Participants included 37,169 individuals who signed up for Koko between mid-August and mid-September of 2016. These users were not asked to reveal their age, gender, or other personal information, so the demographics of this sample are unknown. However, in a separate survey of Koko users from 2017 (N=496), 65% identified as female and the majority were young adults (mean 18.24, SD 5.80). No significant changes occurred in the Koko platform or advertising methods; thus, it is likely that users in 2016 had similar characteristics as those surveyed in 2017. As with other research on other commercial well-being apps [22], Koko users accept a user agreement outlining the privacy policies and how usage data might be used for research purposes. The terms and conditions note that, "we do not require you to provide any personal information" and that, "we may share some or all of this usage with third parties in connection with research, analytics, or similar purposes." Users are told that their usage data may be used, "in aggregate form, that is, as a statistical measure, but not in a manner that would identify you personally."

Figure 1. A screenshot of the Koko platform, as seen on the Kik Messenger service. The agent automatically retrieves a response for the user while peer responses are collected from the network.



Results

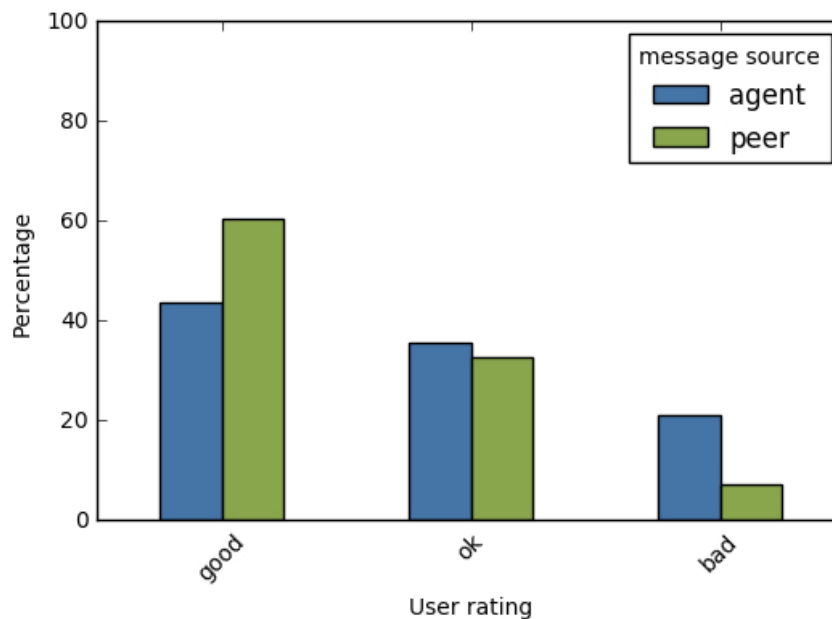
All retrieved responses were located and presented to the user within one second. The actual latency was likely lower but the timestamps on the data we evaluated only resolved to the nearest second. We evaluated user ratings for responses that came from the agent and from peers. In the time period of our analysis, we collected 3770 responses from the bot and 43,596 from peers. Response ratings differed significantly across conditions (see Figure 2). Responses composed by peers were significantly more likely to be rated as good compared to responses that came from the agent ($\chi^2(2)=981.20$, $P<.001$). However, 79.20% (2986/3770) of responses from the system were deemed *ok* or *good*, suggesting users mostly found the results acceptable (see Figure 2).

Conclusions

To automatically retrieve responses, we employed an unsupervised learning method. This approach, while useful to

bootstrap the system and collect user feedback, was not sufficient on its own. Responses from actual peers were rated significantly higher than responses generated from the automated system.

Inspection of the data revealed some serious errors due to gender discrepancies. For example, a match between posts might be incredibly close in all respects, except that the subject of one is a female while the other is a male; this led to situations in which a male user was inadvertently assumed to be a female (see row 3 in Table 1). This issue illustrates how a single phrase or word could make two posts quite dissimilar semantically even when much of the raw content overlaps considerably. The system also struggled when it retrieved responses that contained first-person personal accounts. Some of these accounts should not have been repurposed by the artificial agent (eg, "I've also struggled with an eating disorder...").

Figure 2. Distribution of user ratings for responses that were created by the nonhuman agent or human peers.

However, despite these shortcomings, it is perhaps surprising that the majority of responses from the bot were rated favorably. Even with a very simple, unsupervised model, many of the responses generated by the system were well-received. It is also worth noting that the bot was fairly ambitious in its attempts to empathize with the users; these were nuanced responses that addressed specific elements of the poster's concerns and so there was a great deal of potential for error.

Fortunately, there are many ways this system could be improved upon. In the future, other features could be added to the model to better capture the similarity between incoming and previously archived posts. For example, measures of linguistic style, sentiment, and topic could be included as additional features. Furthermore, the outcome ratings from users could be used to help train supervised models. These models would likely improve as the training set increases in size. As of this writing, there has been a more than tenfold increase in the size of the corpus. As more peers interact with the system, and more outcome data is collected, the models could grow increasingly accurate.

Still, even with larger datasets and improved models, the upper bound of this system remains an open question. When it comes to expressed empathy, people may always prefer humans to agents. To explore this question directly, we conducted a controlled study on the Koko platform that randomly attributed peer responses to either an agent or other humans.

Methods

Participants included 1284 Koko users who joined the platform between January 18 and 23 in 2018. At sign up, a segment of incoming Koko users was randomly assigned to one of two conditions. In the control condition, users were shown responses from their peers as usual. In the experimental condition, users

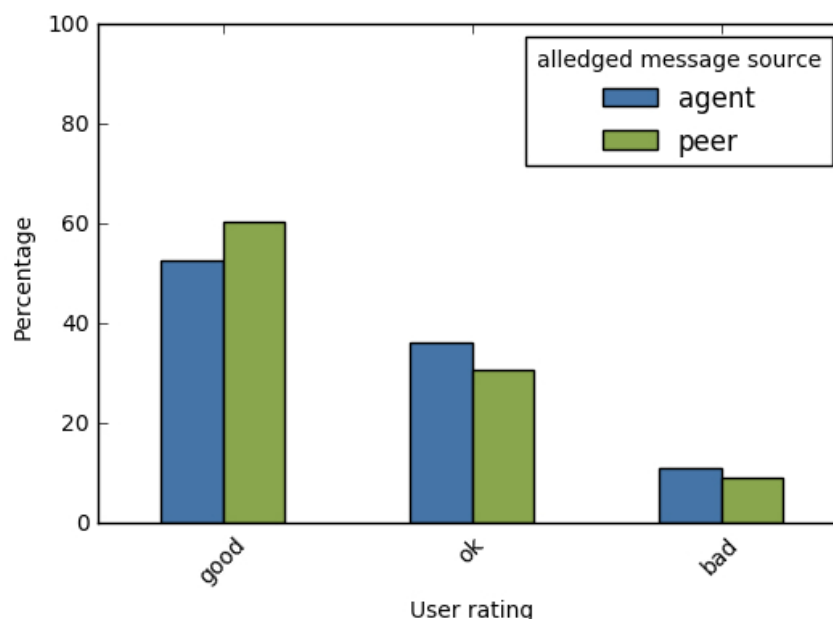
also received peer responses, but they were told they had come from an artificial agent. To limit the potential impact of this experiment on the greater Koko community, we randomly allotted 2/3 of users to the control condition and 1/3 to the experimental condition, as opposed to a 50/50 random split.

The only part of the experience that differed between conditions was the notification that preceded the delivery of the peer response. In the control condition, before returning a response, the bot said, "Someone replied to your post. Let's check it out." In the experimental condition, before returning a response, the bot said, "While you wait for responses, maybe I can help... I'm just a robot and I'm still learning, but here's a thought." The language was intentionally equivocal and submissive, in the hopes that the user would be more likely to forgive the system, should it fail. This approach also served to lower expectations, an approach that can mitigate disappointment when users experience a robot's shortcomings [23]. Users in both conditions rated it on the same three-point scale that was used previously. The manipulation was only performed the first time a user received a response, therefore ratings for subsequent responses were not included in the analysis.

Results

Findings

Users rated responses less favorably when they were told they came from an agent compared to a peer (see Figure 3). A chi-square analysis revealed a significant difference in the distribution of ratings between the two groups ($\chi^2(2)=9.23$, $P<.01$). When users believed an agent composed the response, they were significantly less likely to review it favorably, with only 51.6% (227/440) of users rating the response as *good* compared to 60.6% (510/842) of users rating responses as *good* when told the response came from a peer.

Figure 3. Distribution of ratings when users were told responses came from a peer, as compared to an artificial agent.

Conclusions

Despite the fact that all responses in this experiment were written by peers, and therefore should not differ in quality, the alleged source of these responses led to different perceptions.

This data is preliminary and may be specific to the particular context in which it was collected. However, it is clear that users harbored some resistance to the agent's empathic overtures. Future research is needed to better understand what the source of this resistance could be.

Discussion

Principal Findings

We created a new method through which conversational agents might simulate empathic expression. Specifically, we explored ways in which preexisting peer support data could be repurposed to help nonhuman agents express empathy. During the time in which it was deployed, the majority of the system's responses were well received by users. However, responses that were created by actual peers were deemed significantly more favorable. Better models and additional data could improve the performance of such a system, but an open question remains as to whether nonhuman agents could ever achieve parity with humans on empathic expression.

To explore this latter question, we conducted an experiment to see how the framing of an empathic message might affect how it is perceived. Users rated responses less favorably when they were led to believe they came from a nonhuman agent. This suggests that when it comes to empathic expression, conversational agents might be at a perpetual disadvantage. This is interesting given that the conceptual basis for human and machine comparisons often falls back to the Turing test or the degree to which a machine's behavior would be indistinguishable from that of a human. Our findings suggest that machines might have to do even better than humans to be

considered at the same point of emotional intelligence with empathic expressions.

One unique aspect of our approach was to create a system powered by a corpus of peer-created content. This approach differs from many technological tools that digitize skill-training approaches through didactic modules or interactive features or allow for peer communication through unstructured synchronous (eg, chatrooms or chat platforms) or asynchronous (eg, forums) methods. As these applications essentially digitize previous forms of interaction, many have adapted traditional techniques into either visually-driven technologies such as apps [24], or created chatbots that provide education or instruction in these skills themselves [3,4,14]. Such an approach takes advantage of the years of advances in clinical understanding and science of behavior change but fails to capitalize on the full affordances of new technologies. That is, such technologies allow space, time, and availability of a provider to be transcended but do not significantly revolutionize the delivery of services. Allowing peers to contribute to a platform and then scaling up automation on the basis of their contributions truly moves towards the democratization of delivery and contribution of services. It has been noted elsewhere that health care relies too heavily on "consumable interventions," which are interventions that once used, can never be used again [25]. Technologies are generally nonconsumable interventions in the sense that they can be used repeatedly without exhausting their therapeutic power to help additional people. Peer-powered interventions go a step further; that is, each use of the platform can actually contribute additional benefit to all future users.

Furthermore, other uses of large corpora of empathic chats have demonstrated that such data can provide important contributions to the understanding of actionable strategies to enhance crisis counseling [26] and therapist-client interactions [27]. Such understanding could contribute to the design of automated systems (as was done in this study) and could be applied to human-only interactions (ie, counseling or psychotherapy), or

could promote development of human-machine combinations of supportive tools. Elsewhere the exploration of human-machine combinations based on principles of empathic conversations and cognitive restructuring has been found to be a helpful step in allowing peers to have supportive conversations that mirror some processes in psychotherapeutic interactions [28]. Large corpora, applications of data science, and thoughtful design will likely all increase the quality of empathic expressions in the technologies of the future.

Limitations

There were several limitations to how the system was evaluated and designed. For example, the full end-to-end system was not tested in a controlled study with random assignment. As such, we do not know how our system affected other user behaviors, such as retention and overall engagement with the platform. The system design could also be improved upon. We found some success using unsupervised methods from information retrieval, but more complex models could be employed in the future.

With stronger models, the system could be more confident about when to draw upon preexisting responses and when to use some other approach. Hybrid approaches that combine corpus-based techniques with speech synthesis models could help provide more overall coverage for the system and could be used to create novel utterances that might even outperform those of a peer.

The controlled study had several shortcomings as well, as it was conducted within a very specific context (the Koko application). A similar study should be conducted with various agents within various applications to see if the effect generalizes across settings. Furthermore, we were not able to follow-up with users to assess the credibility of our experimental manipulation. It is possible that users were not fully convinced they were interacting with a machine, despite us claiming otherwise.

Lastly, it is possible the bot's personality and language style influenced the findings. The way the system framed its efforts could have had a huge effect on how its responses were received. In particular, the bot's self-deprecatory stance may have affected user perceptions. Guided by past research [23], we employed this approach in the hope that it might make users more lenient in their judgments of the bot. However, it is also possible that this approach led users to scrutinize the agent more closely and appraise its contributions more critically. Furthermore, past research suggests different agent personalities can have differential effects on how they are perceived [29,30], and so the overall personality of the Koko bot may have also affected user perceptions. Future research is needed to better assess the potential impact of these variables.

Conclusions

Despite its limitations, our approach suggests that it may be possible to simulate empathic expressions by drawing upon a rich corpus of social support data. Theoretically, it might eventually be possible to build a system that draws from millions of candidate responses with near perfect precision. The right response for the right user at the right time could automatically be selected, edited, and returned by the system. This capacity could help agents better serve users within the context of mental health interventions.

Despite these advances, there remains a real question as to whether such a system, even one as idealized as this, could ever outperform a peer. Nonhuman agents do not have lived human experience and so their attempts to express empathy might always appear inauthentic. The effort required to compose a supportive response may be as important as its actual substance. It can be powerful to know that someone else has listened to you, thought deeply about your situation, and then took the time to craft a considerate response. All of this is lost when a machine algorithm automatically generates a response. People often attribute human characteristics to machines and treat them similarly as real people [31], but an agent's empathic expressions may always be perceived slightly differently than a human's. It is possible that expressed empathy is a domain in which it will always be hard, if not impossible, for a robot to match human levels of performance.

Even if agents only provide, at most, a simulacrum of empathy, there are still important ethical implications to consider. In this paper, we examined some of the immediate, short-term effects of artificial empathic expression. Future work should assess how humans might relate to empathic agents over longer time periods, especially within the context of mental health interventions. The potential effects on well-being and social relationships are not well understood and many questions remain. Would people eventually seek emotional support from machines, rather than their friends and family? What might a machine miss in the course of these interactions? How can we get the benefits of empathic machines, without sacrificing existing human-human relationships, or otherwise imperiling the user? As Sherry Turkle warns, there is a risk that a robot that, "begins as a solution ends up a usurper" [32]. The Institute of Electrical and Electronics Engineers (IEEE) report on Ethically Aligned Design outlined similar concerns, arguing that agents that express emotion may have unintended effects on the user, such as over-bonding and misplaced trust [33]. To avoid these pitfalls, while still maximizing the potential benefits of this technology, designers will need to carefully scrutinize how these systems affect the well-being of users over time.

Acknowledgments

We thank all of the Koko users who engaged with the service throughout the duration of this research. We also thank members of the Koko community who provided helpful insights on the design of the platform. SMS is an investigator with the Implementation Research Institute, at the George Washington University in St. Louis, MO, who received an award from the National Institute of Mental Health (5R25MH08091607) and the Department of Veterans Affairs, Health Services Research & Development Service, Quality Enhancement Research Initiative.

Conflicts of Interest

RRM, KK, and RK are employees at Koko, a for-profit enterprise that provides mental health and safety services to large social networks. The Koko peer support service was used to conduct the research described in this paper.

References

1. Luger E, Sellen A. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. 2016 Presented at: ACM SIGCHI Conference on Human Factors in Computing Systems; 2016; New York, NY p. 5286-5297.
2. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. *J Med Internet Res* 2017 Jul 21;19(8):e267 [FREE Full text] [doi: [10.2196/jmir.7023](https://doi.org/10.2196/jmir.7023)] [Medline: [28784594](https://pubmed.ncbi.nlm.nih.gov/28784594/)]
3. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health* 2017 Jun 06;4(2):e19 [FREE Full text] [doi: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785)] [Medline: [28588005](https://pubmed.ncbi.nlm.nih.gov/28588005/)]
4. Ly K, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv* 2017;10:39-46.
5. Sandbank T, Shmueli-Scheuer M, Herzig J, Konopnicki D, Shaul R. EHCTool: managing emotional hotspots for conversational agents. 2017 Presented at: 22nd International Conference on Intelligent User Interfaces; 2017; Limassol, Cyprus p. 125-128.
6. Brave S, Nass C, Hutchinson K. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int J Hum Comput Stud* 2005 Feb;62(2):161-178. [doi: [10.1016/j.ijhcs.2004.11.002](https://doi.org/10.1016/j.ijhcs.2004.11.002)]
7. Elliott R, Bohart AC, Watson JC, Greenberg LS. Empathy. *Psychotherapy (Chic)* 2011 Mar;48(1):43-49. [doi: [10.1037/a0022187](https://doi.org/10.1037/a0022187)] [Medline: [21401273](https://pubmed.ncbi.nlm.nih.gov/21401273/)]
8. Nienhuis JB, Owen J, Valentine JC, Winkeljohn BS, Halford TC, Parazak SE, et al. Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: a meta-analytic review. *Psychother Res* 2016 Jul 07:1-13. [doi: [10.1080/10503307.2016.1204023](https://doi.org/10.1080/10503307.2016.1204023)] [Medline: [27389666](https://pubmed.ncbi.nlm.nih.gov/27389666/)]
9. Hegel F, Spexard T, Wrede B, Horstmann G, Vogt T. Playing a different imitation game: interaction with an empathic android robot. 2006 Presented at: IEEE-RAS International Conference on Humanoid Robots; 2006; Genova, Italy p. 56-61.
10. Riek L, Paul P, Robinson P. When my robot smiles at me: enabling human-robot rapport via real-time head gesture mimicry. *J Multimodal User Interfaces* 2010;3(1):99-108. [doi: [10.1007/s12193-009-0028-2](https://doi.org/10.1007/s12193-009-0028-2)]
11. Siddique F. Zara Returns: Improved personality induction and adaptation by an empathetic virtual agent. 2017 Presented at: ACL System Demonstrations; 2017; Vancouver, Canada p. 121-126.
12. Leite I, Pereira A, Mascarenhas S, Martinho C, Prada R, Paiva A. The influence of empathy in human-robot relations. *Int J Hum Comput Stud* 2013;71(3):250-260. [doi: [10.1016/j.ijhcs.2012.09.005](https://doi.org/10.1016/j.ijhcs.2012.09.005)]
13. Rogers C. *A Way of Being*. New York, NY: Houghton Mifflin Harcourt; 1995.
14. Koko. 2018. URL: <https://itskoko.com/> [accessed 2018-02-15] [WebCite Cache ID 6xG1wgIy7]
15. Morris RR, Schueller SM, Picard RW. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. *J Med Internet Res* 2015;17(3):e72 [FREE Full text] [doi: [10.2196/jmir.4167](https://doi.org/10.2196/jmir.4167)] [Medline: [25835472](https://pubmed.ncbi.nlm.nih.gov/25835472/)]
16. Kshirsagar R, Morris R, Bowman S. Detecting and explaining crisis. 2017 Presented at: Fourth Workshop on Computational Linguistics and Clinical Psychology; 2017; Vancouver, Canada p. 66-73. [doi: [10.18653/v1/W17-3108](https://doi.org/10.18653/v1/W17-3108)]
17. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng* 2017 Jan 30;23(05):649-685. [doi: [10.1017/S1351324916000383](https://doi.org/10.1017/S1351324916000383)]
18. Kononenko O, Baysal O, Holmes R, Godfrey M. Mining modern repositories with Elasticsearch. 2014 Presented at: 11th Working Conference on Mining Software Repositories; 2014; Hyderabad, India p. 328-331.
19. Chowdhury G. *Introduction to modern information retrieval*. London: Facet Publishing; 2010.
20. Kusner M, Sun Y, Kolkin N, Weinberger K. From word embeddings to document distances. 2015 Presented at: International Conference on Machine Learning; 2015; Lille, France p. 957-966.
21. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: 26th International Conference on Neural Information Processing Systems; 2013; Lake Tahoe, USA p. 3111-3119.
22. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Seeing "big" picture: big data methods for exploring relationships between usage, language, and outcome in Internet intervention data. *J Med Internet Res* 2016 Aug 31;18(8):e241 [FREE Full text] [doi: [10.2196/jmir.5725](https://doi.org/10.2196/jmir.5725)] [Medline: [27580524](https://pubmed.ncbi.nlm.nih.gov/27580524/)]
23. Paepcke S, Takayama L. Judging a bot by its cover: an experiment on expectation setting for personal robots. 2010 Presented at: 5th ACM/IEEE international conference on Human-robot interaction; 2010; Nara, Japan p. 45-52.
24. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, 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](https://doi.org/10.2196/mhealth.3713)] [Medline: [25689790](https://pubmed.ncbi.nlm.nih.gov/25689790/)]

25. Muñoz RF. Using evidence-based internet interventions to reduce health disparities worldwide. *J Med Internet Res* 2010;12(5):e60 [FREE Full text] [doi: [10.2196/jmir.1463](https://doi.org/10.2196/jmir.1463)] [Medline: [21169162](https://pubmed.ncbi.nlm.nih.gov/21169162/)]
26. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Trans Assoc Comput Linguist* 2016;4:463-476 [FREE Full text] [Medline: [28344978](https://pubmed.ncbi.nlm.nih.gov/28344978/)]
27. Xiao B, Imel ZE, Georgiou PG, Atkins DC, Narayanan SS. “Rate my therapist”: automated detection of empathy in drug and alcohol counseling via speech and language processing. *PLoS One* 2015;10(12):e0143055 [FREE Full text] [doi: [10.1371/journal.pone.0143055](https://doi.org/10.1371/journal.pone.0143055)] [Medline: [26630392](https://pubmed.ncbi.nlm.nih.gov/26630392/)]
28. O’Leary K, Schueller S, Wobbrock J, Pratt W. Suddenly, we got to become therapists for each other: designing peer support chats for mental health. 2018 Presented at: CHI Conference on Human Factors in Computing Systems; 2018; Montreal, Canada p. 331.
29. Lee K, Peng W, Jin S, Yan C. Can robots manifest personality?: an empirical test of personality recognition, social responses, and social presence in human-robot interaction. *J Commun* 2006;56(4):754-772. [doi: [10.1111/j.1460-2466.2006.00318.x](https://doi.org/10.1111/j.1460-2466.2006.00318.x)]
30. Tapus A, Țăpuș C, Matarić MJ. User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intel Serv Robotics* 2008 Feb 13;1(2):169-183. [doi: [10.1007/s11370-008-0017-4](https://doi.org/10.1007/s11370-008-0017-4)]
31. Nass C, Steuer J, Tauber E. Computers are social actors. 1994 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 1994; Boston, MA p. 72-78.
32. Turkle S. *Alone together: why we expect more from technology and less from each other*. New York, NY: Basic Books; 2012.
33. Chatila R. The IEEE global initiative for ethical considerations in artificial intelligence and autonomous systems standards. *IEEE Robotics and Automation Magazine* 2017;24(1):110. [doi: [10.1109/MRA.2017.2670225](https://doi.org/10.1109/MRA.2017.2670225)]

Abbreviations

IEEE: Institute of Electrical and Electronics Engineers

Tf-idf: term frequency-inverse document frequency

WMD: word mover’s distance

Edited by R Calvo, M Czerwinski, J Torous, G Wadley; submitted 15.02.18; peer-reviewed by A Parks, R Calvo, A Suh; comments to author 14.03.18; revised version received 17.04.18; accepted 23.04.18; published 26.06.18.

Please cite as:

Morris RR, Kouddous K, Kshirsagar R, Schueller SM

Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions

J Med Internet Res 2018;20(6):e10148

URL: <http://www.jmir.org/2018/6/e10148/>

doi: [10.2196/10148](https://doi.org/10.2196/10148)

PMID: [29945856](https://pubmed.ncbi.nlm.nih.gov/29945856/)

©Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, Stephen M Schueller. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.06.2018. 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.

Original Paper

Applying the Principles for Digital Development: Case Study of a Smartphone App to Support Collaborative Care for Rural Patients With Posttraumatic Stress Disorder or Bipolar Disorder

Amy M Bauer¹, MS, MD; Sarah Hodsdon², BA; Jared M Bechtel¹, BS, BA; John C Fortney^{1,3}, PhD

¹Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

²John Snow, Inc, Boston, MA, United States

³Health Services Research and Development Service Center of Innovation for Veteran-Centered and Value-Driven Care, VA Puget Sound Health Care System, Seattle, WA, United States

Corresponding Author:

Amy M Bauer, MS, MD

Department of Psychiatry and Behavioral Sciences

University of Washington

1959 NE Pacific Street

Box 356560

Seattle, WA, 98195 6560

United States

Phone: 1 2062218385

Email: abauer1@uw.edu

Abstract

Background: Despite a proliferation of patient-facing mobile apps for mental disorders, there is little literature guiding efforts to incorporate mobile tools into clinical care delivery and integrate patient-generated data into care processes for patients with complex psychiatric disorders.

Objective: The aim of this study was to seek to gain an understanding of how to incorporate a patient-provider mobile health (mHealth) platform to support the delivery of integrated primary care–based mental health services (Collaborative Care) to rural patients with posttraumatic stress disorder and/or bipolar disorder.

Methods: Using the Principles for Digital Development as a framework, we describe our experience designing, developing, and deploying a mobile system to support Collaborative Care. The system consists of a patient-facing smartphone app that integrates with a Web-based clinical patient registry used by behavioral health care managers and consulting psychiatrists. Throughout development, we engaged representatives from the system's two user types: (1) providers, who use the Web-based registry and (2) patients, who directly use the mobile app. We extracted mobile metadata to describe the early adoption and use of the system by care managers and patients and report preliminary results from an in-app patient feedback survey that includes a System Usability Scale (SUS).

Results: Each of the nine Principles for Digital Development is illustrated with examples. The first 10 patients to use the smartphone app have completed symptom measures on average every 14 days over an average period of 20 weeks. The mean SUS score at week 8 among four patients who completed this measure was 91.9 (range 72.5–100). We present lessons learned about the technical and training requirements for integration into practice that can inform future efforts to incorporate health technologies to improve care for patients with psychiatric conditions.

Conclusions: Adhering to the Principles for Digital Development, we created and deployed an mHealth system to support Collaborative Care for patients with complex psychiatric conditions in rural health centers. Preliminary data among the initial users support high system usability and show promise for sustained use. On the basis of our experience, we propose five additional principles to extend this framework and inform future efforts to incorporate health technologies to improve care for patients with psychiatric conditions: design for public health impact, add value for all users, test the product and the process, acknowledge disruption, and anticipate variability.

(*J Med Internet Res* 2018;20(6):e10048) doi:[10.2196/10048](https://doi.org/10.2196/10048)

KEYWORDS

mHealth; mental health; primary health care; rural health; post-traumatic stress disorders; PTSD; bipolar disorder; depression

Introduction

Background

Although there are thousands of smartphone apps for mental health conditions available on the marketplace [1], mobile tools have yet to have a substantial impact in the delivery of mental health services [2]. The majority of these tools are stand-alone, patient-facing, self-help apps that have not been subject to any evaluation or regulation. Research has demonstrated that very few people sustain use of such self-guided interventions [3-7], which has prompted clinical investigators to explore new models of technology-supported care [5,8-10]. Efforts to incorporate mobile tools into clinical care delivery and integrate patient-generated data into care processes are emerging. However, there is little research to guide these efforts, and few existing tools have demonstrated the capacity to link patients and providers directly [10-12]. Early experience suggests incorporating new health information technology tools into clinical practice is potentially disruptive because it directly impacts providers' task behaviors and requires them to alter routines [13,14]. Adoption of health technologies in general, and patient-facing technologies in particular, thus has lagged [15].

Objectives

To gain an understanding of how to incorporate a patient-provider mobile health (mHealth) platform into the clinical care of rural patients with posttraumatic stress disorder (PTSD) and bipolar disorder, we designed, developed, and deployed a smartphone app that is currently in use within the context of a large clinical trial, the Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT; NCT02738944). SPIRIT is a pragmatic trial by design, and therefore, the procedures and workflow are designed to approximate real-world practice [16-18]. The SPIRIT app supports the primary care-based delivery of mental health services through an evidence-based model of treatment delivery called Collaborative Care, described in additional detail below, which is the intervention offered in one arm of the treatment trial.

Here, we describe how the Principles for Digital Development [19,20], a set of best practices informing the development of technology-enabled programs, apply to the design and deployment processes for the SPIRIT app. We report early data on adoption of the SPIRIT app by care managers and patients, patients' feedback on their use of the SPIRIT app, and lessons learned from the deployment process. This report focuses on the technology and its integration into the clinical model and is not an evaluation of effectiveness of the overall technology-supported care model. Recognizing that the requirements for mHealth tools differ from general requirements for mobile development [11], we propose extensions to the Principles for Digital Development for mHealth tools that are intended to support clinical services. We also propose the development of guidelines for mental health-specific digital design and user experience best practices.

Methods

Clinical Context

The SPIRIT app mobile system was designed to support the effective delivery of Collaborative Care. Collaborative Care is a model for delivering treatments for common mental disorders in primary care settings that is twice as effective as usual depression care [21-23]. For nearly two decades, it has been regarded as a best practice and has been widely disseminated [24-26]. Collaborative Care is defined by core principles that specify that care is team-based and patient-centered. Additionally, Collaborative Care provides measurement-based treatment-to-target, an evidence-based practice that involves routinely monitoring patient outcomes with standardized measures (eg, the Patient Health Questionnaire-9, PHQ-9 [27] for depression) and adjusting treatments when a patient is not improving [28]. Collaborative Care teams provide population-based care, which means providing care manager outreach to all of the patients in a defined population, not just those who show up for clinic visits [29]. This is in contrast to usual primary care in which patients often *fall through the cracks*. The detection and treatment of PTSD and bipolar disorder in primary care is poor [30-38]. Moreover, care is not measurement-based, follow-up is infrequent and ad hoc, and there is no proactive outreach [28,39].

To support the Collaborative Care workflow (Figure 1), the University of Washington developed a Web-based patient registry that providers use to track patient visits and outcomes using standardized measures. The registry differs from an electronic health record (EHR) in its design for effective management of an entire patient population, key features that EHRs lack, and specific support for providers' workflows. The registry, named the Care Management Tracking System (CMTS), supports workflows through reminders for proactive outreach and follow-up for patients who are not engaging in care and flags for patients who have not improved and may benefit from treatment changes.

The SPIRIT app was designed to provide a patient interface with CMTS, which is an entirely clinician-facing tool, and support patient's own self-management and communication with the care manager. Given that empowering patients to improve self-management is a key goal of Collaborative Care, engaging patients in the use of digital technology to facilitate effective, whole-person care is a good fit for the Collaborative Care model [12,40,41]. Potential benefits of the SPIRIT app include improving patient engagement, increasing satisfaction by offering a convenient, asynchronous method for patient-provider communication, and enhancing measurement-based care through timely remote symptom monitoring. Providers may experience improved efficiency through reduced documentation because patients enter their own symptom scores and through reduced need for time-consuming synchronous telephone outreach and follow-up. In addition, the SPIRIT app was designed to increase the

capacity of care managers so they can enroll more patients on their panel. By automating some of their clinical activities and helping them prioritize patients for outreach, the SPIRIT app is intended to increase the reach of the care manager and have a greater impact on population health.

Mobile System Description

The SPIRIT app is a free, password-protected Android app that securely transmits patient-generated data from a patient's smartphone to the care manager through CMTS (Figure 2). It supports two user types: (1) patients, the primary users of the mobile app and (2) providers, who receive outputs from the mobile app on an online patient management system. The SPIRIT app was developed on CommCare, an open source, Software-as-a-Service (SaaS) mobile data collection platform. CommCare was selected for SPIRIT app development based on its strong evidence base in mHealth [42], robust case management functionality, and its turnkey application builder—which provided a graphical user interface (UI) to enable rapid agile design of the system. Data from the SPIRIT app is transmitted securely with encryption standards consistent with the Health Insurance Portability and Accountability Act (HIPAA) of 1996 from the mobile device to the CommCare server. CMTS pulls newly uploaded patient data from the CommCare server via an application programming interface hourly, or immediately if requested by the provider. The SPIRIT app is available in English and Spanish. Details of the system have been previously published [43] and are summarized below.

Patient Interface: Study to Promote Innovation in Rural Integrated Telepsychiatry App

The SPIRIT app is organized into seven modules: Check In, View Progress, Learn More, Reach Out for Help, Safety Plan, Settings, and About the SPIRIT App (Figure 3). The Check In module allows patients to self-monitor and report symptoms to their care manager by completing rating scales for depression (PHQ-9), mania (SPIRIT mania scale), PTSD (PTSD Checklist for DSM-5, PCL-5), and medication adherence. In the View Progress module, patients see a graph of their scores to track their own weekly progress over time. The Learn More module has psychoeducational materials about their condition, psychiatric medications, and tips for managing common side effects. It also includes information about how the Collaborative Care model works and Behavioral Activation, a psychotherapy provided by SPIRIT care managers. Patients can access stories from people living with PTSD or bipolar disorder through links to consumer advocacy websites. If patients choose to customize the SPIRIT app, they can use the Reach Out for Help module to make phone calls directly to supportive, self-programmed contacts such as friends, family, or care providers. The Safety Plan module displays general information about how to contact emergency services such as the National Suicide Prevention Lifeline, Crisis Text Line, and Lifeline Chat for all patients regardless of whether or not they enter a personal plan and a

personal safety plan for patients who choose to enter one. The Settings module facilitates these customizations and also allows them to personalize their Check In day, the time and message content of short message service (SMS) reminders, and select which symptom scales to report on. They can create personalized reminders to take medications or for recurring activities as part of their Behavioral Activation treatment plan and submit new contact information for their care manager if they change their phone number. The About the SPIRIT App module includes frequently asked questions that explain how the SPIRIT app works and about their data privacy within the system. Patients are prompted to provide feedback through an in-app survey (details provided below).

Provider Interface: Care Management Tracking System

The SPIRIT app is directly integrated with CMTS, which care managers use daily in their Collaborative Care workflow to track patient encounters and view progress. To accommodate the SPIRIT app data, we expanded existing CMTS pages to incorporate data from the SPIRIT app and created new features supporting two major functions: (1) registering patients to use the SPIRIT app and manage their account and (2) viewing data their patients enter into the SPIRIT app. To register a patient, a care manager only needs to enter the patient's mobile phone number into a new CMTS mobile registration page that initiates a patient self-registration process. The patient will then receive an SMS with instructions and a link to download the app. In the CMTS mobile registration page, we embedded links to patient handouts and care manager materials about the SPIRIT app (described below). To view patient data, care managers view an alert on their Reminders page (the first page they see after logging in) to indicate when new app data has been imported for any patient on that care manager's caseload. If a patient reports suicidal thoughts on the PHQ-9, a specific alert is triggered so that care managers can quickly identify the patient and determine whether additional support such as telephone outreach is indicated. Additional pages in CMTS allow care managers to determine which patients have new scores and view individual item scores for each measure. Thus, SPIRIT app data is integrated throughout CMTS pages and helps care managers visualize app data at the level of the entire caseload, the individual patient, and the encounter.

App Design Process

The process for designing the SPIRIT app adhered to the Principles for Digital Development and was informed by earlier work with two mobile apps designed for Collaborative Care for depression [19,20]. These principles were created to guide international development organizations, health organizations, and mobile developers in integrating best practices into the design of digital technology-supported service delivery programs. They include nine living guidelines that are intended to be revised over time [20].

Figure 1. Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT) Collaborative Care workflow. CMTS: Care Management Tracking System; PTSD: posttraumatic stress disorder.

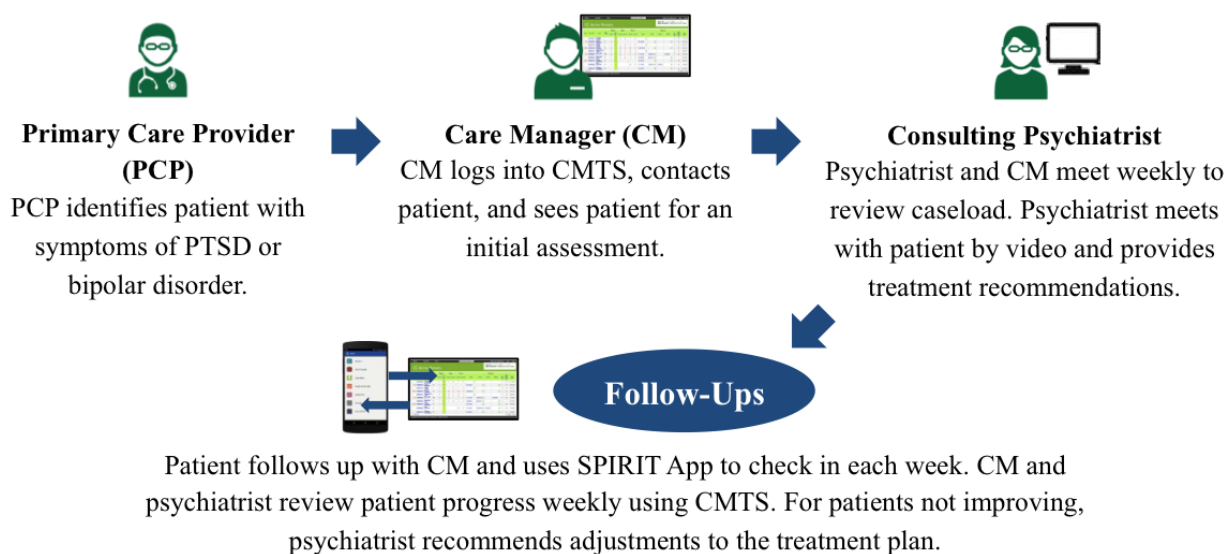


Figure 2. Diagram of the Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT) app mobile system.

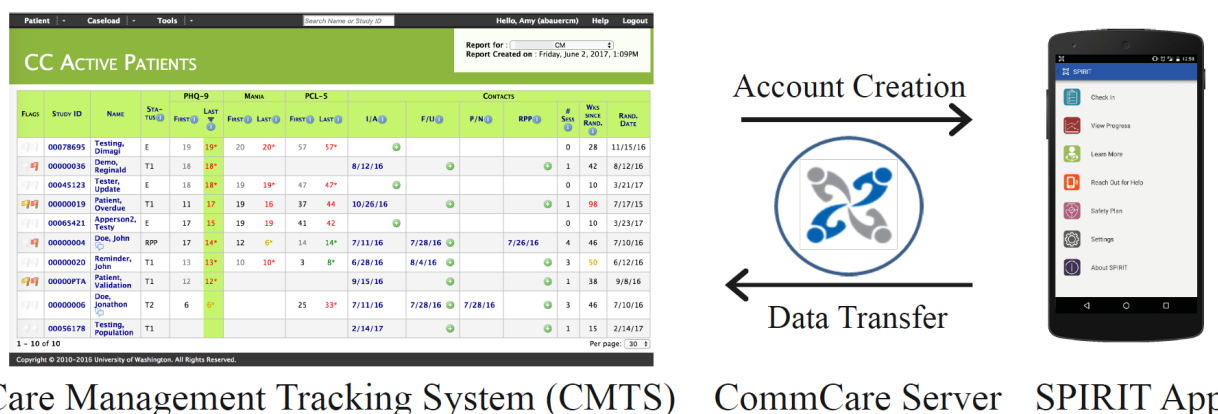
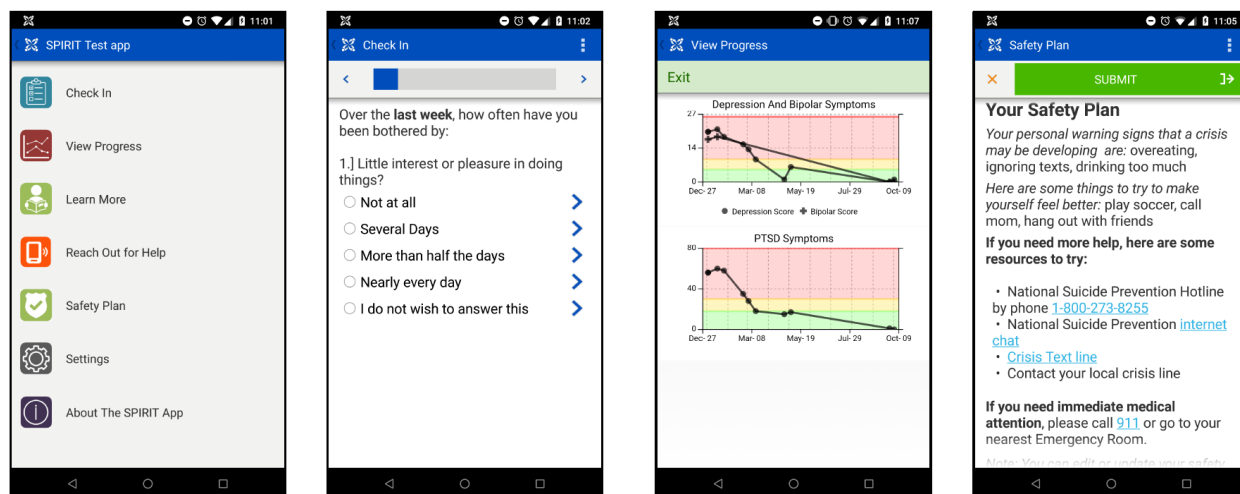


Figure 3. Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT) app screenshots.



Textbox 1. Principles for digital development [19, 20].

1. Design with the user
2. Understand the existing ecosystem
3. Design for scale
4. Build for sustainability
5. Be data driven
6. Use open standards, open data, open source, and open innovation
7. Reuse and improve
8. Address privacy and security
9. Be collaborative

The current set of principles ([Textbox 1](#)) have emerged from a collaborative, community-driven effort that began in the late 2000s with the recognition of struggles with implementation that were faced by digital programs and reflect involvement of diverse organizations such as the World Health Organization, World Bank, United Nations Children's Fund, the United Nations Development Program, the United States Agency for International Development, the Swedish International Development Agency, and the Bill and Melinda Gates Foundation. The Principles for Digital Development have been endorsed by Dimagi, the organization which developed the SPIRIT app [19].

A detailed description of the steps involved in the SPIRIT app design and development, including proof of concept and other preliminary activities, was previously published [43] and are summarized below ([Textbox 2](#)). Service providers such as care managers, health center directors, supervisors, and psychiatrists informed the core content specifications built in CMTS and the SPIRIT app, while patients representative of the final SPIRIT target population were engaged to provide feedback on early concept development, as well as direct feedback and design direction for the app's user experience, including layout, expected use, new features, and visual design. The project faced challenges with identifying and recruiting a larger cadre of representative patient users to engage in early stage design research, limited by partner health centers' current active patient list. However, working through the Consumer Advisory Board (CAB) at these early stages was a critical resource in early assessments and research because its members contributed an active interest in communicating patient perspectives in new models to improve mental health care delivery. In the Results section, we illustrate each of the Principles for Digital Development using examples from the SPIRIT app design and development processes.

Sample and Evaluation Metrics

SPIRIT is a pragmatic clinical trial, and therefore, the procedures and workflow are designed to mirror real-world practice with minimal direct involvement of centralized study staff. Patients receive SPIRIT clinical services for a maximum of 52 weeks. Patients randomized to the Collaborative Care arm of the study can be offered the SPIRIT app by their care manager at any time during the study period. At this time, there are 16 care managers with at least one patient enrolled in the study. We report initial data on the adoption of the SPIRIT app mobile system by five care managers and 10 patients.

Metadata from the CommCare system includes the new user activation date and the time stamp for all data submitted through the SPIRIT app, including each symptom measure. From this mobile metadata, we determined the number of days elapsed since each patient activated the SPIRIT app and the number of symptom measures each patient submitted. Then we calculated the interval of completion of measures as the number of days since activation divided by the number of measures completed. To assess persistence of use, we calculated the number of weeks elapsed between submission of the most recent symptom measure and the date of data extraction.

Patients receive SMS text message notifications to provide feedback through the SPIRIT app at 4, 8, 26, and 40 weeks after activation. The feedback module contains the SUS [46], a widely used 10-item self-report measure of usability for a variety of technologies. Scores range from 0 to 100, and scores above 80 are in the top 10th percentile of usability for all technologies and products tested [47]. Free-text feedback is also elicited in the feedback module. Patients also complete independent research interviews at 6 and 12 months post baseline that include questions about the impact of the SPIRIT app and potential burden. We also ask reasons nonusers did not use the SPIRIT app. Results from these independent research interviews will be reported in a future publication because the SPIRIT study is still ongoing.

Textbox 2. Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT) app design activities.

Proof of concept phase

- Pilot mixed-methods study of mobile health (mHealth) augmentation of Collaborative Care for depression and anxiety [8]
- Partner with Dimagi to conduct design, development, and usability testing of a depression Collaborative Care app
- Partner with one rural health center (interview one care manager, one patient)
- Integrate depression Collaborative Care app with Care Management Tracking System (CMTS)
- Pilot test linkage between depression Collaborative Care app and CMTS (held focus groups with three care managers; interviews with patient users)

Design and development phase

- Conduct focus group #1 with Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT) Consumer Advisory Board (CAB) to propose initial SPIRIT app concept and gather feedback. CAB comprises a group of individuals who reflect the target user profile and patient population of the SPIRIT deployment
- Conduct focus group #2 with SPIRIT CAB to refine SPIRIT app concept
- Partner with 12 rural community health centers in three states
- Define SPIRIT app design high-level scope
- Build SPIRIT app version 0 high-fidelity prototype and storyboard based on feedback from the CAB and proof of concept user focus groups
- Conduct focus group #3 with SPIRIT CAB to present storyboard of care manager and patient personas, low-fidelity prototype, and elicit feedback
- Refine prototype design; update high fidelity prototype
- Develop insight-driven usability testing framework
- Recruit representative users of the target patient population for usability testing; define participation incentives, consent, and testing logistics
- Conduct usability tests with 5 participants, including patients actively in care at SPIRIT's health center partner locations [44,45]
- Incorporate feedback from usability testing into app prototype
- Review prototype with programmatic experts, physicians, and study team; elicit and refine feedback
- Finalize SPIRIT app version 1.0

Deployment phase

- Develop patient and provider education materials
- Announce SPIRIT app launch on study website and newsletter and routine calls with study sites
- Launch SPIRIT app in three states
- Host a live webinar demonstration of SPIRIT app for care managers and study investigators
- Foster peer learning using the SPIRIT app in routine care manager training calls
- Invite feedback about how to make improvements in the system or supporting materials, both through patient engagement and an app-supported feedback form
- Provide SPIRIT information and technology (IT) email support for technical issues
- Offer care manager incentives to enroll their first patient
- Develop self-guided training on the SPIRIT app for new care managers when turnover occurs and to onboard new sites

Results

Application of Principles for Digital Development

The SPIRIT app design and development illustrates the application of the Principles for Digital Development for technology-enabled service delivery programs and provides new insights on how to optimize these principles within a mental health context [19,20].

Principle 1. Design With the End User

The SPIRIT app was created through a human-centered design process that put two groups of end users (patients and providers) at the center of each project stage, including immersion visits to local clinics, user-driven requirements gathering and ideation with multiple stakeholders (such as end users, care managers, supervisors, and consumer advisory groups), 1:1 and group prototype testing and interviews, multiple prototype refinement periods, and implementation driven by local clinics and their care managers [48]. The SPIRIT CAB (consisting of eight consumers and representatives from national consumer advocacy

groups), experts in Collaborative Care, and providers and patients in rural clinics were engaged as domain experts [49] and target users in a participatory design process throughout development (Textbox 2). In the postlaunch deployment phase, we continue to obtain feedback from patients through the in-app feedback survey and from clinics and care managers as part of a team-learning process described under *Principle 9: Be Collaborative*. Our longitudinal engagement has yielded improvements in every aspect of the SPIRIT app mobile platform, including the modules of the app, the CMTS provider interface, and our materials and methods to support training and deployment. Further details on specific changes that have been introduced based on stakeholder feedback are provided in the section below, *Incorporation of specific feedback*, and Multimedia Appendix 1.

Principle 2. Understand the Existing Ecosystem

Understanding the needs, challenges, and perspectives of Collaborative Care team members, especially the patient, was crucial to developing a mobile tool to extend this care delivery. The research team spent considerable time and effort to understand the context and align the technology with user needs and requirements. Contextual factors included federal law governing health information (ie, HIPAA) and the context of the rural settings, including available infrastructure and mobile phone adoption. Although national data at the time indicated high rates of mobile phone ownership [50], there was a lack of information about device ownership in the specific rural regions that SPIRIT targets. Verifying the nature and prevalence of local mobile phone ownership and connectivity was a key challenge, as was user recruitment for local usability testing given that some of the SPIRIT clinics are small and therefore have few eligible patients. To date, 45.4% (169/372) of patients enrolled in SPIRIT indicated that they own an Android device during the baseline patient survey. Given the uncertainty about device ownership before launch, the SPIRIT CAB encouraged us to consider a multiplatform system that included a patient website and an interactive voice response system in addition to an Android app. We carefully considered a multiplatform system; however, we opted against expanding the scope of the mobile system for this initial demonstration given constraints on time and resources and a desire to prioritize quality by developing a robust system.

Principle 3. Define and Plan for Scale

SPIRIT will be the largest clinical trial ever conducted of rural Americans with psychiatric disorders and will provide Collaborative Care services to 500 patients from one of 12 community health centers across Washington, Arkansas, and Michigan. The SPIRIT app links to CMTS, which is used in 22 programs in 24 states throughout the United States and one Canadian province and has helped over 100,000 patients receive better care. The SPIRIT app is available to patients from all of the clinics in the research study. To use the SPIRIT app at large-scale in routine, rural practices, the installation process was simplified and mapped to care managers' workflows so that patients could install the app independently without direct support from the research team. SPIRIT also developed an SMS-driven self-registration process to explore whether this

common communication mechanism is a new approach to helping patients download mHealth apps to their phones privately on their own. This new self-registration process was a key focus for usability testing to ensure that the process was simple. Although the SPIRIT app was designed specifically for use during the study, with minimal modification, the platform could be used at scale in other CMTS implementations, if the experience from SPIRIT supports such expansion.

The SPIRIT app is flexibly designed and is organized into seven modules, as described above. Unlike the depression Collaborative Care app, the SPIRIT app supports multi-condition (depression, bipolar disorder, and/or PTSD) management. The same underlying CommCare platform could be used to support a variety of psychiatric or medical conditions, either by changing the content within the existing modules or by adding or removing modules. As a result, the SPIRIT app is well-positioned for vertical scale to additional clinical sites, as well as horizontal scale into additional psychiatric or medical conditions. Presently, patients select which symptom rating scales to complete; however, it would be possible to enable providers to use CMTS to select rating scales from a menu of choices to tailor care for patients with a variety of conditions. Selecting a mobile platform that enables quick scale-up, reuse, and local adaption was central to SPIRIT development and the team's decision to select CommCare as the underlying technology platform. CommCare's evidence base demonstrates the platform's ability to support both public and private health systems' to scale mHealth solutions both vertically and horizontally, as well as provide "template" apps that can be used in different programmatic domains.

Principle 4. Build for Sustainability

To enable a strong programmatic and technical pathway to sustainability, the SPIRIT app was designed as an augmentation to the Collaborative Care services that are provided within the research study. By integrating with CMTS, the SPIRIT app was built to provide a complementary role within an existing patient management system, in contrast to stand-alone, consumer-facing apps currently dominant in the market. The Patient-Centered Outcomes Research Institute funds SPIRIT and has a strong interest in promoting sustainability of interventions that are supported by research trials. Throughout the SPIRIT study, the research team considered the sustainability of the two clinical care models under study through engagement with policy makers, professional organizations, and health care providers and funders such as the Community Health Plan of Washington, the Michigan Primary Care Association, and Community Health Centers of Arkansas. These partners are actively engaged in developing solutions to challenges such as billing for the clinical services in each of the study arms. To date, sustainability efforts have focused at the higher level of the clinical care models rather than focusing on the specific health technologies to support these. However, as the study progresses, it will be important to include considerations related to the health technologies including CMTS and the SPIRIT app mobile platform. Given that the SPIRIT app is only available for Android devices, it would be important to revisit the possibility of developing a multiplatform system if the platform is implemented for long-term use following the research trial.

Principle 5. Be Data Driven

The SPIRIT app development followed a systematic and data-driven design process, especially with regards to usability testing and measuring the effectiveness with which a user could successfully complete key tasks in the app. To understand representative patients' and care managers' ease of using the SPIRIT app, our usability testing focused on an insight-driven study of the new self-registration process and users' ability to navigate to the core components of the SPIRIT app without training. Four essential usability tasks were defined and tested with five representative users. Predefined metrics and goals were established with success thresholds to determine whether or not the functionality tested was usable or not. No more than three pieces of functionality were tested at a single time or contiguously to limit user burden and avoid missing vital feedback at the end of the test. Feedback informed the final SPIRIT app design. Incorporating all high-priority feedback into the app was constrained by a short development time frame and the emergent need to invest more technical resources in expanding the scope of CommCare's self-registration process, such that we could reduce the number of steps required by a user to successfully register, install, download, and begin using the app. However, this feedback was captured and shared with CommCare's developers for later in the platform's roadmap and patient-as-end-user product developments.

A major impetus for the development of the SPIRIT app mobile system was to generate knowledge about how to incorporate a patient-provider mobile platform into the clinical care of patients with complex psychiatric needs in primary care settings. This emphasis on knowledge generation drives a strong commitment to data collection, analysis, and interpretation. Moreover, Collaborative Care is a clinical model based on principles that concurrently emphasize measurement and accountability and patient-centeredness. The evaluation plan for the SPIRIT app mobile system is similarly holistic and includes both qualitative and quantitative elements. Our extensive stakeholder engagement has generated a wealth of insights that have informed this effort, and additional qualitative evaluation from patients will occur via open-ended questions in the in-app feedback survey and through qualitative research interviews that will be conducted with a subset of SPIRIT participants. A variety of quantitative metrics are available through the CommCare platform, and we are now developing reports to aggregate data on SPIRIT app system usage and considering how best to use these results to inform our deployment efforts near real time. Our commitment to knowledge generation has translated into publishing early lessons while the study is still underway [43].

Principle 6. Use Open Standards, Open Data, Open Source, and Open Innovation

The SPIRIT app is built on CommCare, an open source mobile data collection platform. The SPIRIT app has both benefitted from and contributed to the platform by providing new requirements necessary to open up the platform to patient-as-end-users and resources to invest in further development. The majority of the features in the SPIRIT app were built upon CommCare's existing platform functionality

such as case management, complex logic and branching algorithms, mobile graphing, and SMS-based reminders and data collection. Several of the unique requirements for SPIRIT necessitated the development of new functions such as the self-registration process via SMS and username self-management features (such as password reset and recovery workflows), and a streamlined download experience and UI that required no training compared with prior CommCare deployments in 50 countries. Both of these were subject to rigorous review during usability testing and, as a result, have contributed new infrastructure to the CommCare platform that is now available open source to other health systems around the world.

Principle 7. Reuse and Improve

The SPIRIT app mobile system functions as a patient portal for CMTS, a provider-only registry tool that has been used for nearly a decade to support Collaborative Care. The extensive experience with CMTS as a registry provided a strong foundation for the development of the SPIRIT app system. To leverage and improve upon existing mHealth investments, the SPIRIT app was built on an open source SaaS platform that enables agile software design and rapid prototyping and that has been used and improved through implementations by over 500 organizations in more than 50 countries. We leveraged the strengths of the platform throughout development, as feedback from stakeholders was iteratively integrated into the final design, and the core platform technology supported the breadth of requirements. The SPIRIT app was built from a template design of the depression Collaborative Care app and adapted for new contextual requirements. The flexible, modular design facilitated this adaptation, and a thorough quality assurance process was applied to yield further refinements and improvements. As mentioned previously, the SPIRIT app development also invested resources into the overall platform's capabilities and underlying architecture, so that future programs and applications can reuse and benefit from these improvements. The development team also conducted a review of common consumer-facing apps available on the app store focused on depression, PTSD, and bipolar self-management. Lessons from these deployments informed the design of the SPIRIT app, such as including Learn More modules to provide easy access to information about specific conditions. To support future design, adaptation, and uptake of mHealth for patient-provider collaboration in mental health care, there is an opportunity for SPIRIT to publish a copy of the SPIRIT app to the CommCare Exchange, an online free marketplace of mHealth template apps that enable other individuals and organizations to download a template app from the Exchange and modify it or use it under the terms of a Creative Commons license.

Principle 8. Address Privacy and Security

Protection of patients' privacy has been a priority throughout the development of the SPIRIT app. Unlike many mobile mental health apps, the SPIRIT app requires a username and password so that patients' sensitive data is accessible only to authorized users. Data transmission from the SPIRIT app to the CommCare server is encrypted with standards consistent with HIPAA. Both the CommCare platform and CMTS have methods to restrict

access to patient identifiers. The self-registration process and patient reminders utilize SMS text messages, and therefore, it is necessary to collect the patients' mobile phone number, which the care manager also needs to conduct outreach. We minimized the collection of such identifiable information to the minimum necessary and use a CMTS-generated identifier that is separate from a medical record number to link the SPIRIT app data to the CMTS patient record. In response to care manager requests to adapt the SPIRIT app for use on a tablet in the clinic waiting room, we have given careful consideration to the management of patient accounts and workflows to manage the handoff of the device to consecutive patients to prevent accidental access to an incorrect patient account. Patients use the SPIRIT app on an opt-in basis. We have encountered a limitation in communicating with patients about the protections in place for patient privacy and confidentiality. Care managers are the first-line providers who field patient questions; however, they do not have technical expertise in the system and therefore have needed additional support in responding to patients' technical questions related to data transmission and security.

Principle 9. Be Collaborative

Collaboration among multidisciplinary teams is fundamental to the delivery of Collaborative Care, as well as human-centered design [29,51]. The effective use of health information technologies, including mHealth tools, requires an extension of care teams to include experts in domains such as information technologies, software development, and human-centered design [11]. Effective team-based efforts require the development of new relationships and are facilitated by strong communication and empathy. In SPIRIT, we have 30 clinics from 12 organizations with substantial variability in the qualifications and experience of provider teams, the clinical workflows, and the volume and sociodemographic characteristics of patients. The majority of the interactions with sites have been conducted via videoconferencing because sites are in rural areas, and the study investigator team is distributed across several universities in three states. Similarly, many of the activities in developing the SPIRIT app have been conducted remotely and supplemented with site visits and in-person usability testing. We have embraced a team-learning approach to the development and deployment of the SPIRIT app in which care managers partner with us to explore how the SPIRIT app can improve the care they deliver and their patients' outcomes [52], and patients partner with us to provide direct feedback about what would make a tool more meaningful to use.

Study to Promote Innovation in Rural Integrated Telepsychiatry App Mobile System Adoption

As of January 25, 2018, five care managers and 10 patients have piloted the SPIRIT app mobile system, and another four patients have been registered by their care manager but have not yet completed the activation process. Patients have had access to the system for an average of 20 weeks (range 1-37 weeks) and submitted an average of 24.1 symptom measures (range 3-52 measures). All 10 patients completed the PHQ-9 and PCL-5 in equal frequencies (mean 9.7, range 1-26 measures). For the seven patients who have completed the SPIRIT mania scale, they have completed it in similar frequencies (mean 6.7, range

1-13 measures). Since activating the SPIRIT app, patients have completed symptom measures on average every 14 days (range 7-63 days). The wide range in the interval of symptom measure completion is accounted for because three patients have not submitted symptom measures for more than 16 weeks. All three of these patients submitted their most recent symptom score within 6 weeks of activating the app, suggesting they stopped using the app around that time point. The remaining seven patients have submitted symptom measures within the 2 weeks preceding data extraction, which is consistent with persistent use of the SPIRIT app. This subset of seven persistent app users have used the system for an average of 16 weeks (range 1-37 weeks) and are submitting symptoms measures on average every 10 days (range 7-16 days).

Patient Feedback Survey Results

Nine patients activated the app more than 8 weeks before data extraction, which means that they have been prompted at least twice to provide feedback (at week 4 and week 8). Two of the nine patients provided feedback at week 4, both of whom also provided feedback at week 8. Two additional patients provided feedback at week 8 only. Because more patients responded at week 8, we report SUS scores from this time point. Patients rated usability very highly at week 8 (mean 91.9, range 72.5-100).

In the free-text responses, patients reported that they most liked the SPIRIT app's simplicity, ease of use, convenience, reminders, being in touch with the care team, and prompts to think about oneself. Patients least liked that the symptom measures were a little long and that symptom monitoring was weekly rather than daily. Patients desired to receive feedback from their care team directly through the SPIRIT app. At present, data transmission from CMTS to CommCare is limited to information required for account registration, although two-way communication of clinical information could be developed in the future.

Incorporation of Specific Feedback

Our extensive, longitudinal engagement with stakeholders resulted in a number of improvements to the SPIRIT app, CMTS provider interface, and deployment processes. For each of these domains, [Multimedia Appendix 1](#) lists specific changes that were introduced based on stakeholder feedback, as well as the project phase and description of the findings that led to the change. Taken together, it is apparent that the iterative processes of obtaining and incorporating feedback occurred throughout project phases and resulted in improvements of every aspect of the system and deployment procedures. Importantly, because the focus of stakeholder engagement evolved over the design and development process, the different project phases provided insights into different aspects of the system such that not all project phases contributed to all of the final system components and procedures.

Discussion

Principal Findings

The technology to build mHealth apps to manage mental disorders is increasingly commonplace, yet their integration

into clinical care has lagged. The design of the SPIRIT app adhered to the Principles for Digital Development, which address many important issues in the development of technology-enabled services. Early feedback from patients using the SPIRIT app supports its excellent usability in the initial weeks of adoption; however, more feedback is needed from those patients who were offered the SPIRIT app but have not enrolled or have enrolled but have not submitted scores. A majority of SPIRIT app users are self-monitoring their symptoms on average every 2 weeks, which is considerably more frequently than clinic-based symptom assessment. Moreover, it is promising that most of the SPIRIT app users to date ($n=7$, 70%) have sustained such self-monitoring for 4 or more months.

Despite these encouraging findings, our findings must be interpreted within the context of certain limitations. The SPIRIT app has been developed in the context of a pragmatic research trial designed to mirror real-world rural practice; however, important differences from routine practice remain that could affect the sustainability of the clinical care models. The deployment of novel health technologies into routine clinical practice settings involves additional challenges that must be overcome. These include mapping the technology onto the clinical workflow, while being sensitive to the time investment needed on the part of care providers to fully participate in the design process; educating providers and patients; and developing robust support systems. Moreover, not all providers or patients will embrace digital health tools, even if they are well-designed and integrated into care processes, and further understanding of the limitations of digital mental health tools is needed. In SPIRIT, the overall number of care managers and patients who are currently using the system remains small, as are the response rates to in-app patient feedback surveys. Care managers suggest that lack of Android devices, a barrier we anticipated, has limited SPIRIT app use. This reinforces the notion that a single platform system would be unlikely to be tenable in routine clinical practice. More information is needed about reasons patients discontinue use of the SPIRIT app early in treatment to understand barriers to ongoing use.

Several features have emerged as potential improvements that could be considered in the future. Patients have expressed interest in using the SPIRIT app to communicate directly with their provider. Offering a secure messaging service may be one way to promote sustained use; however, careful consideration would need to be given to educating patients and providers with appropriate expectations for response time and whether the service can be used for crisis management. Patients have also expressed an interest in having a diary function or ability to annotate their symptom scores, and more work is needed to understand how best to structure and aggregate such information and communicate it to care teams. Finally, although the SPIRIT app integrates directly with the Web-based patient registry, neither directly integrates with clinic EHRs. Large-scale adoption of mHealth tools would be facilitated by the development of an integrated suite of health information and technology tools, with patient registry functions to support population health management and patient-facing mobile tools that integrate with EHRs.

As we continue to embark on team learning in collaboration with our clinical partners, we have gathered lessons from these challenges. Our experience with the design, development, and deployment of the SPIRIT app reveals additional considerations for mobile designers and clinical researchers seeking to advance mHealth integration into clinical care. Here we propose five additional principles: design for public health impact, add value for all users, test the product and the process, acknowledge disruption, and anticipate variability.

Design for Public Health Impact

Although the Principles for Digital Development have been informed by a number of public health efforts, we propose an explicit emphasis on optimizing population health by prioritizing the development of technology-supported services to maximize clinical impact for populations who experience the greatest need. To be truly impactful, digital mental health efforts must not focus exclusively on the most technologically savvy, well-resourced, or least impaired members of society [53]. Numerous studies have found that the benefits realized from efforts to improve the quality of mental health service delivery, such as through Collaborative Care, are greatest among the most disadvantaged patients [54-58]. Yet, just as efforts to improve health care quality have prioritized high-need populations, so too should efforts to implement technology-supported mental health service models that meet the unique needs of these populations. In SPIRIT, both clinical and sociodemographic characteristics contribute to the complexity of the study population. PTSD and bipolar disorder cause more disability than nearly all other health conditions worldwide [59]. Moreover, rural Americans with PTSD and bipolar disorder lack access to effective treatments because of severe shortages and geographic maldistribution of mental health specialists [60,61]. The intersection of high clinical complexity and high unmet need yields high potential for public health impact of mHealth.

Add Value for All Users

This is an extension of *Principle 1: Design with the User* and is especially relevant for technology-enabled health care services where technology platforms are used by both patients and providers and thus must meet needs for each group of users simultaneously. Most mobile mental health tools are stand-alone tools that aspire to offer value to consumers or patients (primary users) but do not necessarily offer value to clinicians. In contrast, those that are used to augment clinical services, such as the SPIRIT app, also need to provide value to clinicians (secondary users) by performing useful clinical functions. Conversely, extraneous features that lack clear benefit to users should be eliminated or minimized to limit user burden [11,62]. We aligned the SPIRIT app with an effective clinical care model, Collaborative Care, and focused on those functions that support such care [40]. This emphasis was important in defining and narrowing the scope of the app. For example, although efforts to use passive sensing to infer mental health symptoms are emerging, the validity of most measures is not established. Some patients do not want sensor-based data, and it remains unclear how clinicians would use this information to deliver

measurement-based care [63]. Therefore, we elected not to include sensor data collection in the scope.

Before deploying the SPIRIT app, we envisioned the SPIRIT app for patients' use in the community. Subsequently, we received feedback from SPIRIT care managers interested in using the app on a tablet in the waiting room so that patients enter their own symptom scales before appointments. The SPIRIT providers recognized potential value from using the tool in a different workflow than we had anticipated. To provide additional value to care managers, who are key to adoption of the SPIRIT app, we are working on developing necessary workflows and adaptations.

Test the Product and the Process

We approached quality assurance testing from a technical perspective and focused on specific functions of the SPIRIT app and CMTS linkages. This testing ensured that the registration process, the app, and the data transfer to CMTS were reliable and robust. However, our predeployment testing did not emulate the typical Collaborative Care workflow in which a provider manages a panel of patients and therefore did not serve to pilot test the overall technology-enabled workflow. As such, our testing process was more focused on the technology product than the technology-enabled service process [9]. A more comprehensive testing plan and/or a deployment time line that planned for a soft pilot launch and additional app refinement period would have allowed us to identify improvements in the technology-supported clinical workflow before full-scale launch. For future deployments, we recommend expanding the scope of testing to mirror practice by employing clinicians as testers before deployment during user-acceptance testing and using case-based example patient scenarios, simulation, or role-plays. Such workflow process testing should be accounted for as a separate step in project timelines and budgets for efforts to develop technology-enabled clinical services.

Acknowledge Disruption

When new technologies are deployed in new clinical settings, some disruption in routines is inevitable, whether or not this is intended. Although individual providers may need to alter their behaviors, implementation is even more challenging when multiple members of the health care team are affected because a new technology requires the development of new clinical workflows [13,14,52]. New technologies can disrupt existing organizational routines and relationships, requiring teams to relearn how to work together [52]. Anticipating, acknowledging, and planning for such disruption can help systems to succeed in efforts to adopt new digital technologies. SPIRIT care managers were already learning new workflows and skills to provide care management and behavioral interventions. We sought to minimize additional burden on care managers as they developed competence working with the SPIRIT app. We did not expect care managers to be experts in the technical details of the system. However, as the front-line clinicians interacting with patients, they have been fielding questions about the SPIRIT app. We developed training materials and resources, including sample scripts, for care managers that were easily accessible to equip them for this new responsibility.

Anticipate Variability

Some digital health tools may work better in certain clinical settings, and successful adoption of a new digital health technology in one setting may not generalize to other settings. The complex interaction between a new digital technology and the dynamics of the health care team is a key determinant of the success of deployment and may be more important than the features of the technology itself [52]. In SPIRIT, we have 30 clinics from 12 organizations with substantial variability in the qualifications and experience of provider teams, the clinical workflows, the sites' fidelity to the Collaborative Care clinical model, and the volume and sociodemographic characteristics of patients. All five of the care managers with patients using the SPIRIT app practice in sites that are implementing Collaborative Care well. We anticipate that variability in the interaction between the technology and the dynamics of care teams across sites may lead to greater success with deployment in some clinical contexts than others [52]. This naturally occurring variation provides an additional opportunity for team learning to understand how best to facilitate adoption. Settings also vary in the size and clinical characteristics of the patient populations they serve. Because the practice change required to learn a new system is difficult, this initial hurdle may be overcome more readily if providers believe that the new technology will benefit a majority of their patients and have a "critical mass" of eligible patients. In SPIRIT, this presents a challenge in some small rural clinics serving a handful of patients with PTSD and/or bipolar disorder. We anticipate the need for ongoing efforts to support and promote use of the SPIRIT app in settings with low patient volume or high provider turnover. Integration of mHealth tools into clinical practice may be more successful when a single system can be used to support care for a variety of health conditions or for patients who have multiple chronic illnesses rather than having separate apps for each condition. Tools that support multi-condition management are likewise more patient-centered.

Conclusions

Adhering to the Principles for Digital Development, we created and deployed an mHealth system to support Collaborative Care for patients with complex psychiatric conditions in rural health centers. Early feedback from patients indicates that the SPIRIT app earns high scores for usability among active users. A majority of the initial patients using the SPIRIT app are self-monitoring symptoms on a biweekly basis over a period of several months, which are intervals that are clinically meaningful for supporting measurement-based care. Our findings come from a pragmatic clinical trial designed to reflect real-world clinical practice, and therefore, these patterns occurred with limited involvement of the research team. Our examples of how the Principles for Digital Development applied to the SPIRIT app can serve as a model for clinical researchers and mobile developers. We propose to extend this framework with five additional principles: design for public health impact, add value for all users, test the product and the process, acknowledge disruption, and anticipate variability. These principles can inform future efforts to improve health care quality and outcomes by integrating mobile tools into clinical care pathways in rural practice settings.

Acknowledgments

This research was funded by a Patient-Centered Outcomes Research Institute (PCORI) Award (PCS-1406-19295), The John A. Hartford Foundation (Grant No. 2015-0077: Social Innovation Fund IMPACT Expansion), and the National Institutes of Health (NIH; NCATS KL2 TR000421). The statements are solely the responsibility of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee, The John A. Hartford Foundation, or the NIH. The authors would like to thank the SPIRIT Consumer Advisory Board and the care managers and patients who have shared their valuable insights with them. They would also like to thank Lori Ferro, MHA and Lindsay Baldwin, BA for their assistance.

Conflicts of Interest

SH was previously employed by Dimagi Inc, the developers of the CommCare platform. None of the other authors have any conflicts of interest to disclose.

Multimedia Appendix 1

Examples of specific changes to the Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT) mobile platform based on end-user feedback.

[[PDF File \(Adobe PDF File\). 39KB - jmir_v20i6e10048_app1.pdf](#)]

References

1. Research2guidance. 2013. Mobile health market report 2013-2017 URL: <http://www.giiresearch.com/report/r2g228278-mobile-health-market-report-2011-2016.html> [accessed 2017-08-30] [WebCite Cache ID 6t6jqJjJR]
2. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. *JAMA Psychiatry* 2017 May 01;74(5):437-438. [doi: [10.1001/jamapsychiatry.2017.0262](https://doi.org/10.1001/jamapsychiatry.2017.0262)] [Medline: [28384700](https://pubmed.ncbi.nlm.nih.gov/28384700/)]
3. 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](https://doi.org/10.1080/16506070903318960)] [Medline: [20183695](https://pubmed.ncbi.nlm.nih.gov/20183695/)]
4. 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](https://doi.org/10.2196/jmir.1194)] [Medline: [19403466](https://pubmed.ncbi.nlm.nih.gov/19403466/)]
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](https://doi.org/10.2196/jmir.1602)] [Medline: [21393123](https://pubmed.ncbi.nlm.nih.gov/21393123/)]
6. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. *J Med Internet Res* 2016 Dec 20;18(12):e330 [FREE Full text] [doi: [10.2196/jmir.6482](https://doi.org/10.2196/jmir.6482)] [Medline: [27998876](https://pubmed.ncbi.nlm.nih.gov/27998876/)]
7. Gilbody S, Brabyn S, Lovell K, Kessler D, Devlin T, Smith L, REEACT collaborative. Telephone-supported computerised cognitive-behavioural therapy: REEACT-2 large-scale pragmatic randomised controlled trial. *Br J Psychiatry* 2017 May;210(5):362-367. [doi: [10.1192/bjp.bp.116.192435](https://doi.org/10.1192/bjp.bp.116.192435)] [Medline: [28254959](https://pubmed.ncbi.nlm.nih.gov/28254959/)]
8. Bauer AM, Iles-Shih M, Ghomi RH, Rue T, Grover T, Kincler N, et al. Acceptability of mhealth augmentation of collaborative care: a mixed methods pilot study. *Gen Hosp Psychiatry* 2018 Apr 24;51:22-29. [doi: [10.1016/j.genhosppsych.2017.11.010](https://doi.org/10.1016/j.genhosppsych.2017.11.010)] [Medline: [29272712](https://pubmed.ncbi.nlm.nih.gov/29272712/)]
9. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. *Psychiatr Serv* 2017 May 01;68(5):427-429. [doi: [10.1176/appi.ps.201600541](https://doi.org/10.1176/appi.ps.201600541)] [Medline: [28412890](https://pubmed.ncbi.nlm.nih.gov/28412890/)]
10. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. *JAMA Psychiatry* 2018 Feb 01;75(2):113-114. [doi: [10.1001/jamapsychiatry.2017.3838](https://doi.org/10.1001/jamapsychiatry.2017.3838)] [Medline: [29238805](https://pubmed.ncbi.nlm.nih.gov/29238805/)]
11. Heffernan KJ, Chang S, Maclean ST, Callegari ET, Garland SM, Reavley NJ, et al. Guidelines and recommendations for developing interactive ehealth apps for complex messaging in health promotion. *JMIR Mhealth Uhealth* 2016;4(1):e14 [FREE Full text] [doi: [10.2196/mhealth.4423](https://doi.org/10.2196/mhealth.4423)] [Medline: [26860623](https://pubmed.ncbi.nlm.nih.gov/26860623/)]
12. Hallgren KA, Bauer AM, Atkins DC. Digital technology and clinical decision making in depression treatment: current findings and future opportunities. *Depress Anxiety* 2017 Jun;34(6):494-501. [doi: [10.1002/da.22640](https://doi.org/10.1002/da.22640)] [Medline: [28453916](https://pubmed.ncbi.nlm.nih.gov/28453916/)]
13. Holahan PJ, Aronson ZH, Jurkat M, Schoorman F. Implementing computer technology: a multiorganizational test of Klein and Sorra's model. *JET-M* 2004 Mar;21(1-2):31-50. [doi: [10.1016/j.jengtecman.2003.12.003](https://doi.org/10.1016/j.jengtecman.2003.12.003)]
14. Lluch M. Healthcare professionals' organisational barriers to health information technologies-a literature review. *Int J Med Inform* 2011 Dec;80(12):849-862. [doi: [10.1016/j.ijmedinf.2011.09.005](https://doi.org/10.1016/j.ijmedinf.2011.09.005)] [Medline: [22000677](https://pubmed.ncbi.nlm.nih.gov/22000677/)]
15. DesRoches CM, Charles D, Furukawa MF, Joshi MS, Kralovec P, Mostashari F, et al. Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012. *Health Aff (Millwood)* 2013 Aug;32(8):1478-1485. [doi: [10.1377/hlthaff.2013.0308](https://doi.org/10.1377/hlthaff.2013.0308)] [Medline: [23840052](https://pubmed.ncbi.nlm.nih.gov/23840052/)]

16. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol* 2009 May;62(5):499-505. [doi: [10.1016/j.jclinepi.2009.01.012](https://doi.org/10.1016/j.jclinepi.2009.01.012)] [Medline: [19348976](#)]
17. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009 May;62(5):464-475. [doi: [10.1016/j.jclinepi.2008.12.011](https://doi.org/10.1016/j.jclinepi.2008.12.011)] [Medline: [19348971](#)]
18. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015 May 08;350:h2147. [Medline: [25956159](#)]
19. Waugaman A. UNICEF Innovation. Washington, DC: The Principles for Digital Development Working Group; 2016. From principle to practice: implementing the principles for digital development URL: http://www.unicefstories.org/wp-content/uploads/2013/08/From_Principle_to_Practice.pdf [WebCite Cache ID 6x2V8oliL]
20. Digital Development Principles Working Group. URL: <https://digitalprinciples.org/> [WebCite Cache ID 6wBAnR86P]
21. Archer J, Bower P, Gilbody S, Lovell K, Richards D, Gask L, et al. Collaborative care for depression and anxiety problems. *Cochrane Database Syst Rev* 2012 Oct 17(10):CD006525. [doi: [10.1002/14651858.CD006525.pub2](https://doi.org/10.1002/14651858.CD006525.pub2)] [Medline: [23076925](#)]
22. Thota AB, Sipe TA, Byard GJ, Zometa CS, Hahn RA, McKnight-Eily LR, Community Preventive Services Task Force. Collaborative care to improve the management of depressive disorders: a community guide systematic review and meta-analysis. *Am J Prev Med* 2012 May;42(5):525-538. [doi: [10.1016/j.amepre.2012.01.019](https://doi.org/10.1016/j.amepre.2012.01.019)] [Medline: [22516495](#)]
23. Unützer J, Katon W, Callahan CM, Williams JW, Hunkeler E, Harpole L, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 2002 Dec 11;288(22):2836-2845. [Medline: [12472325](#)]
24. Katon WJ, Unützer J. Health reform and the affordable care act: the importance of mental health treatment to achieving the triple aim. *J Psychosom Res* 2013 Jun;74(6):533-537 [FREE Full text] [doi: [10.1016/j.jpsychores.2013.04.005](https://doi.org/10.1016/j.jpsychores.2013.04.005)] [Medline: [23731753](#)]
25. New Freedom Commission on Mental Health. Achieving the Promise: Transforming Mental Health Care in America. Rockville, MD: DHHS Pub. No. SMA-03-3832; 2003.
26. U.S. Department of Health and Human Services. Mental Health: A Report of the Surgeon General. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, National Institute of Mental Health; 1999.
27. 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](#)]
28. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A Tipping Point for Measurement-Based Care. *Psychiatr Serv* 2017 Feb 01;68(2):179-188. [doi: [10.1176/appi.ps.201500439](https://doi.org/10.1176/appi.ps.201500439)] [Medline: [27582237](#)]
29. American Psychiatric Association and Academy of Psychosomatic Medicine. 2016. Dissemination of integrated care within adult primary care settings: the collaborative care model URL: <https://www.psychiatry.org/psychiatrists/practice/professional-interests/integrated-care/get-trained/about-collaborative-care> [WebCite Cache ID 6wB9vxVev]
30. Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RM. The national depressive and manic-depressive association (DMDA) survey of bipolar members. *J Affect Disord* 1994 Aug;31(4):281-294. [Medline: [7989643](#)]
31. Magruder KM, Frueh BC, Knapp RG, Johnson MR, Vaughan JA, Carson TC, et al. PTSD symptoms, demographic characteristics, and functional status among veterans treated in VA primary care clinics. *J Trauma Stress* 2004 Aug;17(4):293-301. [doi: [10.1023/B:JOTS.0000038477.47249.c8](https://doi.org/10.1023/B:JOTS.0000038477.47249.c8)] [Medline: [15462536](#)]
32. Das AK, Olfson M, Gameroff MJ, Pilowsky DJ, Blanco C, Feder A, et al. Screening for bipolar disorder in a primary care practice. *JAMA* 2005 Feb 23;293(8):956-963. [doi: [10.1001/jama.293.8.956](https://doi.org/10.1001/jama.293.8.956)] [Medline: [15728166](#)]
33. Olfson M, Das AK, Gameroff MJ, Pilowsky D, Feder A, Gross R, et al. Bipolar depression in a low-income primary care clinic. *Am J Psychiatry* 2005 Nov;162(11):2146-2151. [doi: [10.1176/appi.ajp.162.11.2146](https://doi.org/10.1176/appi.ajp.162.11.2146)] [Medline: [16263856](#)]
34. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the national comorbidity survey replication. *Arch Gen Psychiatry* 2005 Jun;62(6):629-640. [doi: [10.1001/archpsyc.62.6.629](https://doi.org/10.1001/archpsyc.62.6.629)] [Medline: [15939840](#)]
35. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the national comorbidity survey replication. *Arch Gen Psychiatry* 2005 Jun;62(6):603-613. [doi: [10.1001/archpsyc.62.6.603](https://doi.org/10.1001/archpsyc.62.6.603)] [Medline: [15939838](#)]
36. Liebschutz J, Saitz R, Brower V, Keane TM, Lloyd-Travaglini C, Averbuch T, et al. PTSD in urban primary care: high prevalence and low physician recognition. *J Gen Intern Med* 2007 Jun;22(6):719-726 [FREE Full text] [doi: [10.1007/s11606-007-0161-0](https://doi.org/10.1007/s11606-007-0161-0)] [Medline: [17503105](#)]
37. Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the national comorbidity survey replication. *Arch Gen Psychiatry* 2007 May;64(5):543-552 [FREE Full text] [doi: [10.1001/archpsyc.64.5.543](https://doi.org/10.1001/archpsyc.64.5.543)] [Medline: [17485606](#)]
38. Stein MB, Roy-Byrne PP, Craske MG, Campbell-Sills L, Lang AJ, Golinelli D, et al. Quality of and patient satisfaction with primary health care for anxiety disorders. *J Clin Psychiatry* 2011 Jul;72(7):970-976 [FREE Full text] [doi: [10.4088/JCP.09m05626blu](https://doi.org/10.4088/JCP.09m05626blu)] [Medline: [21367351](#)]

39. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry* 2008 Dec;69(12):1916-1919. [Medline: [19192467](#)]
40. Bauer AM, Thielke SM, Katon W, Unützer J, Areán P. Aligning health information technologies with effective service delivery models to improve chronic disease care. *Prev Med* 2014 Sep;66:167-172. [doi: [10.1016/j.ypmed.2014.06.017](#)] [Medline: [24963895](#)]
41. Unützer J, Choi Y, Cook IA, Oishi S. A web-based data management system to improve care for depression in a multicenter clinical trial. *Psychiatr Serv* 2002 Jun;53(6):671-3, 678. [doi: [10.1176/ps.53.6.671](#)] [Medline: [12045303](#)]
42. Dimagi Inc. 2016. The commcare evidence base URL: <http://www.dimagi.com/wp-content/uploads/2016/06/CommCare-Evidence-Base-July-2016.pdf?submissionGuid=fb1be3e6-5156-4636-a1ad-97f7a53995d6> [WebCite Cache ID 6x5HpImp1]
43. Bauer AM, Hodsdon S, Hunter S, Choi Y, Bechtel J, Fortney J. Lessons from the Deployment of the SPIRIT App to Support Collaborative Care for Rural Patients with Complex Psychiatric Conditions. In: 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers.: ACM; 2017 Presented at: p. 772-780. [doi: [10.1145/3123024.3125610](#)]
44. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. : ACM; 1993 Presented at: INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; April 24-29, 1993; Amsterdam, The Netherlands. [doi: [10.1145/169059.169166](#)]
45. Nielsen J. Nielsen Norman Group. 2000. Why you only need to test with 5 users URL: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [WebCite Cache ID 6x2VCyASb]
46. Brooke J. SUS - a quick and dirty usability scale. URL: <http://www.usabilitynet.org/trump/documents/Suschart.doc> [WebCite Cache ID 6wwVmabMY]
47. Sauro J. Measuring usability with the system usability scale (SUS). 2011. URL: <https://measuringu.com/sus/> [WebCite Cache ID 6wwVqrYA2]
48. IDEO. DesignKit: what is human-centered design? URL: <http://www.designkit.org/human-centered-design> [WebCite Cache ID 6wBAOoQHE]
49. Allen M, Leung R, McGrenere J, Purves B. Involving domain experts in assistive technology research. *Univ Access Inf Soc* 2008 Jan 23;7(3):145-154. [doi: [10.1007/s10209-008-0112-5](#)]
50. Pew Research Center. 2015. The smartphone difference URL: <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/> [WebCite Cache ID 6YJ3IPbC4]
51. University of Washington AIMS Center. Patient-centered integrated behavioral health care principles & tasks URL: <http://aims.uw.edu/collaborative-care/principles-collaborative-care> [accessed 2018-02-06] [WebCite Cache ID 6x2VIFtz0]
52. Edmondson A. Framing for learning: lessons in successful technology implementation. *Calif Manag Rev* 2003 Jan;45(2):34-54. [doi: [10.2307/41166164](#)]
53. Ben-Zeev D, Atkins DC. Bringing digital mental health to where it is needed most. *Nat Hum Behav* 2017 Oct 23;1(12):849-851. [doi: [10.1038/s41562-017-0232-0](#)]
54. Wells K, Sherbourne C, Schoenbaum M, Ettner S, Duan N, Miranda J, et al. Five-year impact of quality improvement for depression: results of a group-level randomized controlled trial. *Arch Gen Psychiatry* 2004 Apr;61(4):378-386. [doi: [10.1001/archpsyc.61.4.378](#)] [Medline: [15066896](#)]
55. Areán PA, Ayalon L, Hunkeler E, Lin EH, Tang L, Harpole L, IMPACT Investigators. Improving depression care for older, minority patients in primary care. *Med Care* 2005 Apr;43(4):381-390. [Medline: [15778641](#)]
56. Davis TD, Deen T, Bryant-Bedell K, Tate V, Fortney J. Does minority racial-ethnic status moderate outcomes of collaborative care for depression? *Psychiatr Serv* 2011 Nov;62(11):1282-1288. [doi: [10.1176/ps.62.11.pss6211_1282](#)] [Medline: [22211206](#)]
57. Katon W, Russo J, Reed SD, Croicu CA, Ludman E, LaRocco A, et al. A randomized trial of collaborative depression care in obstetrics and gynecology clinics: socioeconomic disadvantage and treatment response. *Am J Psychiatry* 2015 Jan;172(1):32-40 [FREE Full text] [doi: [10.1176/appi.ajp.2014.14020258](#)] [Medline: [25157500](#)]
58. Grote NK, Katon WJ, Russo JE, Lohr MJ, Curran M, Galvin E, et al. A randomized trial of collaborative care for perinatal depression in socioeconomically disadvantaged women: the impact of comorbid posttraumatic stress disorder. *J Clin Psychiatry* 2016 Nov;77(11):1527-1537. [doi: [10.4088/JCP.15m10477](#)] [Medline: [28076671](#)]
59. Alonso J, Petukhova M, Vilagut G, Chatterji S, Heeringa S, Üstün TB, et al. Days out of role due to common physical and mental conditions: results from the WHO World Mental Health surveys. *Mol Psychiatry* 2011 Dec;16(12):1234-1246 [FREE Full text] [doi: [10.1038/mp.2010.101](#)] [Medline: [20938433](#)]
60. Ellis AR, Konrad TR, Thomas KC, Morrissey JP. County-level estimates of mental health professional supply in the United States. *Psychiatr Serv* 2009 Oct;60(10):1315-1322. [doi: [10.1176/ps.2009.60.10.1315](#)] [Medline: [19797370](#)]
61. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. *Psychiatr Serv* 2009 Oct;60(10):1323-1328. [doi: [10.1176/ps.2009.60.10.1323](#)] [Medline: [19797371](#)]
62. Suh H, Shahriaree N, Hekler EB, Kientz JA. Developing and validating the user burden scale: a tool for assessing user burden in computing systems. : ACM; 2016 Presented at: CHI Conference on Human Factors in Computing Systems; May 07-12, 2016; San Jose, CA. [doi: [10.1145/2858036.2858448](#)]

63. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *J Med Internet Res* 2013;15(4):e86 [FREE Full text] [doi: [10.2196/jmir.2583](https://doi.org/10.2196/jmir.2583)] [Medline: [23598614](https://pubmed.ncbi.nlm.nih.gov/23598614/)]

Abbreviations

CAB: Consumer Advisory Board
CMTS: Care Management Tracking System
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
PCL-5: PTSD Checklist for DSM-5
PHQ-9: Patient Health Questionnaire-9
PTSD: posttraumatic stress disorder
SaaS: Software-as-a-Service
SMS: short message service
SPIRIT: Study to Promote Innovation in Rural Telepsychiatry
SUS: System Usability Scale
UI: user interface

Edited by J Torous, G Wadley, R Calvo, M Czerwinski; submitted 09.02.18; peer-reviewed by T Ungar, M Crichton; comments to author 12.03.18; revised version received 22.03.18; accepted 22.03.18; published 06.06.18.

Please cite as:

Bauer AM, Hodsdon S, Bechtel JM, Fortney JC

Applying the Principles for Digital Development: Case Study of a Smartphone App to Support Collaborative Care for Rural Patients With Posttraumatic Stress Disorder or Bipolar Disorder

J Med Internet Res 2018;20(6):e10048

URL: <http://www.jmir.org/2018/6/e10048/>

doi: [10.2196/10048](https://doi.org/10.2196/10048)

PMID:

©Amy M Bauer, Sarah Hodsdon, Jared M Bechtel, John C Fortney. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.06.2018. 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.

Original Paper

User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews

Katarzyna Stawarz^{1*}, PhD; Chris Preist^{1*}, PhD; Debbie Tallon², BSc, MSc; Nicola Wiles^{2,3}, PhD; David Coyle^{4*}, PhD

¹Bristol Interaction Group, Faculty of Engineering, University of Bristol, Bristol, United Kingdom

²Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

³National Institute for Health Research Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, United Kingdom

⁴School of Computer Science, University College Dublin, Dublin, Ireland

*these authors contributed equally

Corresponding Author:

Katarzyna Stawarz, PhD
Bristol Interaction Group
Faculty of Engineering
University of Bristol
Merchant Venturers Building
Woodland Road
Bristol, BS8 1UB
United Kingdom
Phone: 44 1179545289
Email: k.stawarz@bristol.ac.uk

Abstract

Background: Hundreds of mental health apps are available to the general public. With increasing pressures on health care systems, they offer a potential way for people to support their mental health and well-being. However, although many are highly rated by users, few are evidence-based. Equally, our understanding of what makes apps engaging and valuable to users is limited.

Objective: The aim of this paper was to analyze functionality and user opinions of mobile apps purporting to support cognitive behavioral therapy for depression and to explore key factors that have an impact on user experience and support engagement.

Methods: We systematically identified apps described as being based on cognitive behavioral therapy for depression. We then conducted 2 studies. In the first, we analyzed the therapeutic functionality of apps. This corroborated existing work on apps' fidelity to cognitive behavioral therapy theory, but we also extended prior work by examining features designed to support user engagement. Engagement features found in cognitive behavioral therapy apps for depression were compared with those found in a larger group of apps that support mental well-being in a more general sense. Our second study involved a more detailed examination of user experience, through a thematic analysis of publicly available user reviews of cognitive behavioral therapy apps for depression.

Results: We identified 31 apps that purport to be based on cognitive behavioral therapy for depression. Functionality analysis (study 1) showed that they offered an eclectic mix of features, including many not based on cognitive behavioral therapy practice. Cognitive behavioral therapy apps used less varied engagement features compared with 253 other mental well-being apps. The analysis of 1287 user reviews of cognitive behavioral therapy apps for depression (study 2) showed that apps are used in a wide range of contexts, both replacing and augmenting therapy, and allowing users to play an active role in supporting their mental health and well-being. Users, including health professionals, valued and used apps that incorporated both core cognitive behavioral therapy and non-cognitive behavioral therapy elements, but concerns were also expressed regarding the unsupervised use of apps. Positivity was seen as important to engagement, for example, in the context of automatic thoughts, users expressed a preference to capture not just negative but also positive ones. Privacy, security, and trust were crucial to the user experience.

Conclusions: Cognitive behavioral therapy apps for depression need to improve with respect to incorporating evidence-based cognitive behavioral therapy elements. Equally, a positive user experience is dependent on other design factors, including consideration of varying contexts of use. App designers should be able to clearly identify the therapeutic basis of their apps, but they should also draw on evidence-based strategies to support a positive and engaging user experience. The most effective apps

are likely to strike a balance between evidence-based cognitive behavioral therapy strategies and evidence-based design strategies, including the possibility of eclectic therapeutic techniques.

(*J Med Internet Res* 2018;20(6):e10120) doi:[10.2196/10120](https://doi.org/10.2196/10120)

KEYWORDS

mental health; mobile apps; cognitive behavioral therapy; depression; user experience; mHealth

Introduction

Background

Mental health difficulties are a leading cause of disability worldwide [1,2], with depression alone affecting 98.7 million people [2]. Responding to the urgent need to provide more people with access to effective treatments, substantial research has been undertaken on the use of technology to increase access to mental health treatment [3-8]. Much of this work has focused on the development and evaluation of computerized cognitive behavioral therapy (CBT) [5,8-13]. CBT incorporates both behavioral and cognitive aspects and provides a structured approach for recognizing and addressing negative thinking patterns and underlying beliefs [14]. Due to this structured approach, it lends itself well to being adapted to computerized platforms, both as self-directed [9] and therapist-guided [10] treatment.

More recently, mobile apps have provided an alternative to computerized CBT interventions. With 76% of UK and 81% of US adults owning a smartphone [15,16], there is a strong argument for the potential of apps to help in providing flexible access to mental health support [17,18]. Studies have been conducted focusing on the development of mobile apps to support mental health [19-28], and more detailed reviews and analyses of existing research are also available [4,29-31]. This work indicates the potential of appropriately designed apps and could drive future innovation in mental health apps to ultimately deliver large-scale impact on public health. However, given the openness of app stores to developers [32,33], challenges with regulating health apps [34], and the time it typically takes for evidence-based research to make its way into health care practice [35], it is unsurprising that current research is not always reflected in the apps available in app stores.

Recently, several papers have reviewed apps with the aim of assessing the extent to which they are grounded in theory, especially CBT [36,37], or to evaluate the extent of expert involvement [38,39]. They suggest that current apps tend to lack an evidence base [31,38] and often combine evidence-based features with other approaches not supported by research [37]. Furthermore, there does not seem to be any correlation between apps' ratings and popularity and the presence of evidence-based features [37].

We agree that the lack of an evidence base in publicly available apps is a significant cause for concern. However, high ratings of apps defined as inconsistent with evidence [37] suggest that they might be important to users. The existence of these apps provides an important opportunity to investigate and understand factors that facilitate user engagement with mental health apps. Through app reviews submitted to app stores, people using these

apps have provided a large body of data regarding their user experiences, context of use, and features they value. Previous work within the human-computer interaction (HCI) community has demonstrated the benefits of using public reviews to investigate user attitudes toward and experiences of existing apps [40-44]. Researchers have also analyzed user reviews of mood-tracking apps [44], looked at general use of health apps [45], or the types of health apps people with depression use [46]. Alongside efficacy, user experience and engagement are critical factors to the overall effectiveness of mental health technologies [47]; therefore, it is important to investigate what it is that the users themselves value.

Objectives

This paper has 2 key aims: first, to systematically analyze the therapeutic elements and engagement approaches used in apps described as being based on CBT for depression; and second, to analyze publicly available user reviews of these apps to provide a more detailed understanding of the user experience and of what makes apps engaging and valuable to users.

In recent years, a number of important approaches have emerged for examining mobile health apps. For example, the Mobile App Rating Scale developed by Stoyanov et al provides a tool for assessing the quality of mobile health apps [48]. It includes a dedicated section on engagement that allows an assessor to score individual apps based on key engagement features. Chan et al provide a framework specifically for evaluating mobile mental health apps [49] that allows patients and mental health service providers to evaluate apps by their usefulness, usability, and integration with infrastructure. We view our approach as complementary, but distinct from these approaches.

Instead of providing detailed analysis of individual apps, we focused on the thematic analysis and synthesis of user perspectives on engagement and therapeutic features across a range of apps. Analyzing app reviews can provide insight on what end users find engaging in general and how the apps are used, and identify mismatches between what researchers believe to be important and what users actually find engaging. It can also help us to better understand why current apps are highly rated and leverage this understanding in the design of compelling, evidence-based apps going forward. As an approach, it can also offer distinct insights into recent app reviews that have incorporated traditional usability evaluations. For example, Huguet et al [36] applied Nielsen's expert-led heuristic evaluation approach to assess the usability of CBT apps for depression. While usability helps to assess the degree to which users can easily—or with minimal training—use and understand the app, our approach offers complementary insights on *how* and *why* people use apps and the particular features which they find engaging or unengaging.

Methods

Data Collection

Our initial analysis of apps aimed to identify 2 key groups:

1. CBT apps for depression: apps that self-identify as implementing CBT to target depression.
2. Mental well-being apps: a broader group of apps that not only includes apps that aim to address mental health problems such as anxiety and depression, but also issues such as stress, worry, mood, or emotional well-being.

To limit the scope of our study, we deliberately excluded apps targeting more severe disorders such as bipolar disorder or less common disorders such as obsessive-compulsive disorder.

A detailed analysis of CBT apps for depression is the core focus of this paper. We consider both engagement features and therapeutic features (study 1) and user opinions (study 2). The set of apps addressing mental well-being more generally is not subjected to the same level of scrutiny as it is too large and too diverse in terms of therapeutic approaches applied. For these apps, we analyze engagement features only to allow a comparison with engagement approaches found in CBT apps for depression.

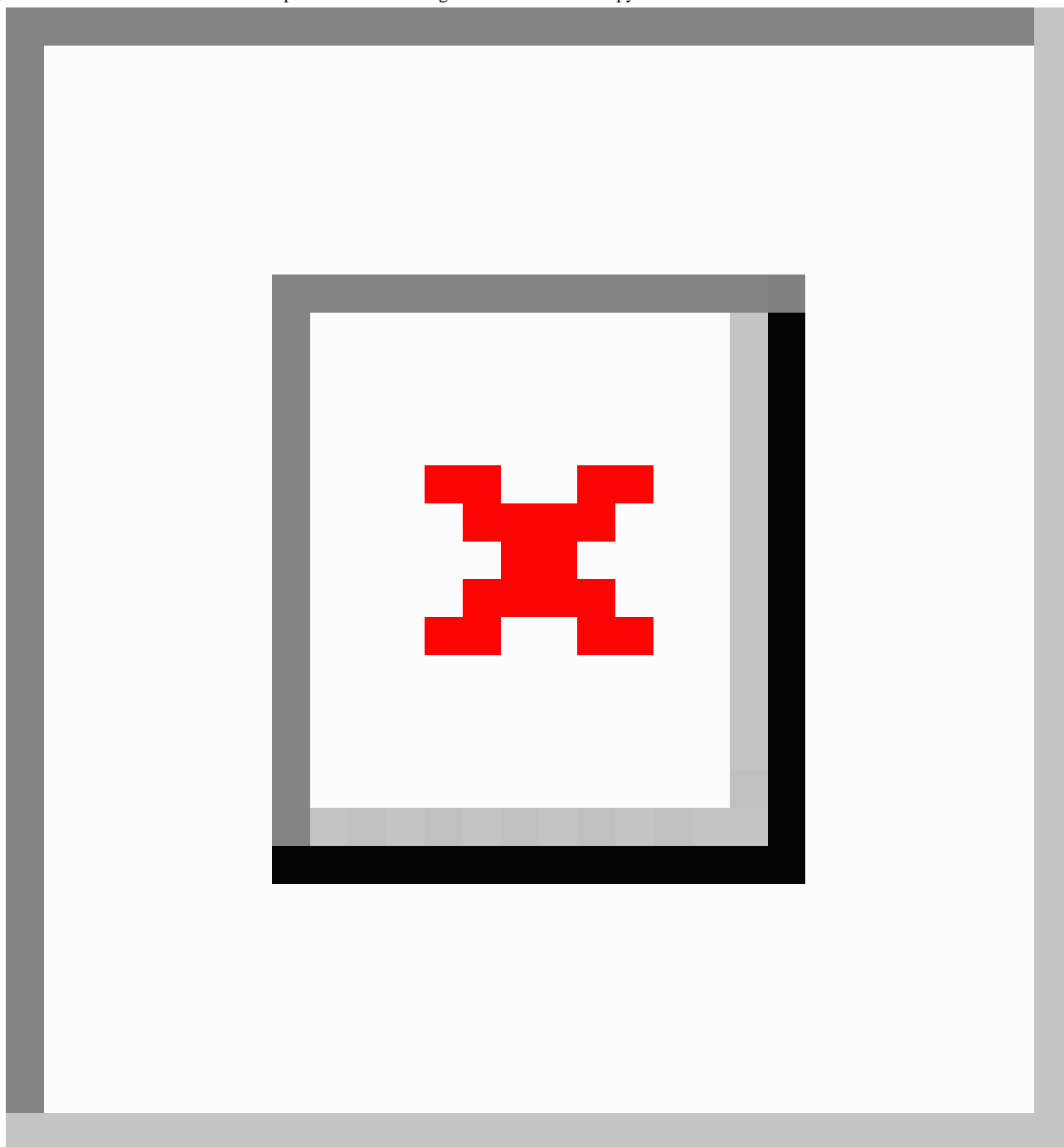
Figure 1 illustrates the overall systematic process used to identify relevant apps.

Phase 1: Initial Keyword Search and Data Clean Up

We first defined the following groups of keywords:

1. General keywords related to mental health and well-being: “mental health,” “mental wellbeing,” “emotional wellbeing,” depression, anxiety, stress, mood, and emotions.
2. Keywords related to CBT: “cognitive behavioral therapy,” “cognitive behavioural therapy,” and CBT.
3. Keywords related to aspects of CBT: “activity diary,” “thought record,” “behavioral activation,” “behavioural activation,” “negative thoughts,” “core beliefs,” and “cognitive restructuring.”

We used scripts [50,51] to automatically download search results for each of these keywords separately from the UK version of Google Play and Apple’s App Store. The searches took place in January 2017. Recorded information included each app’s name, its short description (if available), detailed description, price, average rating, number of user ratings, developer’s details, and app store category. This resulted in 3954 apps (2316 apps from Google Play and 1638 from App Store).

Figure 1. Data extraction and exclusion procedures. CBT: cognitive behavioral therapy.

We then used a custom script to combine the search results from each app store and remove duplicates. Finally, we automatically extracted apps belonging to the following app store categories (deemed to include relevant apps based on a manual check using the keyword “depression”): Health and Fitness, Medical, Lifestyle, Education, and Game Educational. We then combined the results and excluded duplicates of apps available for both platforms. At this point, 1680 unique apps remained.

Phase 2: Potentially Relevant Apps Identified

Next, we undertook the first manual screening. Following the approach used in the study by Shen et al [38], we manually reviewed the 1680 apps by examining each app’s title and short description. This allowed us to identify not relevant apps that

would be excluded and potentially relevant apps that would be included in the next phase. We excluded the following types of apps:

1. Apps specifically addressing less common or more severe mental health disorders, for example, substance misuse, OCD; or other health conditions (eg, diabetes, chronic pain).
2. General health tracking apps and single purpose well-being apps, for example, for mindfulness meditation only.
3. Apps to support mental health professionals and students, and apps that require an access code (eg, that are part of a study, insurance plan, employer wellness scheme).
4. Apps not available in English.

A total of 1297 apps were marked as not relevant and further 16 turned out to be duplicates. This resulted in a set of 367 potentially relevant apps.

Phase 3: Apps With Clearly Identified Functionality

Next, we manually reviewed the full descriptions of all 367 potentially relevant apps. During this process, we identified 43 more apps that met the exclusion criteria described above; 2 apps that provided no information about their functionality; 1 app that was available for smartwatches only; and 37 duplicates. At the end of this phase, we had identified 284 apps.

Phase 4: Apps Using Cognitive Behavioral Therapy and Targeting Depression

The aim of this stage was to identify the final subset of apps that self-identified as (1) focusing on depression and (2) based on CBT. This resulted in 31 apps which we classified as *CBT apps for depression* and which we refer to as such throughout the paper. The remaining 253 apps were classified as *mental well-being apps*.

Study 1: Functionality Analysis

The aim of this study was primarily to examine the functionality of the 31 CBT apps for depression. For each app, we recorded both engagement features and therapeutic features. We defined therapeutic features as functionality that aims to help users manage their mental health and well-being, and engagement features as functionality that encourages regular use, makes app content more appealing, and in general helps users to stay engaged with therapy or the app itself.

For each app, we recorded all features listed on its description page (eg, mood tracking, discussion forums, reminders, etc) or visible on screenshots; this approach has also been used in other app reviews [43]. We also noted mentions of expert involvement in app creation (health professionals, researchers, etc). Features were recorded by the first author and regularly reviewed and discussed with others.

Next, to assess whether the features of 31 CBT apps for depression reflect CBT practice, we asked 2 researchers (a clinical psychologist who is also an accredited CBT therapist, and an HCI researcher experienced in designing technologies to support CBT) to independently match them against a recognized CBT competence framework [14]. For each feature, they indicated (“yes” or “no”) whether it represented one of CBT competencies in relation to treatment for depression. Inter-rater agreement was 73% with raters disagreeing regarding 8 items, including 7 items where the disagreement was between a definitive answer (“yes” or “no”) and a “maybe.” Disagreements were resolved through a discussion, and the final categorization is available in Table 1.

Finally, we also undertook a brief analysis of the engagement features available in 253 mental well-being apps. A detailed review of the therapeutic features and approaches applied across this larger group was beyond the scope of this paper.

Study 2: User Reviews Analysis

The analysis of publicly available app reviews has been successfully used in the past to investigate user attitudes toward

existing apps and their feature requests [40–44]. We adopted this method to better understand users’ attitudes toward CBT apps for depression and which features they use and find the most important.

We used scripts [50,51] to automatically download all reviews for the 31 CBT apps for depression. If the app was available for both Android and iOS devices, we downloaded both sets of reviews. In total, we downloaded 2904 reviews of 24 apps (7 apps had 0 reviews). To identify reviews for the analysis, we followed the approach similar to the one used in the study by Stawarz et al [43]—first, the lead author manually assessed all reviews, recording their sentiment (positive, negative, neutral) and whether each mentioned at least 1 therapeutic feature; this was then discussed with other authors. Next, to qualitatively identify underlying themes, we used thematic analysis [52] to analyze the subset of reviews that mentioned at least 1 therapeutic feature. Coding was done by the first author and regularly discussed with others to allow for better familiarization with the data and reduce potential for bias. This iterative process led to codes gradually being merged into broader categories and researchers identifying overarching themes.

Results

Study 1: Functionality Analysis

Overall, within the set of 31 CBT apps for depression, we identified 26 therapeutic features and 10 engagement features; 4 additional engagement features were also available in the broader group of apps. App functionality is described in the following sections and features are summarized in Tables 1 and 2. Detailed information about the 31 CBT apps for depression, including their user ratings, available features, and expert involvement, are summarized in Multimedia Appendix 1.

Therapeutic Features of Cognitive Behavioral Therapy Apps for Depression

The most common features available in the 31 CBT apps for depression focused on dealing with negative automatic thoughts (48%, 15/31 apps) and negative thinking styles (29%, 9/31 apps), and provided examples of activities users could do to improve their mood (29%, 9/31 apps). They also allowed users to record thoughts and emotions (10%, 3/31 apps), schedule daily activities (10%, 3/31 apps), offered challenges and behavioral experiments (6%, 2/31 apps), or enabled goal setting (6%, 2/31 apps). The apps also offered several non-CBT features, including tools for writing and self-reflection, gratitude and affirmations, various tests and scales, or relaxation tracks. All features are summarized in Table 1.

Comparisons With Cognitive Behavioral Therapy Guidelines

Of 26 identified therapeutic features, 50% (13/26 features) reflected elements from the list of CBT competencies [14]. Overall, 90% of CBT apps for depression (28/31 apps) provided at least 1 CBT feature. 39% (12/31 apps) provided only 1 feature, while 29% (9/31 apps) provided only 2 CBT features; among these apps, features for dealing with automatic negative thoughts or negative thinking patterns were the most common.

Moreover, 13% (4/31) of the apps provided 3 CBT features, 1 app (Cloud Clinic) had 4, and 1 app (Depression CBT Self-Help Guide) had 5. Finally, 1 app (MoodTools - Depression Aid) provided 8 CBT features. On the other hand, 10% (3/31) apps did not provide any CBT features at all. Instead, they mainly provided features such as positive self-talk and gratitude (7%, 2/31 apps), depression scales and self-assessment (7%, 2/31 apps), information about mental health and well-being in general (3%, 1/31 apps), and relaxation tracks (3%, 1/31 apps).

We also checked whether the presence of CBT features was associated with app ratings. Ratings were available for 24 apps with the average rating of 4.1 (on a 5-point scale), ranging from 2 for the Activity Diary app (which had 1 CBT feature) to 5 for MoodMaster Anti-Depression App (which had 2). There was no difference in average scores between apps with no CBT features (average score=4.1, min=4, max=4.2, N=3) and apps with at least 1 such feature (average score=4.1, min=2, max=5, N=21). The average rating for MoodTools, the most comprehensive app, was 4.3.

Expert Involvement

For each app, we recorded whether any experts were involved in the development process. Overall, 45% (14/31) of apps mentioned experts on their description page—health professionals (42%, 13/31 apps) and university researchers (7%, 2/31 apps). Apps with expert involvement provided between 0 and 8 CBT features (N=13, mean=2.3, mode=2), whereas the number of CBT features in apps that did not mention experts ranged from 0 to 5 (N=17, mean=1.6, mode=1). Health professionals were involved in the development of MoodTools, the app with the most CBT elements, and were also mentioned in the description of Self-Esteem Blackboard, an app that did not provide any features matching CBT elements. In contrast, 1 app with 3 CBT features (MoodSentry) was built by a patient who wanted to share the methods that worked for him.

There was also no link between the presence of expert involvement and user ratings. For example, MoodTools and What's Up were the most rated apps with over 2000 user ratings each, and both had the average score of 4.3. However, the former offered 8 CBT features and experts were involved in its creation, whereas What's Up offered only 2 such features and no experts were involved. On the other hand, Cloud Clinic received no ratings at all, despite expert involvement and presence of CBT features; both the lowest rated (Activity Diary) and the highest rated (MoodMaster Anti-Depression App) apps mentioned health professionals.

Engagement Features

For all mental health apps we identified, including 31 CBT apps for depression, we recorded details of features that aim to support user engagement with therapy or the app itself. They are summarized in [Table 2](#).

Among the CBT apps for depression, 58% (18/31) of apps provided explicit engagement features. A quarter (26%, 8/31 apps) enabled sharing, including sharing with friends, family, and therapists. Graphs and charts to illustrate progress (13%, 4/31 apps), audio content (10%, 3/31 apps), and notification and prompts (10%, 3/31 apps) were also available. In addition, 2 of 31 apps (7%) offered personalization features and 2 other apps used gamification. Peer support and professional support can also drive engagement [47]; the former was available in 2 out of 31 (7%) and the latter in only 1 of 31 (3%) apps.

Among the broader group of mental well-being apps, 58.9% (149/253) provided engagement features. Graphs and charts were the most common (available in 16.6% (42/253) of apps) and were used to visualize users' mood and progress. Reports and summaries that often accompanied graphs and charts were available in 1.9% (5/253) of apps. In addition, 9.1% of apps (2/253) provided various notifications and prompts—from reminders to record one's mood or interact with the app in some way to weekly summary emails and the ability to set up medication reminders for antidepressants. Interactive content included video (9.5%, 24/253 apps), audio (7.1%, 18/253 apps), or game elements such as cartoon avatars, badges, or progress bars (2.4%, 6/253 apps). Customization and the ability to add own pictures and videos (2.4%, 6/253 apps each) helped to tailor the experience to users' needs. Moreover, 1.2% (3/253) apps also used a Q&A format to make reading materials more engaging and to help users select the right activities. A total of 15 out of 253 apps (5.9%) enabled contact with therapists, including one-to-one chats, forums where health professionals can answer questions, or even the ability to schedule sessions.

Conclusions of Study 1

CBT apps for depression provided a mix of features. Even though all their descriptions mentioned CBT, only half of all features provided by apps reflected core competencies of CBT [14]. Moreover, most of the apps that did offer CBT features provided only 1 or 2 of them, and 3 apps did not provide any such features at all. This limited evidence base of apps is in line with existing research [31,37,38].

Available CBT features tended to be limited, focusing around mood tracking, recording thoughts and emotions, and dealing with negative thoughts. As a result, they often lacked elements of CBT used in high-intensity interventions for depression, such as addressing core beliefs [14]. The presence or absence of features grounded in CBT practice was not linked with expert involvement in app creation, which raises concerns regarding the responsibility of app creators who may be misleading potentially vulnerable users by mentioning CBT without actually providing it. There was also no clear link in terms of expert involvement and user ratings. However, our results highlighted high user ratings for all apps, regardless of whether they provided CBT features, which corroborates previous findings [37].

Table 1. Therapeutic features available in cognitive behavioral therapy (CBT) apps for depression (N=31), including information on whether each feature reflects CBT practice in relation to depression.

Therapeutic features	Definitions	CBT feature	Apps, n (%)
Dealing with negative automatic thoughts	Identifying and challenging specific negative automatic thoughts about self or the world	Yes	15 (48)
Addressing negative thinking styles	Identifying and challenging thinking styles and patterns; for example, catastrophizing, all-or-nothing thinking	Yes	9 (29)
Example activities	Example of pleasurable activities to do to improve one's mood	Yes	9 (29)
Writing and self-reflection	Diaries and journals	No	6 (19)
Tracking mood	Tracking and annotating moods	Yes	5 (16)
Self-assessment	Tests and scales to assess one's well-being	No	5 (16)
Gratitude and affirmations	Gratitude diary, examples of affirmations, ability to add affirmations, questions encouraging positive thinking about self	No	5 (16)
Information about CBT	Articles, blog posts, and other resources explaining cognitive behavioral therapy, its components, and how it works	Yes	4 (13)
Information about depression	Articles, blog posts, videos, and other resources explaining depression, its symptoms, how it works, and how to deal with it	Yes	3 (10)
Recording thoughts and emotions	Recording information about events, and thoughts and emotions that accompany them	Yes	3 (10)
Scheduling activities	Planning activities	Yes	3 (10)
Relaxation tracks	Calming music, sounds of nature, etc	No	3 (10)
General information about well-being	Articles, blog posts, videos, and other resources about mental health in general, health tips, well-being advice, nutrition, etc	No	3 (10)
Tracking anxiety and worries	Tracking anxiety incidents, worry lists	No	3 (10)
Recording and monitoring daily activities	Recording activities, matching activities with the calendar and mood information	Yes	2 (6)
Challenges and behavioral experiments	Tasks to complete to practice (new) coping skills	Yes	2 (6)
Setting goals	Setting up specific goals to work toward	Yes	2 (6)
Peer support	Ability to join forums or social networks, ask questions, and talk to others	No	2 (6)
Suicide prevention	Links to support services, ability to prepare a crisis plan	Yes	1 (3)
Challenging beliefs	Written exercises and examples of tasks to do to address one's beliefs about the world and self	Yes	1 (3)
Breathing exercises	Written or recorded (audio or video) instructions for breathing exercises	No	1 (3)
Mindfulness	Mindfulness meditation tracks and written exercise instructions; excludes other types of meditation	No	1 (3)
Fun content	Games, jokes, and humorous content to provide distractions and improve one's mood	No	1 (3)
Inspirational quotes	Quotes of famous people to provide motivation and lift one's mood	No	1 (3)
Meditation	Guided meditation, topics to contemplate; excludes mindfulness	No	1 (3)
Physical exercise and yoga	Suggestions for specific exercises or yoga sessions	No	1 (3)

Table 2. Engagement features available in cognitive behavioral therapy (CBT) apps for depression (N=31) and other mental well-being apps (N=253).

Engagement features	CBT apps for depression (N=31), n (%)	Mental well-being apps (N=253), n (%)
Ability to share data directly from the app with others	8 (26)	28 (11.1)
Graphs and charts	4 (13)	42 (16.6)
Notifications and reminders	3 (10)	23 (9.1)
Audio content	3 (10)	18 (7.1)
Peer support	2 (7)	21 (8.3)
Customization	2 (7)	6 (2.4)
Games and gamification	2 (7)	6 (2.4)
Video content	1 (3)	24 (9.5)
Treatment program format (modules)	1 (3)	13 (5.1)
Ability to contact a therapist	1 (3)	15 (5.9)
Reports supporting graphs and charts	—	5 (1.9)
Ability to add pictures and videos	—	6 (2.4)
Chat with a bot	—	4 (1.6)
Q&A interface	—	3 (1.2)

Doherty et al [47] discuss interactive features, professional support, peer support, and customization as key strategies to facilitate engagement with therapy. Although a similar proportion of CBT apps for depression provided engagement features compared with the other mental well-being apps (58% vs 58.9%), these features were less varied. Interactive features such as video content, graphs, and charts; a bot interface that allows users to talk with a “virtual” therapist; or ability to add own pictures were more prevalent in the apps for mental well-being, although a bigger proportion of CBT apps offered audio content and gamification. Contact with peers or professionals was almost nonexistent in CBT apps for depression, possibly because peer support is not part of standard CBT and having an open forum would require moderation to reduce potential risks. However, CBT apps for depression offered more customization options, allowing users more options to adapt the features to their needs.

To better understand high app ratings and reasons why people use and value these apps, our second study focused on users’ attitudes toward app features and their experience of using CBT apps for depression.

Study 2: User Reviews Analysis of Cognitive Behavioral Therapy Apps for Depression

Of the initially collected 2904 user reviews, 1287 reviews from 23 apps mentioned at least 1 therapeutic feature and therefore were included in the analysis; 91.61% (1179/1287) of these reviews were positive, 7.38% (95/1287) were negative, and 1.01% (13/1287) were neutral. The thematic analysis uncovered 4 key themes: different contexts of use; importance of privacy, security, and trust; importance of engagement features; and the attitudes toward therapeutic features not related to CBT. The themes are described below.

Context of Use of Cognitive Behavioral Therapy Apps for Depression

Reviews often mentioned context of use and how the apps fit into users’ lives. For 1 group of commenters, having a “pocket therapist” often meant using the app instead of therapy. It was often motivated by not being able to afford therapy or negative experiences in the past. The apps were also seen as simply better than regular therapy:

I just downloaded this for what’s probably obvious reasons—to try and get better, since I can’t afford therapy right now. [What’s Up]

Another user said:

I have had depression for 3 years now and have found very little help from the [National Health Service]. These application from excel at life have helped me to relax and start to help myself. Couldn’t recommend more highly. [Depression CBT Self-Help Guide]

In contrast, another group of commenters used the apps as an adjunct to treatment, alongside visits to a therapist. Users often commented on apps’ usefulness and how well they enhanced their treatment:

Great app. I always email my entries to myself so that I can print it out and share it with my psychiatrist. [Cognitive Diary CBT Self-Help]

In some cases, users were using the app because it was the therapist who suggested that in the first place:

My therapist actually recommended this app and we trialed it in session, where it was really effective. [CBT Thought Record Diary]

Regardless of why people used the apps, they generally appreciated their role in supporting their mental well-being. Comparisons to a “pocket therapist” or “therapy sessions at the tip of their fingers” were frequent:

Love this app, it's like having a therapist in your pocket. [Depression CBT Self-Help Guide]

Reviews were also written by therapists. They showed that sometimes clients were the ones who found the apps and integrated them into their work with the therapist, and at other times, therapists actively recommend apps to their clients:

I have several therapy clients who use this app rather than writing out a thought log. [iCBT]

However, feedback was not always positive. Some users advised caution and warned that the apps should not be used without supervision, or even at all:

Can't replace a therapist, especially when you're first starting this therapy, but by directing the process, this is the next best thing. [Cognitive Diary CBT Self-Help]

Another user mentioned:

Good design, bad idea. Apps shouldn't be diagnosing medical conditions. Consider suggesting the user seek medical help PRIOR to using the app, and that the app should only be used in conjunction with treatment. Not as a precursor to treatment or as a reason to get treatment. [MoodTools]

Importance of Privacy, Security, and Trust

As the apps play an important role in supporting users' mental well-being, knowing that the service was reliable and their data were secure was important. However, this was not always the case—often apps were unreliable and losing data was experienced as “devastating”:

When the app upgraded, [it] erased ALL my examples in my Log. Anyone who practices CBT or has utilized the app [...] knows that this represents hours of work, and is practically irreplaceable. It is particularly bad news in a mental health app. [...] I was devastated about losing my data. [iCouch CBT]

The analysis also revealed the importance of privacy and security, and users often mentioned them alongside therapeutic features. People appreciated the presence of password protection or security locks and demanded these features when they were not available:

Wish [I] could password protect diary. Some of my entries involve loved ones and I do not wish to cause them stress. I hide the app in my phone [and] my entries are written with little detail. Neither is ideal. [MoodTools]

Another user said:

I used to keep a pen and pad as my diary. But people kept reading my personal thoughts and I felt very betrayed. Thanks to this app I can write how I feel in the midst of a situation. AND it has a PASSCODE to keep intruders out. [What's Up]

Users appreciated the discreet nature of the apps, and often compared them with paper worksheets, highlighting the privacy benefits the apps bring:

I like that it's on my phone and therefore it's always in my pocket, so if I'm stressing out over an issue I can pull it out and deal with the stress immediately. To people around me it just looks like I'm checking my phone, playing a game or writing an email. [MoodKit]

In addition, this also meant that the users were more engaged:

One of my biggest struggles [during therapy] was actually doing the work and capturing things as they occurred. Having those tools in my pocket makes it convenient enough that I can do it any time I need to, and so I have been, and have benefited from it. [MoodKit]

Importance of Engagement Features

Sharing was the most common engagement feature among CBT apps for depression and users appreciated the ability to share data with their therapist. Graphs and charts (more common among the wider set of mental well-being apps) and notifications were mentioned mostly in feature requests:

All I would improve is maybe reminder system. Like a reminder to update journal and to exercise mentally or physically. [MoodTools]

One of the apps (What's Up) provided a forum where users could talk and support each other. This feature polarized the users, who either loved or hated it:

I love the community! The encourage me, give me helpful advices and I have made some new friends! [What's Up]

Another user mentioned:

The forum is horrible. Just a bunch of teenagers giving each other advice on “cutting” techniques or which pills are best for suicide. Obviously not monitored. Very sad. [What's Up]

Comments also highlighted the importance of customization. Users were happy when the app was customizable, and demanded more flexibility when it was not:

It has a lot of helpful wording and allows you to add your own, to personalize it. [Cognitive Diary CBT Self-Help]

Another user said:

There's also not many customization options with statements. For instance, as an atheist, statements like “I can turn this over to God” are just not helpful to me, so I'd like to be able to hide them. [Worry Box]

Therapists using an app with their clients also wanted the ability to customize it, with one observing:

I would have liked to have a way to list your own alternative coping statements along with the canned ones. [Worry Box]

Attitudes Toward Therapeutic Features Not Based on Cognitive Behavioral Therapy

Users frequently commented on therapeutic features not based specifically on CBT, especially relaxation tracks and meditation. Commenters also appreciated examples of activities they could do to improve their mood and the ability to track anxiety episodes and worry (although the majority of those comments referred to Worry Box—an app designed specifically to deal with these issues). Writing in a diary was also often mentioned, although it was not always clear whether the mentions referred to simple journaling or a structured thought diary. The majority of the users valued having both therapeutic features based on CBT and other approaches:

For a free app that provides cbt logs, meditation, relaxation training, the ability to track some of you depression symptoms & provide suggestions to get you up and moving—it's a well designed little app. [Depression CBT Self-Help Guide]

However, some indicated that the use of CBT was important because of its strong scientific or evidence base and did not like more eclectic apps:

This app is presented as a straightforward cognitive therapy app, but is riddled with pseudo-spiritual New Age nonsense. Avoid. [Depression CBT Self-Help Guide]

Some therapists who used an app with their clients were positive about the integration of these different types of features in a single app, with one commenting:

I was looking for an app to use with clients in my clinical work and I really like this one. Having both the relaxation tracks and a way to challenge unhelpful thoughts was great. [Worry Box]

Furthermore, the main “criticism” voiced by many users, together with associated feature and customization requests, was the absence of positivity:

It stresses ONLY negative feelings. I believe that focusing a person solely on their darker aspects only reinforces those aspects of their daily outlook. [MoodTools]

Another user mentioned:

I really didn't like how I could only write unhelpful thoughts. I do have positive thoughts too. I want to write those down. [Thought Diary]

Conclusions of Study 2

The analysis of user reviews showed that users appreciated all therapeutic features, including both the ones based on CBT as well as on other approaches, and so did professionals using the apps, which can explain the lack of correlation between high ratings and presence of evidence-based features reported in other studies (eg, [36,37]). Moreover, users wanted those other, non-CBT features, especially the ones focusing on more positive experiences.

Apps were often mentioned in the context of therapy—as “pocket therapists,” they often replaced or augmented therapy,

allowing users to take an active role in supporting their own mental well-being. Their discreet nature was particularly important. However, reliance on the app and the type of data users entered meant that privacy, security, and trust were important. And when that trust was violated, for example, when the app lost the data, it had serious consequences, leaving users devastated.

Discussion

Principal Findings

We presented 2 studies that investigated factors that make mental health apps engaging: in study 1, we examined engagement features available in CBT apps for depression, and the relationship between the presence of CBT features and expert involvement and the app ratings; in study 2, we thematically analyzed publicly available user reviews to understand user experience and contexts of use. Our results show that apps are used in a wide range of contexts, both replacing and augmenting therapy, and allowing users to take an active role in supporting their mental health and well-being. Users, including health professionals, valued and used apps that incorporated both core CBT and non-CBT elements, but concerns were also expressed regarding the unsupervised use of apps. Positivity was seen as important to engagement, for example, in the context of automatic thoughts, users expressed a preference to capture not just negative but also positive ones. Privacy, security, and trust were crucial to the user experience. We discuss these findings below.

Integration Into Different Therapeutic Practices

The results showed that apps were used as part of different therapeutic practices—as part of therapy as well as a tool for self-management. A mix of different features and varying contexts of use provides a challenge to app developers, but at the same time opens up opportunities for integrating apps into therapeutic practice. The prevalence of features related to key CBT concepts (eg, negative automatic thoughts, mood tracking) that do not necessarily require input from a therapist and are often covered in self-directed computerized CBT suggests that apps may, in particular, be a useful addition to low-intensity CBT. They may be particularly effective in facilitating engagement with CBT homework (which often requires recording thoughts and emotions, planning activities, or tracking mood), which is a desirable outcome as regular engagement with such CBT exercises increases the effectiveness of therapy [53,54]. Moreover, the results show that apps have important advantages with regard to integration into practice—they are always at hand, are more private and discreet than paper worksheets, and enable easy sharing of data with the therapist. Taking a smartphone out when in a shopping queue or on a bus is perfectly “normal” behavior, that is, it is familiar to the individual and (usually) acceptable to those around them. Therefore, people can integrate such exercises into everyday life and reduce concerns about “getting caught” completing therapy tasks in public [55].

Engagement With Therapy

Engagement with therapy or therapeutic content can be achieved through different means, including interactive features, peer or professional support, or customization [47]. Each of these approaches was found to some degree across the apps we examined, although CBT apps for depression lacked features enabling contact with others and more interactive app-based engagement techniques such as bots or the ability to upload own content. Study 2 findings suggest that reminders to use the app, in particular, are something that users would value. However, care must be taken when designing such reminders to ensure they are effective and not annoying [56]. As with other health apps (eg, menstruation trackers [57]), they also need to ensure privacy—it was clear from the user comments that people shared their phones with others and thus would want to keep private the fact that they use an app to support mental health. In exploring this area, many lessons could be drawn from existing literature on behavior change [21], health promotion [7], or medication adherence apps [43].

It is common in CBT interventions that therapists provide personalized examples or tailor exercises based on a client's needs [14]. Study 1 showed that personalization features were more prevalent among CBT apps for depression, and study 2 showed that both users and therapists appreciated the ability to customize apps, which reflects this personalized nature of therapy. This raises questions of how best to support customization of exercises. One option is for the customization to be done by the therapist, and appropriate content uploaded into the client's app. A second option involves a therapist and client working together to customize the app, for example, to set specific exercises for the client to complete between therapeutic sessions. Alternatively, the user (outside a therapeutic relationship) could make such modifications themselves, which some apps already allow. However, when customizations have the potential to impact the therapeutic effectiveness of an app, putting such abilities in the hands of the user would require responsibility on the part of the app designer, to help ensure that such customizations are therapeutically appropriate. The means of implementing this is an important subject for future research.

Building on an Evidence Base From Multiple Fields

The analysis of our data suggests that many end users, including therapists, valued flexibility in the use of therapeutic approaches. Similar to the finding of the study by Kertz et al [37], we found that many apps mixed both CBT and non-CBT features. Some users felt this was inappropriate, particularly when they thought the latter were not evidence-based. However, they were a minority and a number of commenters expressed a desire for features beyond standard CBT, such as the ability to record positive emotions. Although positive logs are used in CBT, the apps tend to focus on negative automatic thoughts and negative thinking styles. Moreover, some therapists seemed to approve the combination of CBT techniques with, for example, relaxation audio. Many users find such a blend to be more engaging, even if not faithful to a core CBT model [14]. Given the importance of engagement in achieving effective outcomes, this raises

important questions for the design of mental health apps and potential benefits of non-CBT features in CBT-focused apps.

Prior research on eclecticism emphasizes that it should not be construed as antitheoretical. Rather, eclectic approaches should be guided by some integrating framework that gives coherence to the overall therapeutic process. Within the context of traditional, face-to-face mental health interventions, several such frameworks have emerged [58]. Debate regarding the efficacy of eclecticism versus adherence to a core intervention model has a long history and is the subject of ongoing research [58–60]. A detailed discussion of this literature is beyond the scope of this paper. However, better understanding of how eclecticism can be effectively supported through apps, either with or without therapist support, is an important subject for future research. Promising initial work has been carried out by Mohr et al [26], who developed IntelliCare—an eclectic suite of apps that provide a wide range of features (their Thought Challenger was among the set of 31 CBT apps we identified). In addition, the most comprehensive app in our set, MoodTools, has been developed by health professionals, and is a good example of how different approaches can be blended. The potential benefits of combining CBT with approaches from Positive Psychology [61,62] is another clear message from our work. Moreover, although it is crucial that apps should draw on clinical theory, this alone is not enough. They also need to be based on research grounded in other domains such as HCI, which provide evidence of effective approaches to building desirable and engaging computer systems [23,47].

Responsibility Toward the Users

With potential users expressing different needs, and therefore using the apps in different ways, the above discussion raises the question of the responsibility held by app designers or sellers. Regulating health apps is a challenge [34], raising several ethical and practical issues. Should designers take responsibility for the ways in which their app is used? Should they somehow police it? It is clear that some users felt that certain apps should not be used by someone untrained in CBT or outside a therapeutic relationship. Assuming this view is correct, how can designers ensure that their app is not used inappropriately? Is there potential for the identification and warning of inappropriate use patterns? Alternatively, it may be appropriate to block unsupervised use of apps, for example, through access codes provided by professional therapists to their clients, as used by the Pesky gNATs app [63] and others.

This responsibility also extends to the stability and reliability of the app. While an app failure in general can be inconvenient and annoying, it can have serious consequences in the context of mental health apps—someone who has come to rely on an app for emotional support can find a failure “devastating.” This is also an issue with other health-related apps, for example, app updates can lead to a loss of scheduled reminders from medication adherence apps [43]. Therefore, reliability and backward compatibility of upgrades should be tested more thoroughly than for other types of apps. New providers entering the market should think seriously whether they can take on such responsibility.

This responsibility arguably applies to the app stores as well. They already encourage thorough testing of the apps before launch [32,33], routinely examine the technical implementation of apps, and block apps that do not meet the required technical standards. If an app is available in the health section, and offers mental health support, should the app stores be expected to police its quality from a health or clinical perspective? Perhaps such app listings should be required to explicitly state which health professionals (if any) were involved in its design, what evidence (if any) is available for the techniques it offers, and also provide guidance as to which contexts of use the developers feel are appropriate. Guidelines for evaluating health apps [64], objective app guides [65], and dedicated health app stores (eg, the one curated by the UK National Health Service [66]) are a good first step, but they do not solve the potential issues with apps widely available in commercial app stores.

Limitations

To assess whether the features of the 31 CBT apps for depression reflect CBT practice, we compared them with a recognized CBT Competence Framework [14]. It is important to note that this competence framework was developed for assessing competency of face-to-face therapists, not for CBT apps. Given the current absence of a widely recognized competence framework for CBT apps for depression, we believe this was the best approach.

Our focus on app stores means that the results reflect apps that are commonly available, rather than apps that are the current state-of-the-art apps in research. We believe that by exploring the experience of current users, our work complements other research and can help to inform future designs. Similarly, given

the focus on user reviews in study 2, there is an obvious potential for selection bias toward extreme ratings and positive reviews [67,68]. However, we believe that focusing on specific features and context of use allowed us to reduce this bias while still providing relevant insights. Although the reviews were coded by 1 author, the coding and the data were regularly discussed with the rest of the team to ensure everyone is familiar with the data, which is an acceptable approach in qualitative analysis [52,69,70]. Overall, although not definitive, our approach allowed access to user experience data that would otherwise be extremely difficult to obtain. It provided data on experience across a large number of different apps, which people have used over different durations, as part of their daily life. In-depth studies of individual apps do provide stronger data, but not the same breadth of coverage. They also prescribe how a user in a study is expected to use the app, and so provide less insight into the development of in-the-wild usage patterns in response to personal situations.

Conclusions and Future Work

Drawing from a rich pool of public app reviews, our research shows that users use apps alongside and instead of therapy, with the same app being used in both contexts. It also suggests that features not considered evidence-based may be key to facilitating user engagement. The challenges and opportunities we have identified open up new avenues for research. Future work should explore approaches to integrating apps into different therapeutic practices, facilitating engagement, finding a balance between drawing from clinical and design research, and exploring different approaches toward responsibility and accountability, and the role of app stores as gatekeepers.

Acknowledgments

This report is independent research funded by the National Institute for Health Research (Programme Grants for Applied Research, Integrated therapist and online CBT for depression in primary care, RP-PG-0514-20012). This study was also supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health. The authors would like to thank Dr David Kessler for his helpful comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of 31 CBT apps for depression.

[[XLSX File \(Microsoft Excel File\)](#), 12KB - [jmir_v20i6e10120_app1.xlsx](#)]

References

1. Whitton AE, Proudfoot J, Clarke J, Birch MR, 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 Mar 04;2(1):e3 [[FREE Full text](#)] [doi: [10.2196/mental.3573](#)] [Medline: [26543909](#)]
2. World Health Organization. 2008. The global burden of disease: 2004 update URL: http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/
3. Firth J, Torous J, Nicholas J, Carney R, Prata A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry* 2017 Oct;16(3):287-298 [[FREE Full text](#)] [doi: [10.1002/wps.20472](#)] [Medline: [28941113](#)]

4. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *J Affect Disord* 2017 Aug 15;218:15-22 [[FREE Full text](#)] [doi: [10.1016/j.jad.2017.04.046](#)] [Medline: [28456072](#)]
5. Foroushani PS, Schneider J, Assareh N. Meta-review of the effectiveness of computerised CBT in treating depression. *BMC Psychiatry* 2011 Aug 12;11:131 [[FREE Full text](#)] [doi: [10.1186/1471-244X-11-131](#)] [Medline: [21838902](#)]
6. Grist R, Porter J, Stallard P. Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. *J Med Internet Res* 2017 May 25;19(5):e176 [[FREE Full text](#)] [doi: [10.2196/jmir.7332](#)] [Medline: [28546138](#)]
7. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. 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](#)]
8. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression? A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. *BMC Psychiatry* 2013 Apr 15;13:113 [[FREE Full text](#)] [doi: [10.1186/1471-244X-13-113](#)] [Medline: [23587347](#)]
9. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. *JAMA Psychiatry* 2017 Apr 01;74(4):351-359. [doi: [10.1001/jamapsychiatry.2017.0044](#)] [Medline: [28241179](#)]
10. 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](#)]
11. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. *Health Expect* 2009 Mar;12(1):45-59. [doi: [10.1111/j.1369-7625.2008.00531.x](#)] [Medline: [19250152](#)]
12. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. *BMJ* 2015 Nov 11;351:h5627 [[FREE Full text](#)] [Medline: [26559241](#)]
13. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. *BMJ Open* 2015;5(11):e008581. [doi: [10.1136/bmjopen-2015-008581](#)] [Medline: [26621513](#)]
14. University College London. Cognitive and Behavioural Therapy URL: <https://www.ucl.ac.uk/pals/research/cehp/research-groups/core/competence-frameworks/cognitive-and-behavioural-therapy> [accessed 2018-02-08] [[WebCite Cache ID 6x52vkXl2](#)]
15. Ofcom. 2017 Aug 03. Communications Market Report 2017 URL: https://www.ofcom.org.uk/data/assets/pdf_file/0017/105074/cmr-2017-uk.pdf [accessed 2018-02-08] [[WebCite Cache ID 6x548sapR](#)]
16. Lella A. ComScore. 2017. U.S. Smartphone Penetration Surpassed 80 Percent in 2016 URL: <https://www.comscore.com/Insights/Blog/US-Smartphone-Penetration-Surpassed-80-Percent-in-2016> [accessed 2018-02-08]
17. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. *J Med Internet Res* 2016 Dec 20;18(12):e330 [[FREE Full text](#)] [doi: [10.2196/jmir.6482](#)] [Medline: [27998876](#)]
18. Tang W, Kreindler D. Supporting Homework Compliance in Cognitive Behavioural Therapy: Essential Features of Mobile Apps. *JMIR Ment Health* 2017 Jun 08;4(2):e20 [[FREE Full text](#)] [doi: [10.2196/mental.5283](#)] [Medline: [28596145](#)]
19. Bardram JE, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM Press; 2013 Presented at: CHI '13; Apr 27 - May 02, 2013; Paris, France p. 2627-2636. [doi: [10.1145/2470654.2481364](#)]
20. Barry M, Doherty K, Marcano Belisario J, Car J, Morrison C, Doherty G. mHealth for Maternal Mental Health: Everyday Wisdom in Ethical Design. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM Press; 2017 Presented at: CHI '17; May 06-11, 2017; Denver, CO, USA p. 2708-2756. [doi: [10.1145/3025453.3025918](#)]
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 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1001363](#)] [Medline: [23458994](#)]
22. Kelley C, Lee B, Wilcox L. Self-tracking for Mental Wellness: Understanding Expert Perspectives and Student Experiences. In: *Proc SIGCHI Conf Hum Factor Comput Syst*. New York, NY, USA: ACM Press; 2017 Presented at: The 35th SIGCHI Conference on Human Factors in Computing Systems; May 06-11, 2017; Denver, CO, USA p. 629-641. [doi: [10.1145/3025453.3025750](#)]
23. Matthews M, Doherty G. In the mood: engaging teenagers in psychotherapy using mobile phones. In: *Proc SIGCHI Conf Hum Factor Comput Syst*. New York, NY, USA: ACM Press; 2011 Presented at: The 29th SIGCHI Conference on Human Factors in Computing Systems; May 07-12, 2011; Vancouver, BC, Canada p. 2947-2956. [doi: [10.1145/1978942.1979379](#)]
24. Matthews M, Murnane E, Snyder J. Quantifying the Changeable Self: The Role of Self-Tracking in Coming to Terms With and Managing Bipolar Disorder. *Human-Computer Interaction* 2017;32(5-6):413-446. [doi: [10.1080/07370024.2017.1294983](#)]

25. Mohr DC, Montague E, Stiles-Shields C, Kaiser SM, Brenner C, Carty-Fickes E, et al. MedLink: a mobile intervention to address failure points in treatment of depression in general medicine. In: *Int Conf Pervasive Comput Technol Healthc*. Brussels, Belgium: ICST; 2015 May Presented at: The 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2015; Istanbul, Turkey p. 100-107. [doi: [10.4108/icst.pervasivehealth.2015.259042](https://doi.org/10.4108/icst.pervasivehealth.2015.259042)]
26. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: An Eclectic, Skills-Based App Suite for the Treatment of Depression and Anxiety. *J Med Internet Res* 2017 Jan 05;19(1):e10 [FREE Full text] [doi: [10.2196/jmir.6645](https://doi.org/10.2196/jmir.6645)] [Medline: [28057609](https://pubmed.ncbi.nlm.nih.gov/28057609/)]
27. Rennick-Egglestone S, Knowles S, Toms G, Bee P, Lovell K, Bower P. Health Technologies “In the Wild”: Experiences of Engagement with Computerised CBT. In: *Proc SIGCHI Conf Hum Factor Comput Syst*. New York, NY, USA: ACM Press; 2016 Presented at: The 34th SIGCHI Conference on Human Factors in Computing Systems; May 07-12, 2016; San Jose, CA, USA p. 2124-2135. [doi: [10.1145/2858036.2858128](https://doi.org/10.1145/2858036.2858128)]
28. Topham P, Caleb-Solly P, Matthews P, Farmer A, Mash C. Mental Health App Design: A Journey From Concept to Completion. In: *Prof of MobileHCI*. New York, NY, USA: ACM Press; 2015 Presented at: The 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct; Aug 24-27, 2015; Copenhagen, Denmark p. 582-591. [doi: [10.1145/2786567.2787136](https://doi.org/10.1145/2786567.2787136)]
29. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. *JMIR Ment Health* 2016 Mar 01;3(1):e7 [FREE Full text] [doi: [10.2196/mental.4984](https://doi.org/10.2196/mental.4984)] [Medline: [26932350](https://pubmed.ncbi.nlm.nih.gov/26932350/)]
30. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. *J Med Internet Res* 2013 Nov 15;15(11):e247 [FREE Full text] [doi: [10.2196/jmir.2791](https://doi.org/10.2196/jmir.2791)] [Medline: [24240579](https://pubmed.ncbi.nlm.nih.gov/24240579/)]
31. Lui JHL, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. *Professional Psychology: Research and Practice* 2017;48(3):199-210. [doi: [10.1037/pro0000122](https://doi.org/10.1037/pro0000122)]
32. Apple Inc. Submitting iOS apps to the App Store URL: <https://developer.apple.com/ios/submit/> [accessed 2018-02-13] [WebCite Cache ID 6xCo7OGON]
33. Google Developers. Launch. How to publish, manage, and distribute your app worldwide URL: <https://developer.android.com/distribute/best-practices/launch/index.html> [accessed 2018-02-13] [WebCite Cache ID 6xCnz6Cp1]
34. Vincent CJ, Niezen G, O’Kane AA, Stawarz K. Can standards and regulations keep up with health technology? *JMIR Mhealth Uhealth* 2015 Jun 03;3(2):e64 [FREE Full text] [doi: [10.2196/mhealth.3918](https://doi.org/10.2196/mhealth.3918)] [Medline: [26041730](https://pubmed.ncbi.nlm.nih.gov/26041730/)]
35. Kristensen N, Nymann C, Konradsen H. Implementing research results in clinical practice- the experiences of healthcare professionals. *BMC Health Serv Res* 2016 Feb 10;16:48 [FREE Full text] [doi: [10.1186/s12913-016-1292-y](https://doi.org/10.1186/s12913-016-1292-y)] [Medline: [26860594](https://pubmed.ncbi.nlm.nih.gov/26860594/)]
36. 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 May 02;11(5):e0154248 [FREE Full text] [doi: [10.1371/journal.pone.0154248](https://doi.org/10.1371/journal.pone.0154248)] [Medline: [27135410](https://pubmed.ncbi.nlm.nih.gov/27135410/)]
37. Kertz SJ, MacLaren Kelly J, Stevens KT, Schrock M, Danitz SB. A Review of Free iPhone Applications Designed to Target Anxiety and Worry. *J Technol Behav Sci* 2017 Jun;2(2):61-70. [doi: [10.1007/s41347-016-0006-y](https://doi.org/10.1007/s41347-016-0006-y)]
38. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth* 2015 Feb 16;3(1):e16 [FREE Full text] [doi: [10.2196/mhealth.3713](https://doi.org/10.2196/mhealth.3713)] [Medline: [25689790](https://pubmed.ncbi.nlm.nih.gov/25689790/)]
39. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: There is an app for that. A systematic review of anxiety apps. *Depress Anxiety* 2017;34(6):518-525. [doi: [10.1002/da.22654](https://doi.org/10.1002/da.22654)] [Medline: [28504859](https://pubmed.ncbi.nlm.nih.gov/28504859/)]
40. Jacob C, Veerappa V, Harrison R. What Are You Complaining About?: A Study of Online Reviews of Mobile Applications. In: *Proc BCS-HCI*. London, UK: British Computer Society; 2013 Presented at: The 27th International BCS Human Computer Interaction Conference; Sep 9-13, 2013; London, UK.
41. Jacob C, Harrison R. Retrieving and analyzing mobile apps feature requests from online reviews. In: *Proc IEEE MSR*. Piscataway, NJ, USA: IEEE Press; 2013 Presented at: The 10th Working Conference on Mining Software Repositories; May 18-19, 2013; San Francisco, CA, USA p. 41-44. [doi: [10.1109/MSR.2013.6624001](https://doi.org/10.1109/MSR.2013.6624001)]
42. Khalid H. On identifying user complaints of iOS apps. In: *Proceedings of the 2013 International Conference on Software Engineering*. Piscataway, NJ, USA: IEEE Press; 2013 Presented at: ICSE '13; May 18-26, 2013; San Francisco, CA, USA p. 1474-1476.
43. Stawarz K, Cox AL, Blandford A. Don't forget your pill! Designing Effective Medication Reminder Apps That Support Users' Daily Routines. In: *Proc SIGCHI Conf Hum Factor Comput Syst*. New York, NY, USA: ACM Press; 2014 Presented at: The 32nd SIGCHI Conference on Human Factors in Computing Systems; Apr 26 - May 01, 2014; Toronto, Ontario, Canada p. 2269-2278. [doi: [10.1145/2556288.2557079](https://doi.org/10.1145/2556288.2557079)]
44. Caldeira C, Chen Y, Chan L, Pham V, Chen Y, Zheng K. Mobile apps for mood tracking: an analysis of features and user reviews. In: *Proc AMIA Annual Symposium*. Bethesda, MD, USA: American Medical Informatics Association; 2017 Presented at: American Medical Informatics Association Annual Symposium; Nov 4-8, 2017; Washington, DC, USA.

45. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR Mhealth Uhealth* 2015 Nov 04;3(4):e101 [[FREE Full text](#)] [doi: [10.2196/mhealth.4924](#)] [Medline: [26537656](#)]
46. Rubanovich CK, Mohr DC, Schueller SM. Health App Use Among Individuals With Symptoms of Depression and Anxiety: A Survey Study With Thematic Coding. *JMIR Ment Health* 2017 Jun 23;4(2):e22 [[FREE Full text](#)] [doi: [10.2196/mental.7603](#)] [Medline: [28645891](#)]
47. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: *Proc SIGCHI Conf Hum Factor Comput Syst*. New York, NY, USA: ACM Press; 2012 Presented at: The 30th SIGCHI Conference on Human Factors in Computing Systems; May 05-10, 2012; Austin, TX, USA p. 1421-1430. [doi: [10.1145/2207676.2208602](#)]
48. 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 Mar 11;3(1):e27 [[FREE Full text](#)] [doi: [10.2196/mhealth.3422](#)] [Medline: [25760773](#)]
49. Chan S, Torous J, Hinton L, Yellowlees P. Towards a Framework for Evaluating Mobile Mental Health Apps. *Telemed J E Health* 2015 Dec;21(12):1038-1041. [doi: [10.1089/tmj.2015.0002](#)] [Medline: [26171663](#)]
50. Olano F. Github. Node.js scraper to get data from Google Play URL: <https://github.com/facundoolano/google-play-scraper> [accessed 2018-02-08] [[WebCite Cache ID 6x4rQ5dYf](#)]
51. Olano F. Github. Scrape data from the itunes app store URL: <https://github.com/facundoolano/app-store-scraper> [accessed 2018-02-08] [[WebCite Cache ID 6x52S5TmC](#)]
52. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;3(2):77-101. [doi: [10.1191/1478088706qp063oa](#)]
53. 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](#)]
54. Westra HA, Dozois DJ, Marcus M. Expectancy, homework compliance, and initial change in cognitive-behavioral therapy for anxiety. *J Consult Clin Psychol* 2007 Jun;75(3):363-373. [doi: [10.1037/0022-006X.75.3.363](#)] [Medline: [17563153](#)]
55. Barnes M, Sherlock S, Thomas L, Kessler D, Kuyken W, Owen-Smith A, et al. No pain, no gain: depressed clients' experiences of cognitive behavioural therapy. *Br J Clin Psychol* 2013 Nov;52(4):347-364. [doi: [10.1111/bjc.12021](#)] [Medline: [24117909](#)]
56. Mehrotra A, Pejovic V, Vermeulen J, Hendley R, Musolesi M. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In: *Proc SIGCHI Conf Hum Factor Comput Syst*. New York, NY, USA: ACM Press; 2016 Presented at: The 34th SIGCHI Conference on Human Factors in Computing Systems; May 07-12, 2016; San Jose, CA, USA p. 1021-1032. [doi: [10.1145/2858036.2858566](#)]
57. Epstein DA, Lee NB, Kang JH, Agapie E, Schroeder J, Pina LR, et al. Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools. In: *Proc SIGCHI Conf Hum Factor Comput Syst*. New York, New York, USA: ACM Press; 2017 Presented at: The 35th SIGCHI Conference on Human Factors in Computing Systems; Apr 27 - May 02, 2017; Denver, CO, USA p. 6876-6888.
58. Feltham C, Hanley T, Winter L, editors. *Integrative Eclectic Approaches*. In: *The SAGE Handbook of Counselling and Psychotherapy*, 4th ed. London, UK: SAGE; 2017:319-360.
59. Norcross JC. *Handbook of eclectic psychotherapy*. Michigan, USA: Brunner/Mazel; 1986.
60. Norcross JC. A primer on psychotherapy integration. In: Norcross JC, Goldfried MR, editors. *Handbook of Psychotherapy Integration*, 2nd ed. Oxford, UK: Oxford University Press; 2005:3-23.
61. Hamilton N, Kitzman H, Guyotte S. Enhancing Health and Emotion: Mindfulness as a Missing Link Between Cognitive Therapy and Positive Psychology. *J Cogn Psychother* 2006;20(2):123-134 [[FREE Full text](#)] [doi: [10.1891/jcop.20.2.123](#)]
62. Karwoski L, Garratt GM, Ilardi SS. On the Integration of Cognitive-Behavioral Therapy for Depression and Positive Psychology. *J Cogn Psychother* 2006;20(2):159-170. [doi: [10.1891/jcop.20.2.159](#)]
63. Handaxe Limited. App Store. Pesky gNATs URL: <https://itunes.apple.com/ie/app/pesky-gnats/id917307943?mt=8> [[WebCite Cache ID 6x54ubn9A](#)]
64. MindTech. Toolkit for Appraising Digital Mental Health Products - Technology theme URL: <http://mindtech.org.uk/projects/71-digital-mental-health-tools-evaluation-criteria-technology-theme.html> [accessed 2018-02-08]
65. One Mind. PsyberGuide. URL: <https://psyberguide.org/> [[WebCite Cache ID 6xzHXPJMC](#)]
66. NHS. Library of mental health apps URL: <https://apps.beta.nhs.uk/?category=Mental%20Health> [accessed 2018-02-08] [[WebCite Cache ID 6x57BtjBs](#)]
67. Chevalier JA, Mayzlin D. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* 2006 Aug;43(3):345-354. [doi: [10.1509/jmkr.43.3.345](#)]
68. Hu N, Pavlou PA, Zhang J. Can online reviews reveal a product's true quality?: Empirical findings and analytical modeling of online word-of-mouth communication. In: *Proc ACM EC*. New York, New York, USA: ACM Press; 2006 Presented at: The 7th ACM conference on Electronic Commerce; Jun 11-15, 2006; Ann Arbor, MI, USA p. 324-330. [doi: [10.1145/1134707.1134743](#)]
69. Braun V, Clarke V. *Successful qualitative research: A practical guide for beginners*. London, UK: SAGE; 2013.

70. Blandford A, Furniss D, Makri S. Qualitative HCI Research: Going Behind the Scenes. Synthesis Lectures on Human-Centered Informatics 2016;9(1):1-115. [doi: [10.2200/S00706ED1V01Y201602HCI034](https://doi.org/10.2200/S00706ED1V01Y201602HCI034)]

Abbreviations

CBT: cognitive behavioral therapy

HCI: human-computer interaction

Edited by G Wadley, R Calvo, M Czerwinski, J Torous; submitted 15.02.18; peer-reviewed by K O'Leary, S Rennick-Egglestone, Y Lee, I Choi, A Basu; comments to author 08.03.18; revised version received 22.03.18; accepted 22.03.18; published 06.06.18.

Please cite as:

Stawarz K, Preist C, Tallon D, Wiles N, Coyle D

User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews
J Med Internet Res 2018;20(6):e10120

URL: <http://www.jmir.org/2018/6/e10120/>

doi: [10.2196/10120](https://doi.org/10.2196/10120)

PMID:

©Katarzyna Stawarz, Chris Preist, Debbie Tallon, Nicola Wiles, David Coyle. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.06.2018. 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.

Original Paper

Discovery of and Interest in Health Apps Among Those With Mental Health Needs: Survey and Focus Group Study

Stephen M Schueller¹, PhD (Clin Psych); Martha Neary¹, MSc; Kristen O'Loughlin¹, BS; Elizabeth C Adkins¹, MA

Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:

Stephen M Schueller, PhD (Clin Psych)
Center for Behavioral Intervention Technologies
Department of Preventive Medicine, Feinberg School of Medicine
Northwestern University
750 N. Lake Shore Drive
10th Floor, CBITs
Chicago, IL, 60614
United States
Phone: 1 312 503 1232
Email: schueller@northwestern.edu

Abstract

Background: A large number of health apps are available directly to consumers through app marketplaces. Little information is known, however, about how consumers search for these apps and which factors influence their uptake, adoption, and long-term use.

Objective: The aim of this study was to understand what people look for when they search for health apps and the aspects and features of those apps that consumers find appealing.

Methods: Participants were recruited from Northwestern University's Center for Behavioral Intervention Technologies' research registry of individuals with mental health needs. Most participants (n=811) completed a survey asking about their use and interest in health and mental health apps. Local participants were also invited to participate in focus groups. A total of 7 focus groups were conducted with 30 participants that collected more detailed information about their use and interest in health and mental health apps.

Results: Survey participants commonly found health apps through social media (45.1%, 366/811), personal searches (42.7%, 346/811), or word of mouth (36.9%, 299/811), as opposed to professional sources such as medical providers (24.6%, 200/811). From the focus groups, common themes related to uptake and use of health apps included the importance of personal use before adoption, specific features that users found desirable, and trusted sources either developing or promoting the apps.

Conclusions: As the number of mental health and health apps continue to increase, it is imperative to better understand the factors that impact people's adoption and use of such technologies. Our findings indicated that a number of factors—ease of use, aesthetics, and individual experience—drove adoption and use and highlighted areas of focus for app developers and disseminators.

(*J Med Internet Res* 2018;20(6):e10141) doi:[10.2196/10141](https://doi.org/10.2196/10141)

KEYWORDS

mHealth; mental health; mobile apps; consumer preference; focus groups

Introduction

Background

The number of publicly available mental health apps continues to expand at a breakneck pace. One estimate, as of 2017, proposed that nearly 325,000 health apps are available across the most common app stores (Google Play and iOS), a 25%

increase from the previous year [1]. The exact number of mental health apps varies by definitions of what constitutes a *mental health* app, ranging from possibilities including general wellness apps to only disorder-specific mental health apps. Estimates suggest that about 7% of the market is focused on mental health [2], which would imply that 22,750 mental health apps exist. Indeed, more than 600 apps focus on depression, and 200 apps focus on suicide alone [3]. In light of this, various strategies

have been proposed for how professionals and researchers might search for and understand these products [4]. However, there has been little investigation into how consumers with mental health needs search for and select health apps. In this study, first we conducted a survey of people with mental health needs and then conducted in-person focus groups to gain more detailed information. The goals of both the survey and focus groups were to better understand how people find health apps and information they use to guide their decision as to which apps to use.

Despite the wide availability of mental health apps, their impact on addressing the burden of mental health has been seriously lacking. This is largely because of the limited uptake and adoption of such tools both in routine care settings and by users in direct-to-consumer models. Furthermore, even those who do download a mental health app are unlikely to persist with that app over time. Two examples of publicly available apps with published information on their use are PTSD Coach [5-7] and IntelliCare [8]. In an evaluation of the uptake of PTSD Coach in the wild over 3 years, it was downloaded 153,834 times with 61.1% of people using the app within a day of installation, but only 41.6% using it after a month and 19.4% using it after 6 months [9]. In the case of IntelliCare, in the first year of its availability, 5210 people downloaded a total of 10,131 IntelliCare apps [8]. About half of the users continued to use the apps a day after installation, whereas a month after installation, rates of usage of each IntelliCare app ranged from 12.02% to 23.30% of the people who initially downloaded each respective app. It is worth noting that these 2 examples likely represent a best case for mental health apps, as these apps have empirical studies supporting their effectiveness [6,10]; were developed by well-respected government agencies and an academic medical school via government research funding; and are mentioned regularly in lay and professional audiences. Many mental health apps do not enjoy these benefits. About one-fourth of health apps downloaded are never opened, and 50% of health apps receive fewer than 500 downloads [11]. Although many might suggest that evidence supporting their efficacy or the development team should drive user adoption, the degree to which users value these factors in making choices to adopt a mental health app is unclear.

It has been suggested elsewhere that app adoption is a heuristic process that is guided by various informational cues [12]. Huang and Bashir [12] examined mental health apps intended to reduce anxiety and found that app ratings and reviews on app marketplaces correlated positively with the number of installs, whereas app price correlated negatively with the number of installs. Furthermore, app titles directly related to anxiety disorders or specific symptoms had lower rates of installation than apps with descriptions of the activities contained within the app such as mindfulness or journaling. A limitation of Huang and Bashir's [12] study, however, was that it used observational data obtained directly from the app stores to understand what influences people's adoption of mental health apps. Further information could be gleaned by directly asking consumers interested in mental health apps.

The most influential model related to the adoption of digital tools is the technology acceptance model (TAM) [13]. TAM

proposes that the behavioral intention to use a technology precedes its use, which would result in uptake and adoption. TAM has been expanded to highlight the factors that influence behavioral intention including performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price, value, and habit in the unified theory of acceptance and use of technology (UTAUT) [14]. Although UTAUT is conceptually useful to highlight the constructs that impact people's intentions and use, each of these constructs need to be operationalized with regards to specific technologies and populations. For example, cost has been noted to be an important determinant of health app adoption [15], but it also seems feasible that people might be willing to pay for apps that confer true health benefits. In the context of mental health, stigma might impact people's willingness to talk about app use specifically or mental health generally that might impact social influence. As such, in this study, we aim to explore behaviors and perspectives as they relate to mental health apps in a population for which use of such apps would be relevant.

Finally, it is important to understand what information consumers are looking for to improve efforts to provide consumers with guidance for identifying and selecting mental health apps. Indeed, a recent report from a working group from the National Institute of Mental Health on opportunities and challenges of technology in clinical research concluded that "there is a need for rigorous evaluation and development of an evaluation structure of these apps" [16]. Several evaluation structures have been proposed either drawing from expert consensus such as the American Psychiatric Association's App Evaluation Model [17] or from the synthesis of existing app rating structures such as the Enlight evaluation framework [18]. These structures share a multifaceted structure that considers several elements of the app such as research evidence, ease of use, therapeutic persuasiveness, privacy and security, and aesthetics. However, these models do not consider what factors are most important to consumers.

Objectives

The field needs to better understand what consumers are looking for to build better products that incorporate those qualities and combine evidence-based practices that will result in effective and desirable mental health apps. Furthermore, understanding how people search for apps and what influences their decision to use an app may be helpful in presenting information about apps in persuasive ways to drive uptake and long-term use. This study addressed these issues through asking people about these questions using a survey and focus group methodology. Combining surveys with focus groups combines strengths of both approaches by collecting a large sample of respondents in surveys but eliciting more detailed and nuanced information in focus groups.

Methods

Study Design

We conducted a survey and focus groups to understand how people with mental health needs search for health apps and what information is valuable to consumers in making a decision as to the quality or desirability of particular apps. All participants

were recruited from a research registry maintained by the Center for Behavioral Intervention Technologies (CBITs), which contains people who are willing to be contacted for future research opportunities. This research registry is framed as an opportunity to be involved in research on the use of technology to improve psychological well-being and improve general health with a particular focus on depression and anxiety. The survey was designed to take between 30 and 45 min to complete and could be completed remotely in exchange for entry into a lottery for a US \$50 Amazon gift card. The focus groups lasted 90 min and are described in more detail below.

Recruitment

An email blast was sent to members of the CBITs research registry, which contains 5100 members. Registry members living in Chicago were invited to complete a survey and a focus group but could complete either if they preferred. Registry members living outside Chicago were invited to complete the survey only. The survey link remained live for 8 weeks from a period of October through December 2017, at which point recruitment was suspended because of the high number of respondents. Inclusion criteria were ownership of a smartphone and being comfortable speaking in English. All recruitment and study procedures were approved by the Institutional Review Board of Northwestern University.

Survey Sample

Of the 5100 registry members sent the survey link (both inside and outside of Chicago), 940 opened the survey, representing a response rate of 18.43% (940/5100). Moreover, 932 of these consented to participate and 811 completed the survey. Of the survey respondents, 79.5% were female (645/811), 18.3% were male (149/811), and 2.1% did not specify gender (17/811). The age range was 18 to 84 years (mean 36.1, SD 13.5). The majority of the sample was well educated, as outlined in Table 1. Although we did not ask about mental health symptoms in the context of our study, this information is collected when people enroll in the registry. The registry has elevated levels of symptoms of depression and anxiety with average scores of

14.3 (SD 5.4) on the Patient Health Questionnaire-8 and average scores of 12.2 (SD 5.4) on the Generalized Anxiety Disorder-7. Scores greater than 9 on the Patient Health Questionnaire and the Generalized Anxiety Disorder-7 are indicative of moderate depression or moderate anxiety and are recommended levels for referring people to treatment. We did ask whether survey respondents were receiving mental health treatments, and 57.9% (469/810) indicated they were, including 36.0% (292/810) receiving therapy and 51.1% (414/810) receiving medication. Thus, participants appeared to have mental health needs with many receiving mental health treatments.

These participants had experience with health apps generally. The average number of apps participants reported having on their phone was 54.14 (SD 50.89), with 3.12 (SD 4.35) of these or approximately 6% being health related (3.12/54.14). About one-third of participants (33.8%, 274/811) reported using a health app at least more days than not over the past week. A considerable minority reported they had not used a health app at all over the past week (28.8%, 234/811). Thus, although it seems that health app ownership was high, health app use was not. For mental health apps specifically, about one-third of the sample (33.9%, 275/811) indicated they had mental health apps on their phones. We discuss the results of the survey below.

Focus Group Sample

In total, 163 eligible prospective participants expressed interest in a focus group, and a random selection were invited to a group. Seven focus groups were conducted with a total of 30 participants (23 females and 7 males) and an average of 4 participants per focus group (range of 3-6 participants in each group). Just under half (47%, 14/30) of the focus group sample had also completed the survey. Participants ranged in age from 21 to 72 years (mean 43.3, SD 14.3). The sample had varying levels of formal education, as outlined in Table 2. Participants reported high levels of confidence using a smartphone, demonstrated by level of agreement with the statement "I feel confident using a smartphone and downloading and using apps"; 27 participants strongly agreed, 2 agreed, and 1 neither agreed nor disagreed.

Table 1. Highest level of formal education completed by survey respondents.

Level of education	n (%)
Less than high school	5 (0.6)
High school graduate	38 (4.7)
Some college, no degree	178 (21.9)
Associate's degree	62 (7.6)
Bachelor's degree	306 (37.7)
Master's degree	193 (23.8)
PhD	29 (3.6)
Total	811 (100.0)

Table 2. Highest level of formal education completed by focus group participants.

Level of education	n (%)
Less than high school	1 (3)
High school graduate	0 (0)
Some college, no degree	7 (3)
Associate's degree	2 (7)
Bachelor's degree	11 (37)
Master's degree	6 (20)
PhD	3 (10)
Total	30 (100)

Focus Group Procedures

The groups were held at Northwestern University's CBITs office space. Participants received US \$30 Amazon credit for their participation. Focus groups were semistructured, and facilitators (2 per group) took a flexible approach; questions were asked to guide the group through the relevant topics, whereas unanticipated ideas that emerged in the discussion were also pursued. The focus groups' aim was to focus on mental health apps, and although we did discuss health apps generally, mental health topics roughly accounted for two-thirds of discussion with the groups. The full semistructured focus group guide is included in [Multimedia Appendix 1](#). In brief, the focus groups were divided into 3 parts. The first was a discussion of mobile apps for health and participant's experiences (both positive and negative) of using health apps. The second part was a discussion of mobile apps specifically for mental health. Again, participants were asked to share both positive and negative experiences using mental health apps as well as ideas on where to look for mental health apps and what information is important to them when choosing an app. The third and final part of the focus group focused specifically on PsyberGuide [19], a Web-based resource that identifies and reviews mental health apps. This study's authors are responsible for the operation of and content on PsyberGuide with funding from One Mind, a nonprofit organization under which PsyberGuide is managed. Participants were guided through the website on a projector and gave feedback on content, design, and navigation. Learnings from the PsyberGuide portion of the interview were much more specific (eg, positive reactions to the PsyberGuide's nonprofit status, increased desire for features to improve navigation and discoverability of apps, and a strong negative reaction to the word "product" as it conveyed commercial interests). These comments guided changes to the site content. We do not report on specific reactions to PsyberGuide further below, but PsyberGuide is available online.

Results

Survey Results

The most common source by which participants identified mental health apps was through social media (45.1%, 366/811) followed closely behind by their own searches (42.7%, 346/811). Common places people searched for apps were the app stores,

Google searches, and Web forums such as Reddit. Although a considerable percentage did indicate that their medical providers were providing information about specific apps (24.6%, 200/811), even more participants indicated that a friend or family member helped them identify apps (36.9%, 299/811). As such, it seems like informal sources of information are relied on more than formal of sources of information in identifying mental health apps.

We also asked participants about the relative importance of a variety of features that might impact their adoption and sustained use of mental health apps. Participants responded on a 5-point Likert scale ranging from "not at all important" to "very important." In [Table 3](#), we display the number of participants who indicated that a feature was either "important" or "very important" to them, ordered in terms of rankings of most important to least important feature based on average responses. In general, the most important features related to the use of the app: is it easy to use and understandable? Issues related to privacy and data security (especially on the app side in terms of encryption compared with the user side in terms of a password) also appeared to be important.

We also explored what kept participants from downloading mental health apps to better understand barriers to uptake and adoption. The most common response was that participants were unsure how effective an app would be (31.4%, 255/811), although many fewer participants indicated that lack of research support contributed to this decision (6.6%, 54/811). Another highly endorsed barrier was about lack of knowledge regarding how to find an app or knowing which app to download (27.3%, 222/811). In general, other concerns were much lower including cost (13.7%, 111/811), lack of interest (11.1%, 90/811), privacy and data security (10.7%, 87/811), lack of time to use apps (6.6%, 54/811), lack of space on one's device (6.0%, 49/811), and/or usability issues (5.0%, 41/811).

Finally, we asked participants about what they liked about current mental health apps. Findings from these questions were largely consistent with the patterns found across other questions. The most common response was related to ease of use (27.0%, 219/811), visual appeal (18.2%, 148/811), simple language (17.4%, 141/811), and content (14.4%, 117/811). Here, participants did not indicate fun (7.7%, 63/811) or name of the app (4.7%, 38/811) being particularly appealing aspects.

Table 3. Importance of features in mental health apps.

Feature	Responses, n		Total responses, n (%)
	Important	Very important	Important and very important
Content	324	412	736 (90.8)
Ease of use	321	406	727 (89.6)
Cost	222	420	642 (79.2)
Encryption	201	401	602 (74.2)
Interactive features	314	284	598 (73.7)
Customization	323	252	575 (70.9)
Privacy policy	195	377	572 (70.5)
Direct research evidence	271	293	564 (69.5)
Indirect research evidence	301	241	542 (66.8)
Simple language	277	215	492 (60.7)
User ratings	314	168	482 (59.4)
User reviews	293	183	476 (58.7)
Visual appeal	288	162	450 (55.5)
App description	268	161	429 (52.9)
Developer	208	199	407 (50.2)
Fun	238	159	397 (48.9)
Password protected	162	205	367 (45.2)
Graphics	214	124	338 (41.7)
Name	90	35	125 (15.4)

Focus Group Results

Data Analysis

All sessions were audio-recorded and transcribed for coding. We conducted an inductive thematic analysis [20]. After sessions were transcribed, all transcripts were read with first memos and then open codes were created. After each transcript, codes were reviewed, which helped facilitate coding of subsequent transcripts. After all transcripts were memoed, coded, and initial themes identified, transcripts were read over again to identify which themes could be revised or combined. Themes were discussed among the study team, including the lead author who led the thematic analysis and the remaining authors who conducted the focus groups. We present the results of this process and provide quotes as specific examples of each theme within the results. We identify participants by number (eg, P1, P2, ...) and which focus group each participant was associated with (eg, FG1, FG2, ...).

Themes

We identified several themes related to people's discovery and interest in health apps including the importance of personal use before adoption, desired features, and trusted sources. We discuss each of these themes along with related subthemes below.

Trusted Sources

One important source of information about which app to use was to lean on the recommendations of "trusted sources." However, participants offered very different definitions of what a trusted source might be. Many participants identified "trusted sources" as people that they have an ongoing relationship with, be it a friend, colleague, or health care provider. For example, one participant stated:

If I'm gonna spend actual money or even stuff like that, I would want at the very minimum a recommendation from a friend, a person I trust, somebody saying, "I really like this one." [P8, FG2]

These participants indicated that such people might be more likely to make recommendations that reflect their preferences or needs or built off of something that had worked for them in the past.

However, participants also acknowledged the importance of professional or advocacy organizations in leading people toward effective products because of the perception that such groups would present less biased views or based recommendations on consensus and reviews of a variety of different apps:

I think it would be helpful, too, to have like the American Psychiatric Association or something, one of those, the licensure bodies or whatever—if they had official recommendations or backing, that would be nice to know. [P12, FG3]

Finally, people generally indicated that connections to academic institutions or medical centers boost the credibility of apps. However, there were also some concerns about whether such organizations could produce apps that would provide the desired levels of usability and user experience. As one participant put it:

It's a medical institution that made this app? It's gonna be super shitty and really hard to deal with? [P7, FG2]

Therefore, although such institutions may get a benefit from potential users in terms of expectations regarding effectiveness and safety, the trade-off in negative expectations toward usability and user experience means that these institutions need to ensure that they are comparable with similar apps created by other developers. In the end, however, although participants indicated such trusted sources were useful to inform initial uptake of apps, they seemed less important in supporting their long-term use.

Personal Use Guides Adoption

Despite the varied sources, participants reported they would rely on to make decisions regarding downloading apps, ultimately their own impressions and use tended to drive adoption. As such, in the searching phase, participants reported that other user reviews or screenshots were some of the most persuasive. One participant said:

The screenshots are probably going to be as important [as the developer], to kind of just see what the user interface is. [P19, FG5]

Several participants commented on capabilities on both the Google Play and Apple iOS store to be able to see screenshots of apps and commented that these screenshots were extremely helpful to get a feel of user interface elements that would guide their decision to download the app.

However, a common theme for personal use is that many people do not simply pick one app and then use it. Instead, it was common for participants to report identifying multiple apps, downloading several, and then trying out those apps to be able to do direct head-to-head comparisons:

I have a tendency to go find many other apps of the same thing, and decide which one I like, to be honest with you. [P15, FG4]

Participants noted that this was useful because with many of these apps, they were not sure which features they were looking for until they used it, and aspects about aesthetics, usability, or usefulness of particular features would become more apparent when it could be compared with other options.

Although cost was not a deterrent for participants, many participants did mention the need to preview the app before committing to pay for it. As P28, FG7 put it “free always wins” with the ability to view content before any purchases being a large factor in that decision. If there was a cost associated with an app, participants preferred in-app purchases or subscriptions that unlocked additional content compared with those that had even a small fee associated with it from the start:

So, if they don't have the free trial and they want money, I'm not even gonna look at it. I'm not gonna pay for something before I've gotten the chance to see if it's gonna work for me or not; free always wins. [P11, FG3]

It is worth noting with regards to cost that participants did have thoughts about the value of apps with ongoing costs such as subscriptions. Although participants reported that they would pay *some* ongoing cost for an app they perceived as useful, many participants voiced some sort of limit to how much they would be willing to spend. For example, one participant mentioned one app that:

...they gave the option to pay \$50.00 a year. And I did that, because I liked the idea of what they were trying to do, kind of create a social community of people. [P3, FG1]

Other participants stated:

I wouldn't spend \$100.00 on any app for a year. [P13, FG3]

...well, no, I'm not likely to buy a \$60.00 a year app. Screw that. Never mind. [P7, FG2]

Thus, although free may be a strong determinant of an initial decision to at least download and try an app, cost might figure differently when long-term use and benefit is considered.

Features

In general, participants wanted apps that were useful, easy to use, and aesthetically pleasing. Across participants, there were commonly reported desired features within apps including tracking, analytics (eg, reports and insights based on tracked data), data sharing, and notifications. Data sharing referred to opportunities to send and share data with others either directly, through social features or social media, or to other apps. In fact, participants saw apps' ability to function for multiple uses or to connect to other data sources as related to usefulness and ease of use as it could reduce the burden for the user for data entry or increase meaningfulness of data through connections to other information. One of the most commonly discussed apps during the focus groups was *Clue*, a period and ovulation tracker. Many participants commented on *Clue*'s ability to track a variety of symptoms related to one's mental state such as focus, distraction, calm, and stress. This ability was useful to make connections between one's cycle and mental health and to notice other patterns in one's mental health more generally.

Usability was a major concern of participants that tended to differentiate those apps that would enjoy long-term use to those that would be quickly discarded. One subtheme within usability was the discoverability of different features. Many participants decried complicated multifeatured apps with “busy” home screens and the need to go through several screens. One participant said:

And for me, it's just too overwhelming and too discombobulating. I just want to tap in and get the information that I need without clicking and searching for dear life. [P14, FG4]

Another usability subtheme was the intuitiveness of apps, either through using paradigms or models that were similar to other commonly used apps or using language or visual elements that made the app quick to learn and use. The last subtheme in usability was bugs and technical difficulties. Many participants reported many apps associated with medical institutions have issues such as crashes, poor displays on their devices, or high demands on their phones memory. Usability was also strongly related to the other theme of personal use guiding adoption; participants reported that they were not willing to work through an app with significant usability concerns even if they could see it being beneficial.

Finally, participants preferred visually appealing apps, although the sentiment of P13, FG3 that “It has to be cute” was not universal among our participants, many commented on different aspects of aesthetics including color schemes, images, and the use of visual metaphors.

Discussion

Principal Findings

The large number of mental health apps means that consumers are faced with a considerable challenge to find any particular app. As consumer strategies for finding and selecting apps will likely bias downloads and use toward particular products, it is important to know how and why consumers make their selections. The results of our survey and focus group were largely consistent showing that, in general, content within apps (eg, aesthetics, features, and functionality) was the largest determinant to encourage people to download and use health apps. Although notions of credibility and issues of privacy and security were important, these aspects were often assumed to be present when “trusted sources” were involved in app development. There were also places where the results were potentially discrepant, which highlights some interesting areas for future work. For example, the survey data revealed one hesitation to adoption was uncertainty regarding the effectiveness of digital tools, but the focus group participants did not seem to think that research evidence was extremely compelling. It is possible that our survey responses reflect a broad question on whether digital tools for mental health could even reasonably be effective, especially with some concerns on the mental health impact of technologies such as smartphones and social media more broadly.

It is worth noting that people rely on relatively informal means of identifying apps, relying on Web searches, social media, and word of mouth. Consumer strategies are not wholly different from strategies recommended by Boudreaux et al [4] with the exception that consumers were unlikely to review the scientific literature and do not have professional connections to rely on. As such endorsements, such as that Apple made of *Calm* for the app of the year [21], might have a strong impact on people's uptake and use of such products. However, it is worth noting that such endorsements are not based on research evidence and as such do not necessarily mean that such products are the most beneficial. Future work should help promote standards related to the promotion of health apps to ensure that effective tools make their ways into the hands of consumers. Furthermore, our

focus groups identified “trusted sources” as a strong influence of people's decision to use tools, but the survey results participants rarely received information about apps from professional sources such as providers. It is useful to consider ways to better involve providers in conversations around mobile apps. This might involve learning from providers what types of tools they would be interested in recommending or training providers as to what tools are safe, efficacious, and evidence-based. Indeed, some evidence suggests that people are likely to follow-up on their providers' recommendations, especially in mental health [11], and aligning consumers' needs and preferences with providers' knowledge and recommendations will likely be a key piece of adoption in practice.

In terms of adoption of apps, a considerable amount of discussion in the focus groups revolved around early use and especially the first-time user experience (or what is referred to as the “FTUX”). Some apps mitigated this concern by relegating more advanced features or content to premium versions, which had the added benefit of a revenue stream for the app. Aspects that users were especially mindful of in the early experience were the usability, aesthetics, and visual interface elements. It is worth noting that for many mental health apps, it is unlikely that a single use would lead to the proposed benefit (eg, reduced depression or anxiety), as addressing many mental health concerns requires sustained behavior change. In light of this, developers should consider how to give users appropriate previews of the apps that not only give a sense of the look and feel of the app and the functionality of its feature but also whether or not the app is likely to lead to the proposed benefit. A stronger focus on the proximal outcomes of success early in the app journey might be critical for setting appropriate expectations and promoting long-term use.

Relatedly, it is worth noting that although users were strongly motivated by information about whether or not an app would help them, this information was not necessarily research evidence. As such, even though researchers have noted that many apps are not based on evidence-based principles [22] and that most have not undergo rigorous research evaluation [23], this is unlikely to impact consumer behavior. Consumers were much more likely to be interested in a variety of sources of information about expected benefit, including user reviews, anecdotes from friends or family members, or doctor recommendations. Therefore, it might be useful to conduct more structured evaluations leveraging similar logic (eg, n-of-1 designs, multiple baselines) to expand the type of evidence that consumers are looking for. Currently, several efforts to identify and review mental health apps exist (for a review, see the paper by Neary and Schueller [24]); it is possible that such reviews or certification from such sources could represent a “trusted source” of information, but we did not evaluate this directly in our survey or focus groups.

An interesting consideration is if one app could accomplish everything users need in the mental health space. Several participants commented positively on interoperability that allowed connection to other apps or leveraged features such as Apple's Health or Google Fit. However, the downside to having an app accomplish multiple features is increased complexity

that might reduce usability. As 2 examples of how this could be addressed, we can look at the Department of Defense apps (eg, PTSD Coach mentioned earlier) or IntelliCare [10]. The Department of Defense has produced several apps, many of which have similar interface elements or features. IntelliCare, on the other hand, has several apps that are interconnected through a hub app. The example of Clue provides another potential alternative in which mental health features are integrated into another health app. It could be that participants talked about health and mental health in this way because we asked questions specifically about health apps and mental health apps. However, it is worth noting that several common health apps, such as apps for physical activity and diet, might have connections to mental health, given the research showing linkages between these areas [25-27]. Apps for sleep may also have an important role to play in mental health, given that sleep is a common symptom of many mental health issues [28,29] and a target of several mental health interventions in both traditional [30] and digital formats [31]. Indeed, participants in our focus groups often cycled back to references to health apps while discussing mental health apps to find examples, make comparisons, or discuss commonalities. Not surprisingly, people talked about health and mental health apps similarly; the distinction between mental health apps and health apps might be more salient to mental health researchers and practitioners than to consumers. Going a step further, it could be that the wave of the future is neither health nor mental health apps, but more mental health app features integrated into commonly used apps like our calendars, conversational agents, or messaging platforms or apps focused on health more broadly. Integrating mental health in the operating system or basic applications would reduce the need to seek and find apps but would concentrate power in the hands of fewer developers who would then have a stronger influence on our health and mental health.

Limitations

These findings, however, were not without their limitations, which are worth acknowledging to ensure that conclusions are accurately represented. First, participants in our study came from an established research registry and might not be representative of the more general population. Individuals willing to enroll in a research registry may be different from participants in other research studies because of their interest

to participate in multiple research studies and willingness to be recontacted. The response rate of 18% was somewhat below average for Web-based surveys [32]; however, we did suspend recruitment after 8 weeks of data collection, which might have artificially reduced total participation. Still, our respondents might be biased toward those who are likely to respond more quickly, but it is unclear how such a characteristic might influence the results. Our respondents did tend to be well-educated and female; however, this also mirrors the characteristics of people who tend to enroll in research studies [33] and use technology-based mental health tools more broadly [34]. Furthermore, this registry, in particular, was connected with a group focused on research in digital mental health and as such participants may have more knowledge of digital mental health apps than other people. Nevertheless, given that apps are not currently widely recommended by health professionals, it is likely that interest drives adoption and this population might be representative of likely users. Another limitation is that it is unclear if when people were responding to survey questions, they were being descriptive (ie, how they would characterize the apps they use now) or prescriptive (ie, what they would like to see in apps that they would download and use). Therefore, we cannot conclude if answers relate to the current state of health and mental health apps or what consumers would be interested in seeing in future health and mental health apps.

Conclusions

Mental health apps are a rapidly growing area with little indication that the speed of development will slow down. Although regulatory developments like approval from the United States Federal Drug Administration might impact their adoption, advances in this area are still too new to fully understand their long-term impact for both the marketplace and consumers. As such, better understanding factors that drive people's decisions to download and use apps is an important step toward sustainable and impactful benefits from such technologies. Our findings highlighted a number of factors—ease of use, aesthetics, and individual experience—and also indicated that evidence-base and usefulness are not equivalent in the eyes of consumers. These findings can inform aspects of the design and dissemination of such products and hopefully impact efforts to ensure consumers get trusted and effective products.

Acknowledgments

SMS receives funding from One Mind to direct and lead PsyberGuide, a nonprofit project focused on identifying and evaluating mental health apps. RedCap access was made possible through Northwestern University's Clinical and Translation Sciences Institute (NUCATS) through funding from the National Center for Advancing Translation Sciences, Grant number UL1TR001422. SMS is an investigator with the Implementation Research Institute (IRI) at the George Washington University in St. Louis through an award from the National Institute of Mental Health (5R25MH08091607) and the Department of Veterans Affairs, Health Services Research & Development Service, Quality Enhancement Research Initiative (QUERI).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Focus group scheduling, including semistructured interview.

[[PDF File \(Adobe PDF File\), 68KB - jmir_v20i6e10141_app1.pdf](#)]

References

1. Research2Guidance. 2017. 325,000 mobile health apps available in 2017 - Android now the leading mHealth platform URL: <https://research2guidance.com/325000-mobile-health-apps-available-in-2017> [accessed 2018-02-15] [WebCite Cache ID 6xFvItqJA]
2. Van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. *Depress Anxiety* 2017 Jun;34(6):526-539. [doi: [10.1002/da.22657](#)] [Medline: [28569409](#)]
3. 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](#)]
4. 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](#)]
5. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. *Mil Med* 2014 Jan;179(1):12-18. [doi: [10.7205/MILMED-D-13-00271](#)] [Medline: [24402979](#)]
6. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. *J Consult Clin Psychol* 2017 Mar;85(3):267-273. [doi: [10.1037/ccp0000163](#)] [Medline: [28221061](#)]
7. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: A pilot randomized controlled trial with community trauma survivors. *Psychol Trauma* 2016 May;8(3):384-392. [doi: [10.1037/tra0000092](#)] [Medline: [27046668](#)]
8. 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](#)]
9. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the Wild: Using Novel Data to Examine the Reach, Use, and Impact of PTSD Coach. *JMIR Ment Health* 2015 Mar 25;2(1):e7 [FREE Full text] [doi: [10.2196/mental.3935](#)] [Medline: [26543913](#)]
10. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: An Eclectic, Skills-Based App Suite for the Treatment of Depression and Anxiety. *J Med Internet Res* 2017 Jan 05;19(1):e10 [FREE Full text] [doi: [10.2196/jmir.6645](#)] [Medline: [28057609](#)]
11. IMS Institute for Healthcare Informatics. 2015 Sep. Patient adoption of mHealth: Use, evidence and remaining barriers to mainstream acceptance URL: http://moodle.univ-lille2.fr/pluginfile.php/215343/mod_resource/content/0/IHI_Patient_Adoption_of_mHealth.pdf [WebCite Cache ID 6xFwMgV1z]
12. Huang HY, Bashir M. Users' Adoption of Mental Health Apps: Examining the Impact of Information Cues. *JMIR Mhealth Uhealth* 2017 Jun 28;5(6):e83 [FREE Full text] [doi: [10.2196/mhealth.6827](#)] [Medline: [28659256](#)]
13. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Manag Sci* 2000 Feb;46(2):186-204. [doi: [10.1287/mnsc.46.2.186.11926](#)]
14. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. *MIS Quarterly* 2003 Sep;27(3):425-478 [FREE Full text]
15. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR Mhealth Uhealth* 2015 Nov 04;3(4):e101 [FREE Full text] [doi: [10.2196/mhealth.4924](#)] [Medline: [26537656](#)]
16. National Institute of Mental Health/National Institutes of Health. Opportunities and Challenges of Developing Information Technologies on Behavioral and Social Science Clinical Research URL: <https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/opportunities-and-challenges-of-developing-information-technologies-on-behavioral-and-social-science-clinical-research.shtml> [accessed 2018-02-15] [WebCite Cache ID 6xFwTqmnW]
17. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen TQ, Luo J, et al. A Hierarchical Framework for Evaluation and Informed Decision Making Regarding Smartphone Apps for Clinical Care. *Psychiatr Serv* 2018 May 01;69(5):498-500. [doi: [10.1176/appi.ps.201700423](#)] [Medline: [29446337](#)]
18. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A Comprehensive Quality and Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions. *J Med Internet Res* 2017 Mar 21;19(3):e82 [FREE Full text] [doi: [10.2196/jmir.7270](#)] [Medline: [28325712](#)]
19. Psyberguide. URL: <https://psyberguide.org/> [accessed 2018-05-11] [WebCite Cache ID 6zKkAmhKI]
20. Guest M, MacQueen KM, Namey EE. Applied thematic analysis. Los Angeles, CA: Sage; 2012.

21. Slate. 2017 Dec 9. Apple's App of the Year Makes Sense If You've Been Paying Attention to the News URL: http://www.slate.com/articles/technology/technology/2017/12/apple_app_store_calm_app_of_the_year_is_a_fitting_choice_for_2017.html [WebCite Cache ID 6xFwYyZj6]
22. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth* 2015 Feb 16;3(1):e16 [FREE Full text] [doi: [10.2196/mhealth.3713](https://doi.org/10.2196/mhealth.3713)] [Medline: [25689790](https://pubmed.ncbi.nlm.nih.gov/25689790/)]
23. Buijink AW, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. *Evid Based Med* 2013 Jun;18(3):90-92. [doi: [10.1136/eb-2012-100885](https://doi.org/10.1136/eb-2012-100885)] [Medline: [22923708](https://pubmed.ncbi.nlm.nih.gov/22923708/)]
24. Neary M, Schueller SM. State of the Field of Mental Health Apps. *Cogn Behav Pract* 2018;1-7. [doi: [10.1016/j.cbpra.2018.01.002](https://doi.org/10.1016/j.cbpra.2018.01.002)]
25. Stathopoulou G, Powers MB, Berry AC, Smits JAJ, Otto MW. Exercise Interventions for Mental Health: A Quantitative and Qualitative Review. *Clin Psychol Sci & Pract* 2006 May;13(2):179-193. [doi: [10.1111/j.1468-2850.2006.00021.x](https://doi.org/10.1111/j.1468-2850.2006.00021.x)]
26. Paluska SA, Schwenk TL. Physical activity and mental health: Current concepts. *Sports Med* 2000 Mar;29(3):167-180. [Medline: [10739267](https://pubmed.ncbi.nlm.nih.gov/10739267/)]
27. O'Neil A, Quirk SE, Housden S, Brennan SL, Williams LJ, Pasco JA, et al. Relationship between diet and mental health in children and adolescents: a systematic review. *Am J Public Health* 2014 Oct;104(10):e31-e42. [doi: [10.2105/AJPH.2014.302110](https://doi.org/10.2105/AJPH.2014.302110)] [Medline: [25208008](https://pubmed.ncbi.nlm.nih.gov/25208008/)]
28. Harvey AG. Insomnia: symptom or diagnosis? *Clin Psychol Rev* 2001 Oct;21(7):1037-1059. [Medline: [11584515](https://pubmed.ncbi.nlm.nih.gov/11584515/)]
29. Ford DE, Kamerow DB. Epidemiologic study of sleep disturbances and psychiatric disorders. An opportunity for prevention? *JAMA* 1989 Sep 15;262(11):1479-1484. [Medline: [2769898](https://pubmed.ncbi.nlm.nih.gov/2769898/)]
30. Taylor DJ, Pruiksma KE. Cognitive and behavioural therapy for insomnia (CBT-I) in psychiatric populations: A systematic review. *Int Rev Psychiatry* 2014 Apr;26(2):205-213. [Medline: [24892895](https://pubmed.ncbi.nlm.nih.gov/24892895/)]
31. Baron KG, Duffecy J, Reid K, Begale M, Caccamo L. Technology-Assisted Behavioral Intervention to Extend Sleep Duration: Development and Design of the Sleep Bunny Mobile App. *JMIR Ment Health* 2018 Jan 10;5(1):e3 [FREE Full text] [doi: [10.2196/mental.8634](https://doi.org/10.2196/mental.8634)] [Medline: [29321122](https://pubmed.ncbi.nlm.nih.gov/29321122/)]
32. Cook C, Heath F, Thompson RL. A meta-analysis of response rates in Web- or Internet-based surveys. *Educ Psychol Meas* 2000;60(6):821-836. [doi: [10.1177/00131640021970934](https://doi.org/10.1177/00131640021970934)]
33. Curtis R, Presser S, Singer E. The effects of response rate changes on the index of consumer sentiment. *Public Opin Q* 2000;64(4):413-428. [Medline: [11171024](https://pubmed.ncbi.nlm.nih.gov/11171024/)]
34. Lindner P, Nyström MB, Hassmén P, Andersson G, Carlbring P. Who seeks ICBT for depression and how do they get there? Effects of recruitment source on patient demographics and clinical characteristics. *Internet Interv* 2015 May;2(2):221-225. [doi: [10.1016/j.invent.2015.04.002](https://doi.org/10.1016/j.invent.2015.04.002)]

Abbreviations

CBITs: Center for Behavioral Intervention Technologies

TAM: technology acceptance model

UTAUT: unified theory of acceptance and use of technology

Edited by R Calvo, M Czerwinski, J Torous, G Wadley; submitted 15.02.18; peer-reviewed by T Fleming, V Strotbaum, J Meng; comments to author 08.03.18; revised version received 04.04.18; accepted 05.04.18; published 11.06.18.

Please cite as:

Schueller SM, Neary M, O'Loughlin K, Adkins EC

Discovery of and Interest in Health Apps Among Those With Mental Health Needs: Survey and Focus Group Study

J Med Internet Res 2018;20(6):e10141

URL: <http://www.jmir.org/2018/6/e10141/>

doi: [10.2196/10141](https://doi.org/10.2196/10141)

PMID: [29891468](https://pubmed.ncbi.nlm.nih.gov/29891468/)

©Stephen M Schueller, Martha Neary, Kristen O'Loughlin, Elizabeth C Adkins. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.06.2018. 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.

Original Paper

Detecting Recovery Problems Just in Time: Application of Automated Linguistic Analysis and Supervised Machine Learning to an Online Substance Abuse Forum

Rachel Kornfield^{1*}, MA; Prathusha K Sarma^{2*}, MS; Dhavan V Shah¹, PhD; Fiona McTavish³, MA; Gina Landucci³, BS; Klaren Pe-Romashko³, MS; David H Gustafson³, PhD

¹School of Journalism and Mass Communication, University of Wisconsin-Madison, Madison, WI, United States

²Department of Electrical & Computer Engineering, University of Wisconsin-Madison, Madison, WI, United States

³Center for Health Enhancement System Studies, University of Wisconsin-Madison, Madison, WI, United States

*these authors contributed equally

Corresponding Author:

Rachel Kornfield, MA

School of Journalism and Mass Communication

University of Wisconsin-Madison

5115 Vilas Hall

821 University Avenue

Madison, WI, 53706

United States

Phone: 1 4153355356

Fax: 1 608 890 1438

Email: rkornfield@gmail.com

Abstract

Background: Online discussion forums allow those in addiction recovery to seek help through text-based messages, including when facing triggers to drink or use drugs. Trained staff (or “moderators”) may participate within these forums to offer guidance and support when participants are struggling but must expend considerable effort to continually review new content. Demands on moderators limit the scalability of evidence-based digital health interventions.

Objective: Automated identification of recovery problems could allow moderators to engage in more timely and efficient ways with participants who are struggling. This paper aimed to investigate whether computational linguistics and supervised machine learning can be applied to successfully flag, in real time, those discussion forum messages that moderators find most concerning.

Methods: Training data came from a trial of a mobile phone-based health intervention for individuals in recovery from alcohol use disorder, with human coders labeling discussion forum messages according to whether or not authors mentioned problems in their recovery process. Linguistic features of these messages were extracted via several computational techniques: (1) a Bag-of-Words approach, (2) the dictionary-based Linguistic Inquiry and Word Count program, and (3) a hybrid approach combining the most important features from both Bag-of-Words and Linguistic Inquiry and Word Count. These features were applied within binary classifiers leveraging several methods of supervised machine learning: support vector machines, decision trees, and boosted decision trees. Classifiers were evaluated in data from a later deployment of the recovery support intervention.

Results: To distinguish recovery problem disclosures, the Bag-of-Words approach relied on domain-specific language, including words explicitly linked to substance use and mental health (“drink,” “relapse,” “depression,” and so on), whereas the Linguistic Inquiry and Word Count approach relied on language characteristics such as tone, affect, insight, and presence of quantifiers and time references, as well as pronouns. A boosted decision tree classifier, utilizing features from both Bag-of-Words and Linguistic Inquiry and Word Count performed best in identifying problems disclosed within the discussion forum, achieving 88% sensitivity and 82% specificity in a separate cohort of patients in recovery.

Conclusions: Differences in language use can distinguish messages disclosing recovery problems from other message types. Incorporating machine learning models based on language use allows real-time flagging of concerning content such that trained staff may engage more efficiently and focus their attention on time-sensitive issues.

(*J Med Internet Res* 2018;20(6):e10136) doi:[10.2196/10136](https://doi.org/10.2196/10136)

KEYWORDS

self-help groups; substance-related disorders; supervised machine learning; social support; health communication

Introduction

Background

Digital health interventions have proliferated in recent years [1], and evidence suggests they can improve management of mental health issues, including substance use disorders (SUDs) [2,3]. Once their design is established, digital interventions can be disseminated with less expense and effort than face-to-face ones [4]. Such scalability is crucial for addressing SUDs, as demand for treatment dramatically outstrips available services [5]. In addition, although SUDs are chronic and relapsing [6,7], the help conveyed through technologies is ongoing and accessible. One recent clinical trial demonstrated that, relative to a control group, individuals who accessed a mobile phone-based recovery system reported reduction in risky drinking days by more than half over a year [8].

Substantial human labor also supports many effective digital health interventions. Some evidence suggests that, relative to interventions that lack human guidance, those that combine computerized tools with human support and coaching can enhance engagement and improve effectiveness of interventions [9]. The need for human expertise extends to interventions featuring peer-to-peer communication. Digital peer-to-peer interventions have involved “moderators” in various ways, including spurring and guiding discussion; monitoring forums for problematic content; and, crucially, providing just-in-time support to patients who are struggling, including through escalating contact or recommending treatment [10,11]. Through just-in-time support, moderators contribute to efficiency of health services at a systems level, making additional attention and resources available to those who most need them, while maintaining less intensive support for those at a lower risk level. Yet, attending to changing needs of an online health community poses a considerable challenge as participants can produce a massive volume of text exchanges [12]. Demands on staff represent a key hurdle in scaling up digital health interventions [13]. In this paper, we describe how automated linguistic analysis of text-based exchanges, and supervised machine learning, may play a role in managing moderator workflow in a technology-based recovery support system.

Our approach builds on the power of language as a signal of mental health risk, with linguistic cues being increasingly discernable through computational methods. Over the past several decades, researchers have amassed an extensive body of literature showing the promise of language to reveal individuals’ psychological traits, thoughts, feelings, and likely behaviors [14], including in social media contexts [15]. As similar ideas can be conveyed in different ways, individuals’ risk profiles emerge not only from the explicit content of their communication (ie, what topics authors are talking or writing about) but also from the style of their language (ie, *how* authors say what they say). In this study, we investigate how recovery challenges may emerge both through the individual words that authors use within a discussion forum as well as through general

psycholinguistic dimensions of their messages (eg, affect, cognitive mechanisms), as captured through a dictionary-based approach. Leveraging these linguistic features, our goal is to find classifiers that can accurately label messages as conveying or not conveying recovery problems, allowing us to prioritize this content for review and intervention.

We consider several computational linguistic and machine learning approaches. First, we extract linguistic features of messages using 3 techniques: (1) a Bag-of-Words (BoW) approach representing each message in terms of word occurrences, (2) the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker Conglomerates, Austin, TX) [16], which computes rates of language use within validated dictionaries corresponding to psychological and linguistic concepts, and (3) a novel hybrid approach combining important features from BoW and LIWC. We propose that BoW and LIWC have complementary strengths, with BoW attending to important words specific to the dataset (eg, those related to substance use), whereas LIWC attends to relevant psychological states (eg, anxiety, self-focus). We expect that a hybrid approach, capitalizing on the strengths of each, should outperform either LIWC or BoW. We test these techniques in the context of supervised machine learning models that have been utilized in social media contexts: support vector machines (SVM), decision trees, and boosted decision trees.

In interpreting performance of our computational linguistic and machine learning approaches, we consider some particularities of the domain of addiction recovery support, namely: (1) a low tolerance for false negatives, (2) a preference for understandability of the method to stakeholders, and (3) efficiency in processing language and classifying messages in real-time. In other words, in addition to considering overall accuracy of each classifier, we ask: Does it miss too many worrisome messages to be useful to forum moderators? Does it have face validity to a team of health professionals? And can we successfully implement it in real-time? To establish the utility and robustness of our approach, we test our classifiers in a separate iteration of our mobile intervention involving a cohort of primary care patients with SUDs. We finally discuss implications of our findings for future research and system design, including how to improve model performance, and how classification can serve as the basis for directing attention and resources to those who need them.

Online Support Forums

SUDs are among the most common mental health disorders in the United States, with over 20 million adults affected in 2013 [5]. SUDs precipitate distress for sufferers and their communities, as well as serious health consequences [17,18]. Although many individuals with SUDs make attempts to stop using substances, resumption of risky substance use, or relapse, is extremely common [19]. With intensive SUDs treatments being time-limited, it is crucial to find ways to extend recovery support to prevent relapse in the long term [20].

Mobile phones and internet use are now ubiquitous in the United States [21], with one consequence being that individuals with mental health challenges can access social support despite physical distance and at any time of day [22]. Often, this support comes from others who share the same mental health concern, as occurs via digital peer-to-peer forums where participants seek help on an “as needed” basis, and provide it to others [23,24]. Such forums typically involve anonymity or pseudo-anonymity, allowing for candid disclosure of personal and stigmatized issues and experiences [25]. Content analyses show that participants in SUDs forums disclose a variety of recovery challenges, prompting exchange of informational and emotional support [26-28].

The Role of Moderators

Although discussion forums offer a valuable arena for peer-to-peer exchange, moderators can also play a key role. For instance, those in recovery must manage their *own* health issues, limiting the time and energy that can be applied to help others [29]. In addition, although peers can offer first-hand experience related to coping and recovery, they may lack expertise necessary to guide decision making about clinical issues [30]. In contrast, moderators often have knowledge of intervention components and health behavior change processes and may recognize instances where contact or treatment is appropriate [31,32,10]. Moderators may additionally engage in pseudo-therapeutic activities such as offering emotion-focused support or assisting participants in reassessing dysfunctional perspectives [11] and may be more effective than peers in motivating individuals earlier in their behavior change process [33].

The presence of moderators in digital health forums has been associated with benefits. Notably, studies have found greater participation and expressiveness in moderated health forums relative to unmoderated ones [34,35]. In a mobile SUDs intervention for drug court participants, trained staff played a central role in discussion networks, with many participants communicating only with staff [36]. Prior work has also found that staff can enhance the success of digital mental health interventions regardless of formal clinical training [37].

Machine Learning Applications to Moderator Engagement

To “scale up” digital interventions, designers must take steps to support and streamline moderators’ work. Fortunately, such efforts can make use of extensive data generated as participants engage with technologies. A rapidly growing research area centers on leveraging the digital traces of participants’ activities to gain insights into the changing contexts within which participants are embedded and the psychological states they experience [38,39]. Digital trace data collected through mobile phones may include sensor data (eg, geolocation, accelerometry), as well as patterns of engagement with the intervention itself, and the content of messages exchanged.

By capturing spontaneous, first-hand accounts of authors’ beliefs, feelings, and experiences, text-based messages offer particularly powerful insights into wellness, including the risk of mental health-related outcomes [15]. For instance, prior

research has shown that linguistic qualities such as self-focus (as conveyed in pronoun use) can distinguish those who go on to post about suicidal ideation [40], and that negative affective language and swearing can identify individuals who go on to relapse in alcohol recovery [41]. These approaches rely on automated linguistic analysis as described in greater detail below.

Text-based features of user-submitted messages can now be efficiently extracted through a range of computational approaches. One of the most common approaches, BoW, involves representing each message in terms of occurrences of individual words, or “unigrams.” After throwing out extremely common words, and grouping together words with the same stem, a message is represented as a vector formed by the occurrence rate of each stem, relative to that stem’s overall occurrence in the full set of messages. In contrast, dictionary-based approaches search within a message for lists of words corresponding to relevant concepts. For instance, LIWC searches for words representing discussion topics (eg, health, family), psychological dimensions (eg, affect, cognition), and linguistic characteristics (eg, pronouns, conjunctions). LIWC then computes the percent of words in a given message that fall in each category. LIWC has been widely used in research, with studies showing that its categories predict health-related states including suicidality, depression, and dementia [42-44].

Relevant to this study, recent work also uses the above approaches to detect self-disclosure in online forums, defined as messages wherein participants convey personally relevant thoughts, feelings, and experiences [45-47]. In the context of support forums, self-disclosures offer a promising opportunity for intervention (eg, by moderators), because participants are revealing and working through personal issues, and may be actively seeking help [46]. The prior literature suggests that self-disclosure messages involve telltale linguistic cues that aid automatic detection. For example, one study identified a number of LIWC categories predictive of self-disclosure sensitivity, including third person pronouns and discussion of family, sex, death, and negative affect [48]. In another study, individual words conveying affect (eg, “happy,” “love,” and “hate”) were characteristic of mental health-related self-disclosure [49].

Human expertise can also play an important role in guiding the development of language-based models. In supervised machine learning, an expert will designate a subset of messages as belonging to a category of interest (such as mental health risk), and the features of labeled messages are then used to predict whether an unlabeled message would fall in the same category. Labeled data can be generated in a number of ways. For instance, naturally occurring response patterns can be used, such as where Huh and colleagues [12] labeled as problematic those messages to which moderators had previously responded in a health support forum, using their linguistic features to classify new messages that moderators would likely be interested in. Alternately, human judgment can be used to generate each label in the training set, as was implemented in efforts to detect suicidality in an online discussion forum for youth [50,51]. This approach recognizes that moderators’ response patterns do not always clearly follow from the level risk a message indicates. For instance, responding to a message need not represent

concern, but could reflect interest in a particular topic or investment in an ongoing relationship. Using SVM, boosted decision trees, and other models, researchers were able to achieve F-scores over 0.9 in identifying messages urgently requiring response [50,51].

This Study

In online addiction recovery forums, messages can be posted at any time of day or night, and some convey serious or time-sensitive problems. To offer timely help, moderators must continually review new content, but this task is demanding, with potentially dozens or hundreds of new posts to consider every day. Therefore, this study focuses on automatic detection of messages suggesting risk. For instance, interpersonal conflict, legal issues, personal traumas, or encounters with substance use cues could all represent threats to recovery in a substance abuse context [52,53]. Furthermore, these events could inspire psychological states associated with relapse, such as negative affect, cravings, or reduced self-efficacy [54]. Although circumstances and states can be conveyed in a variety of ways, prior literature leads us to anticipate that common language elements should emerge making recovery problems amenable to detection.

This paper contributes to the literature on digital SUDs interventions on several fronts. First, this work has practical application to efficiently capturing concerning content, so that forum moderators can respond in time. Efficient engagement by experts has been identified as a priority for extending the effectiveness of digital mental health interventions [13]. Second, the methodological contribution of our work involves comparing common computational linguistics and machine learning approaches and determining which are suited to the context of mental health risk in support forums.

As far as linguistic analysis, we compare performance of 2 techniques and their hybridization. First, BoW is driven by word-level usage in a given dataset and may therefore have an advantage for recovery-specific words (eg, “drink”). In contrast, our dictionary-based approach, LIWC, characterizes messages along general psychological and linguistic dimensions. Through building on prior knowledge about how words relate to established psychological constructs, LIWC offers potential efficiency, interpretability, and theoretical traction; however, its distinct disadvantage is that its dictionaries are not recovery specific. Thus, although LIWC contains a general category for “health,” it lacks dictionaries corresponding to concepts like “relapse” or “cravings.” Given these trade-offs, it is unclear whether BoW or LIWC will perform best.

Importantly, LIWC and BoW differ in their treatment of common words. The BoW framework retains words that are distinctive of the data at hand. Words with consistently high use across contexts, such as “I” and “we,” are considered insignificant within the BoW framework and typically discarded. In contrast, LIWC computes usage rates of these and other so-called “function” words (eg, pronouns, conjunctions, prepositions), which lack content but hold sentences together [55]. Despite their apparent banality, function words have proved powerful in predicting well-being, with pronouns receiving substantial attention in the mental health domain as

a gauge of social integration [40,56,57]. Not surprisingly, personal pronouns also indicate self-disclosure, as they can show that individuals are talking about themselves [49]. In comparing BoW with LIWC and hybrid approaches, we therefore pay particular attention to performance improvements related to function words. In a more general sense, we aim to identify linguistic features most central to manifesting recovery problems, including discussion of substance use triggers, affective states, cognitive processes, and function words.

We also attempt to identify well-performing machine learning approaches. We focus on decision trees with and without boosting, as well as SVM, approaches with good performance in prior social media data [58,59].

Finally, we consider our results in relation to several key features of the domain of recovery support. First, recovery support is an arena where false negatives may be problematic, as missing an opportunity to intervene could allow a problem to escalate, even precipitating relapse. Therefore, in generating gold standard data, we emphasize the importance of establishing a reliable definition of “recovery problems” that is broad enough to capture potentially concerning content. We also reflect our concern about false negatives by prioritizing sensitivity in weighing classifier performance. Second, we seek machine learning methods that can offer insights into the particular language patterns associated with recovery. Decision trees may have an advantage in this regard, as they provide a visualization of the mechanisms of classification that may be helpful to establish face validity among stakeholders [59]. Finally, computational linguistics approaches have different implications for implementing classification in real-time, which we discuss.

Methods

Intervention

Data for this study came from a mobile phone-based intervention that provides on-demand services for recovery maintenance and relapse prevention. These services include informational pages, self-management tools (eg, self-help meeting directories, surveys), and peer-to-peer discussion forums. The intervention has been described in detail elsewhere, and it demonstrated efficacy in reducing risky drinking days by more than half relative to a control group [8]. We used data from 2 studies of the system: (1) a clinical trial involving individuals discharged from alcohol treatment (study 1) [8] and (2) an implementation study in primary care, involving individuals who used either alcohol or illicit drugs (study 2) [60]. The institutional review board at the University of Wisconsin-Madison approved both studies. Study participants provided informed consent for collection and use of their data for research (not shared beyond the team). These data included a log of all uses of the intervention and the content of communications exchanged within the intervention.

Study participants were provided with a mobile phone loaded with the intervention: either the Palm Pre with the Palm OS (Palm, Inc, Sunnyvale, CA) or an HTC Evo running Android 4.4 (HTC Corporation, Taiwan). In study 1, 130 participants posted on the forum. They were 56.2% (73/130) male and had

a mean age of 38 years (SD 9.7). Participants wrote approximately 20 messages each (average length: 31 words). In study 2227 participants posted on the forum, and they were 53.3% (121/227) male, with a mean age of 42 years (SD 10.7). Participants wrote approximately 69 messages each (average length: 29 words).

This study focuses on text-based messages that were exchanged in the system's discussion forum, where participants could either start new threads on a topic of their choosing or respond to existing threads. All forum messages were visible to those on study, but study 1 forums were gender segregated. Moreover, 3 members of the research team also monitored the forums (authors GL, FM, and KP). Although the moderators lack clinical background, they are experts in digital health support for self-management of chronic conditions, including addiction recovery.

As mentioned earlier, gold standard data in this study substantially differ from those used in some prior work using moderators' natural response patterns [12]. We instead developed and applied a standardized, reliable codebook for capturing recovery risk. The first author first conducted an initial interview with the 3 moderators to understand their role in the forum and which messages would be considered worthy of intervention, and then consulted with them throughout the hand-labeling process to ensure our process captured messages of concern.

Computational Linguistics

We represented discussion forum messages using a BoW model, the LIWC program, and a hybrid approach.

The BoW approach represents each message in a feature space characterized by word counts. Common words were discarded, and remaining words were reduced to their stems using the Lancaster Stemmer from the NLTK stem package in Python and the NLTK word_punct tokenizer. For example, the stem "drink" would capture "drinking," "drinkin," "drinker," "drinks," and so on. We also wrote an additional filter to remove emoticons and other nonstandard characters. After grouping words according to their stems, Term Frequency-Inverse Document Frequency (TF-IDF) weighting was applied to calculate the occurrence rate of each specific stem in a message, offset by the importance of the stem in the entire corpus. Specifically, TF-IDF for a term is expressed as the term frequency (the number of times a word appears in a document divided by the total number of words in that document) multiplied by inverse document frequency (log of: total number of messages in a corpus divided by the number containing the term), thus adjusting for the fact that some words appear more frequently than others in general [61]. Once computed, the TF-IDF weights are used to form a vector representation of each message. After discarding common words, our BoW representation utilized 4247 unique unigrams as features.

The LIWC 2015 program computes rates of using words that fall within approximately 90 categories representing linguistic characteristics (eg, personal pronouns), topics of discussion (eg, family), affect (eg, anger), and cognitive processes (eg, insight) [16]. Each category corresponds to a predetermined dictionary

of related words and word stems. Therefore, each message is represented as a 90-dimensional vector, with each dimension corresponding to a category such as "pronouns" and "positive affect." The value in each dimension is computed as the number of words from the message belonging to that category divided by the total number of words in the message. For example, "personal pronoun" is one of the features scored by LIWC. In the message "I am doing well," 1 out of the 4 words are personal pronouns, and so the LIWC score would be 1 out of 4 words or 25%.

In a hybrid approach, we exploit linguistic features from both BoW and LIWC. In other words, for a given message, word frequencies of the most important features from the TF-IDF matrix and the percentages falling in the most important linguistic categories from LIWC are stacked together to form a single feature vector. Given that combining too many features can inhibit performance by introducing noise [62], we utilized a subset of features from each representation. After ranking features according to their importance for a random forest model [63], we picked up to 10% of the most relevant features from BoW and LIWC to form a new feature set. Feature importance is calculated using the Gini Impurity measure, defined as the sum across the number of splits over all trees containing a feature, divided by number of samples in each split [64]. The hybrid approach included 310 features.

Machine Learning Techniques

With numeric representations of each message in our training set, and a corresponding label (recovery problem or no recovery problem), we trained 3 candidate binary classifiers for our task: SVM, decision trees, and boosted decision trees. SVM is a widely used technique and involves defining an optimal hyperplane to distinguish between items falling in classes of interest [65]. Decision trees involve segmenting the feature space into a number of simple regions [66]. In a series of decision steps, represented as branches, observations are made about an item (eg, the frequency at which a particular word is used within the message), leading to corresponding conclusions about the appropriate class (represented in the leaves). Finally, a related approach, boosted decision trees, involves an ensemble of decision trees where each tree learns by fitting the residual of the trees before it, allowing iterative improvement in performance. Python scikit-learn was used for machine learning [67].

As our datasets feature unbalanced classes (ie, messages including "recovery problems" are outnumbered by messages without them), we compensated for this imbalance by oversampling from the minority class. Specifically, we used the Synthetic Minority Oversampling Technique to generate synthetic samples from the minority class [68]. Rather than creating exact copies, the algorithm samples 2 or more similar instances, with similarity being calculated by a distance measure, (eg, Euclidean, Cosine), and then slightly perturbs these instances to create synthetic samples.

Once our classes were balanced, we trained our classifiers using labeled training data from study 1, the clinical trial for those completing alcohol treatment (n=2581), and calculated parameters for each machine learning model using k-fold cross

validation. Next, we tested the best performing models in labeled messages sampled from study 2 ($n=800$) with its primary care population. We report F-scores, as well as sensitivity (the proportion of correctly identified true positives), specificity (the proportion of correctly identified true negatives), and area under curve (AUC). We also describe example decision trees that illustrate classification logic.

Results

Identifying Recovery Problems

Our conversations with moderators first revealed that they recognized a wide range of issues and circumstances as warranting a response (relationship troubles, cravings, etc). Moderators expressed a fear of missing an important message, reporting a preference to have uninteresting messages flagged (false positives) than to miss actual problems (false negatives). Supporting our strategy of hand-labeling problem messages versus using prior response patterns as gold standard data, moderators also reported that contextual considerations influence their likelihood of responding on the forum. For instance, they might be unlikely to respond if participants had already received competent help from peers, or if they had personally had recent contact with participants outside the forum (eg, by phone call or private message). Moderators also stressed that they sometimes miss concerning messages inadvertently.

Guided by this feedback, 3 coders independently reviewed a preliminary set of 200 messages to identify ones they thought disclosed recovery challenges, broadly construed, and then mutually discussed their decisions. Coders arrived at consensus around a rule for coding the entire dataset, when “the writer describes a potential threat to well-being or recovery efforts.” We further specified that the message may express either feeling vulnerable (eg, “I’ve been clean for about 7 months but even now I still feel like maybe I won’t make it”) or may outline a specific incident (eg, “it’s not looking good, they are talking 0 to 5, and that’s not days [in jail]. It’s got my head all f.... up.”). The coding rule also specified that the code should be applied even if the writer conveys that he or she has skills or abilities to handle a given problem (ie, a message may convey both a threat and mastery of that threat at the same time). Thus, by making the coding rule quite general, we avoided some subjectivity involved in making determinations about problems’ seriousness. The first author next overlapped with each other coder on a set of 100 messages, allowing computation of interrater reliability, with average Cohen kappa of .77 for the 2 overlap sets deemed acceptably high [69].

Table 1. F-scores reported by 3 classifiers on the test data from study 2.

Classifier	BoW ^a	LIWC ^b	Hybrid
SVM ^c	0.76	0.71	0.76
Decision tree	0.8	0.75	0.77
Boosted decision tree	0.8	0.83	0.85

^aBoW: Bag-of-Words.

^bLIWC: Linguistic Inquiry and Word Count.

^cSVM: support vector machines.

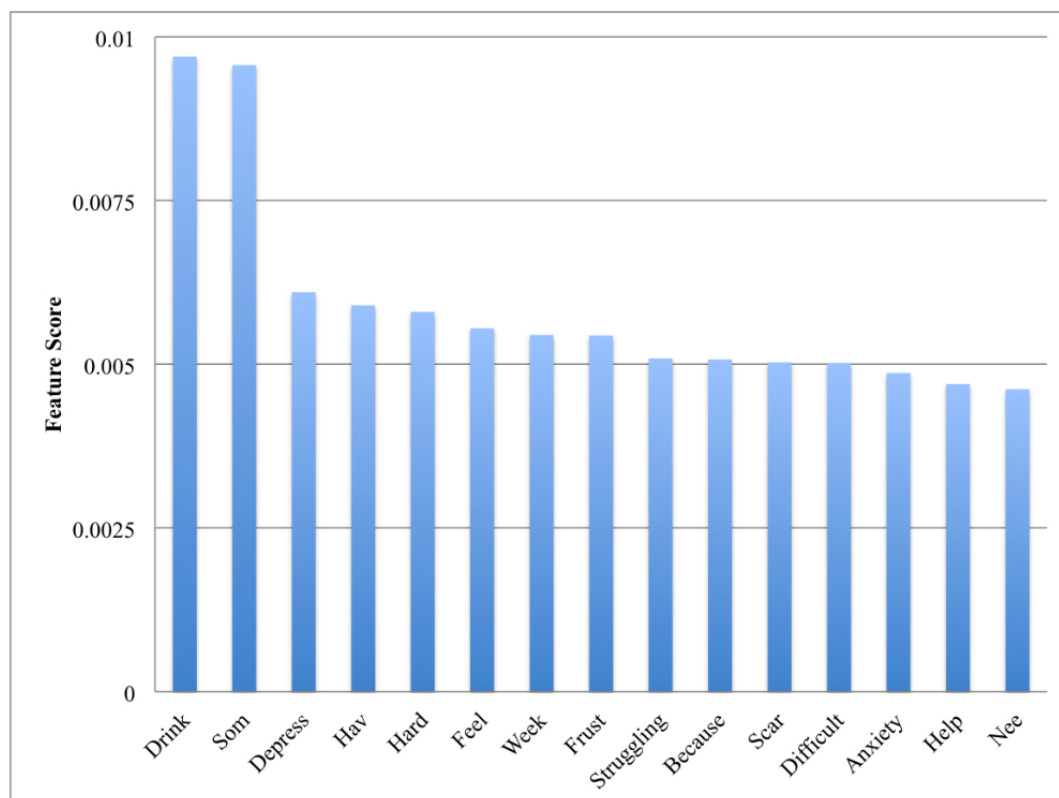
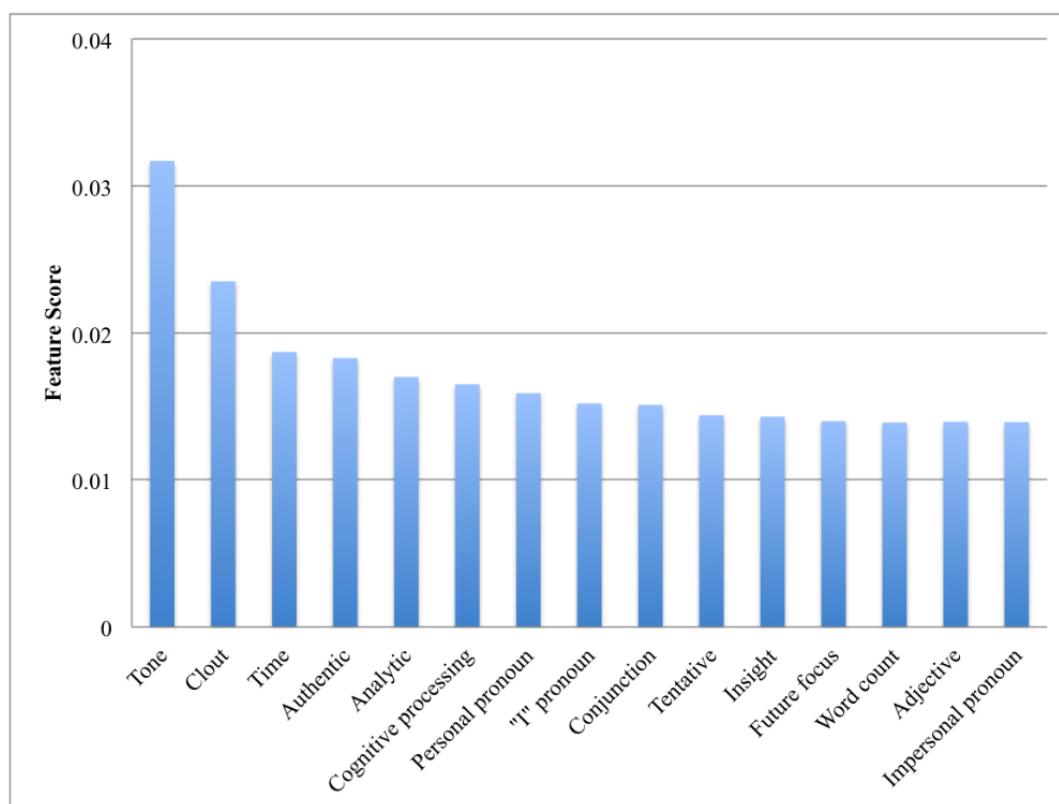
Thus, our codebook captured recovery problems broadly construed. Results of hand-labeling revealed that of the 2581 messages posted to the forum over the course of the study 1, 388 (15%) disclosed some recovery problem. Review of these messages revealed themes including negative affect, cravings, and discouragement. Some described sleep problems, legal issues, medical concerns, unemployment, interpersonal conflict, financial worries, or housing. In a few cases, the writer simply shared that he or she was “struggling” or having a “hard time.” Some messages relayed relapse. In contrast, messages not relaying recovery problems included small talk, affirmations, bonding, reports of doing well or feeling good, or giving support to others.

Supervised Machine Learning

To choose an optimal classifier and its parameters, we performed 10-fold cross validation on labeled data from study 1, partitioned into 70% training and 30% test datasets. Error metrics used were the average F-scores and AUC scores. Moreover, a total of 3 basic classifiers were considered (1) SVMs with linear and Gaussian kernels, (2) decision trees, and (3) boosted decision trees. Our results indicated that SVM performed worst with improvements in decision trees and best performance in boosted decision trees where we achieved F-scores of 0.88, 0.89, and 0.94 for the BoW, LIWC, and hybrid approaches, respectively. For the decision tree classifiers, we used tree depth of 3 and a minimum of 10 samples per leaf at termination when using the BoW feature space. When using the LIWC feature space, we used the same tree depth but a minimum of 8 samples per leaf at termination. For the hybrid feature space, we used a slightly deeper tree (depth=4) with a minimum of 11 samples per leaf at termination. Boosting utilized an average of 175 estimators across the 3 feature spaces.

Having set parameters, we trained on all data from study 1 and applied all 3 classifiers to test data in study 2. Recall that study 2 contained messages posted by a separate cohort of individuals with substance use disorders (in contrast to study 1 in which all individuals had alcohol abuse issues). F-scores for SVMs, decision trees, and boosted decision trees in test data are provided in Table 1.

Figures 1 and 2 show the top features extracted from the BoW and LIWC representations, respectively. For BoW, top features included words with the stems: drink, som (eg, some), because, hard, depress, feel, and hav (eg, have). For LIWC, top features are tone, clout, time, authenticity, analytic words, and insight words. Moreover, 3 top categories include pronoun forms.

Figure 1. Fifteen most important feature words in the Bag-of-Words (BoW) framework.**Figure 2.** Fifteen most important features in the Linguistic Inquiry and Word Count (LIWC) framework.

To understand whether demographic characteristics (gender, age) would influence how recovery problems were expressed in language, we conducted additional experiments in the 2581 messages from study 1. In these experiments, we left 1 gender or age out of the training set, reserving this gender or age for a

testing set. We used decision tree classifiers with feature representation from LIWC to test this question, finding an F-score=0.76 when training on the 1618 messages posted by women and testing on the messages posted by men, which is identical to cross-validation results achieved with the

gender-mixed study 1 sample (F-score=0.76). We used the same approach for age, first leaving out 486 messages from those under 30 years, then 758 messages by those in their 30s, then 881 messages by those posted in their 40s, and finally 309 messages posted by those 50 years or older. The F-scores achieved were 0.77, 0.78, 0.77, and 0.73, respectively. Thus, they were roughly consistent with full study 1 cross-validation, although slightly lower for the 50 years or older group despite the training data being largest.

We produced decision trees for each approach to represent the relationship between language features in predicting recovery problems. For our models that involved boosting, multiple trees impact each classification decision, so any individual tree will provide only a small window into the logic of classification. **Figures 3** and **4** depict truncated exemplar decision trees for the BoW and LIWC approaches. Text in speech bubbles represents messages that would be correctly classified as recovery problem (red) or not a recovery problem (green) by following the associated path. In **Figure 3**, we can see that the BoW decision tree begins with the stem “lot,” with messages having an absence of the word “lot” (0.0 rate of “lot”) following the “true” branch, and messages with presence of the stem “lot” following the “false” branch and being labeled as “recovery problem” (eg, “I’ve been drinking a lot lately”). For those messages not mentioning “lot,” we next look for the stem “thank,” the presence of which leads to a “no recovery problem” label. For those without “lot” or “thank,” we look for “where,” the

presence of which would lead to a “recovery problem” label (eg, “Fighting with my bf again and I don’t know where to go”).

Figure 4 shows the exemplar LIWC tree, which begins with the category of feeling words, producing a categorization of “recovery problem” when paired with time words (eg, “I’ve been feeling not myself for the past week”) but a “no recovery problem” label when mentions of time are below a minimum threshold (eg, “I’m feeling ok.”). For messages without feeling words, the “recovery problem” label would be applied where anger words appear with quantity words (eg, “I’m so pissed!”).

As boosted decision trees performed better than other classifiers, error analysis was summarized in detail for this classifier, with **Table 2** providing specificity, sensitivity, and AUC achieved in the test data for each language processing approach. Results reveal that performance was somewhat improved for hybrid over LIWC and for LIWC over BoW (**Table 2**). More specifically, the hybrid outperforms the LIWC approach in terms of the F-score and the specificity, but not sensitivity, a point we return to below. The hybrid approach makes for an especially robust classifier as seen from the receiver operating characteristics (ROC) curves in **Figure 5**. BoW had the lowest sensitivity. For example, the following message was correctly identified by the hybrid and LIWC approaches and missed by BoW: “This is the hardest thing I have ever done. I just wish I felt better bout recovery. I’m nervous I’m gonna go back to my old ways.”

Figure 3. Example decision tree using features from the Bag-of-Words (BoW) approach. Feature importance was calculated using the Gini Impurity measure.

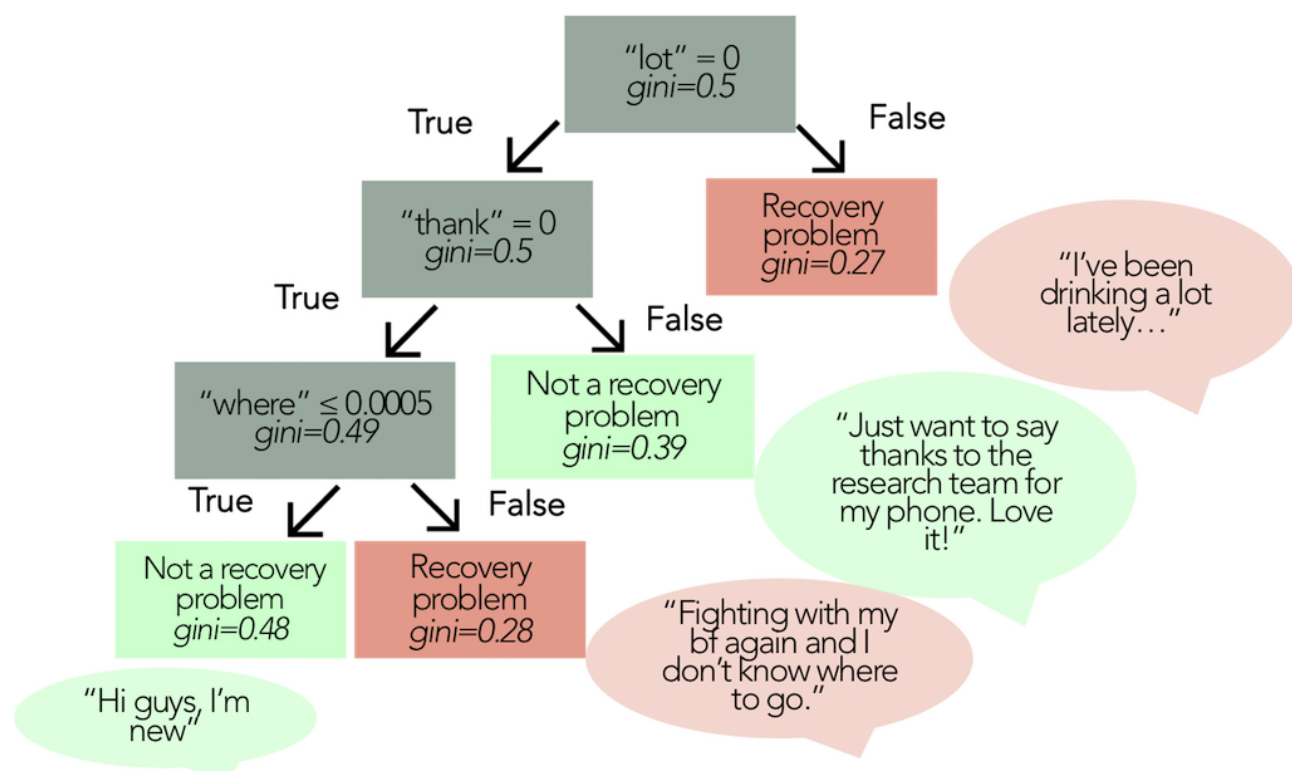


Figure 4. Example decision tree using features from the Linguistic Inquiry and Word Count (LIWC) approach. Feature importance was calculated using the Gini Impurity measure.

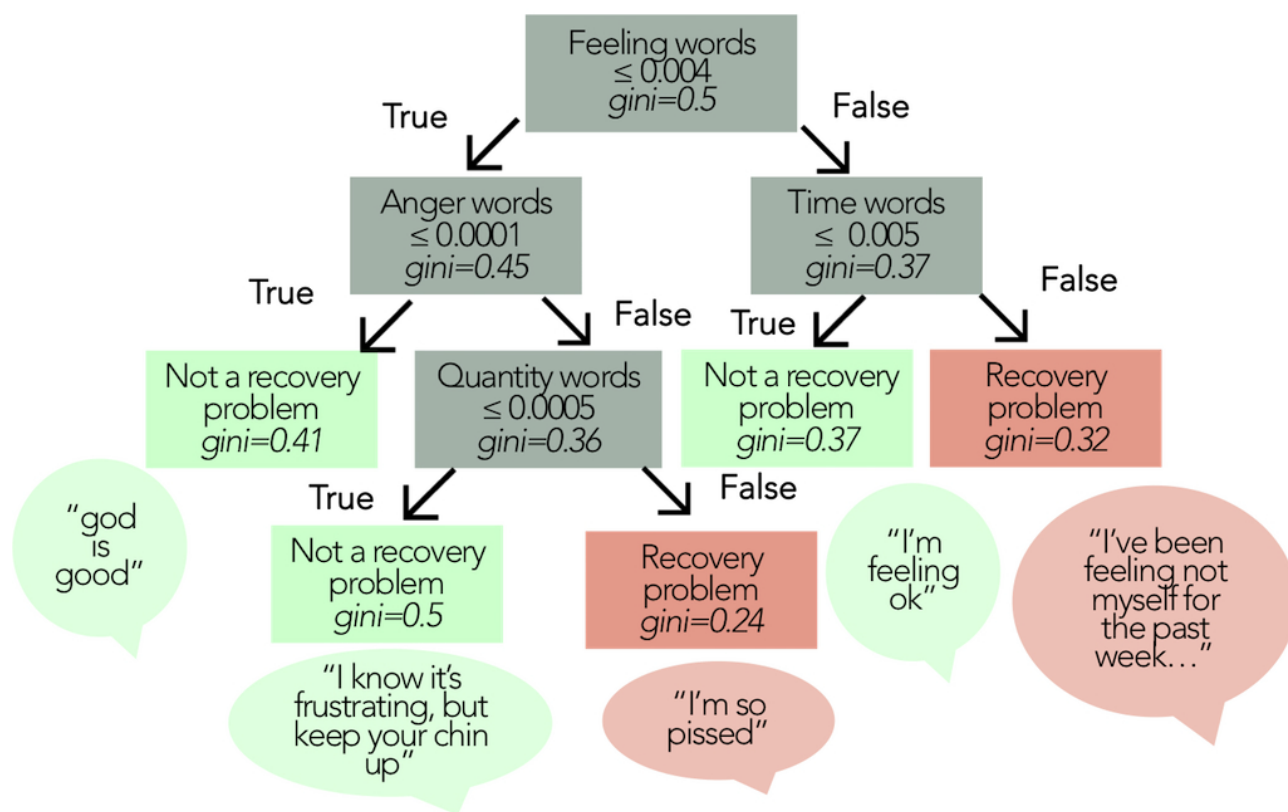


Table 2. Error analysis in study 2 for boosted decision trees using 3 language processing approaches.

Language processing approach	Sensitivity	Specificity	AUC ^a
BoW ^b	0.87	0.78	0.85
LIWC ^c	0.91	0.78	0.88
Hybrid	0.88	0.82	0.92

^aAUC: area under curve.

^bBoW: Bag-of-Words.

^cLIWC: Linguistic Inquiry and Word Count.

Figure 5. Receiver operating characteristic (ROC) curves for boosted decision tree classifiers on the Bag-of-Words (BoW; left), Linguistic Inquiry and Word Count (LIWC; middle), and hybrid (right) feature spaces.

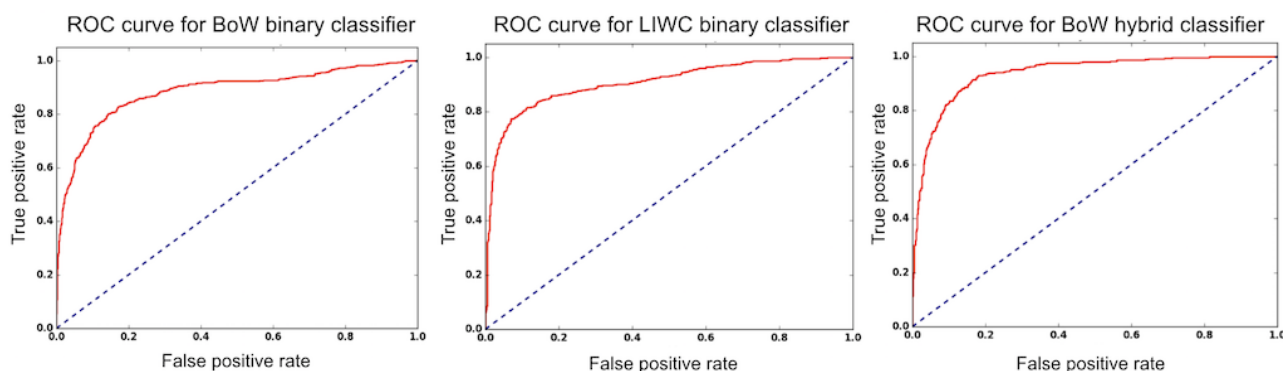
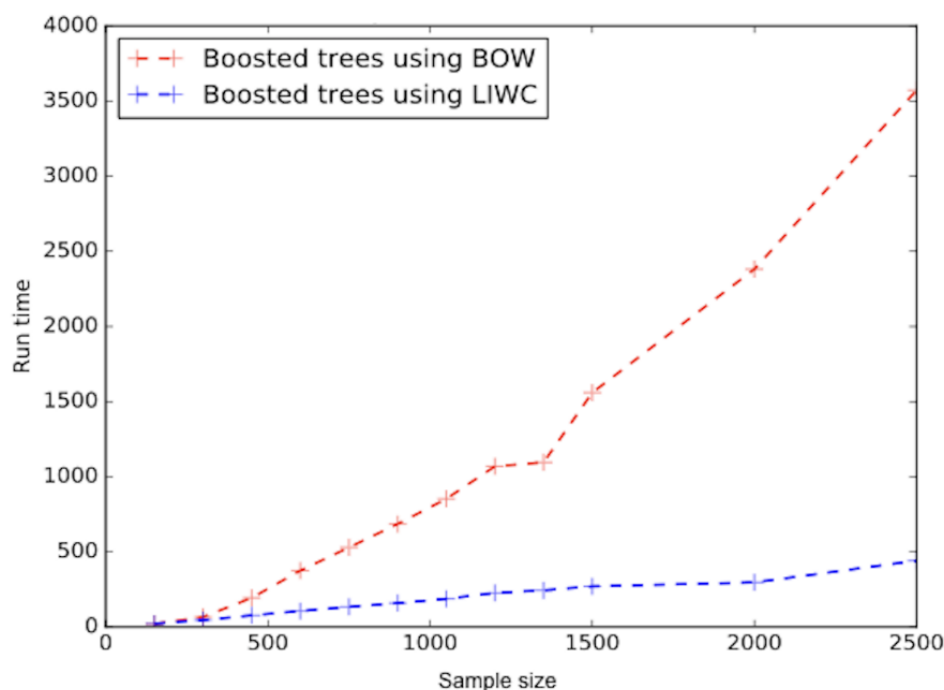


Figure 6. Training times of boosted decision tree classifiers on Bag-of-Words (BoW) and Linguistic Inquiry and Word Count (LIWC) feature spaces.

To understand this classification error, it is helpful to note that the message makes heavy use of personal pronouns such as “I,” which appear in the top 10% of the LIWC features of importance in our datasets but contains just 1 highly ranked unigram from BoW (“hard”). By most metrics, the LIWC and hybrid approach outperform a BoW approach using a boosted decision tree model.

Finally, from [Figure 6](#) we can see the training time for boosted decision trees on LIWC and BoW feature spaces, showing a huge speed advantage for LIWC, a result that was consistent across all classifiers.

Discussion

Principal Findings

The burden on mental health services has fueled recent growth of digital interventions, many of which involve text-based forums connecting a network of peers. Forums often operate with assistance from a moderator who steps in as needed, such as when participants’ problems demand more formal intervention or could overwhelm peers’ abilities to help. Although moderators play an important role, their time-consuming work limits the scalability of digital interventions. This study demonstrates a solution that facilitates moderation efficiency while reducing the possibility of overlooking messages of concern: a machine learning-based model to automatically flag messages that disclose recovery problems. For this work, we used several machine learning approaches, with boosted decision trees performing best, while also offering a view into the logic of classification that may be helpful in establishing face validity.

We also represented our data through a number of computational linguistics techniques. Although the BoW approach captured domain-specific language, it performed somewhat worse than

LIWC, a dictionary-based approach capturing psycholinguistic features. LIWC may do well in this context because recovery problems have important affective dimensions; prior literature shows that LIWC may perform well in cases where affect is a dominant theme [42]. We further found that a hybrid approach, leveraging a combination of features from the dictionary-based LIWC program and BoW, performed best for classifying our test data with regard to AUC and F-score. However, these improvements were only marginally improved over LIWC alone. LIWC achieved a similar F-score of 0.83 (compared with 0.85 for the hybrid) and actually had a higher sensitivity of 0.91.

Ideally, analysts often seek solutions that maximize performance as measured by the F-score, which in this case points toward the hybrid approach. However, there are times when an analyst might prefer greater sensitivity (avoiding false negatives) over improved specificity (avoiding false positives), including perhaps the context of addiction treatment and other health contexts where missing problematic messages could be costly. Given our desire for high sensitivity, LIWC may even be a preferable option over the hybrid. To put this in practical terms, LIWC correctly classified 116 out of 127 true positives in our study 2 test data, compared with 112 classified by the hybrid approach. These additional 4 messages came at the expense of an additional 29 false negatives. However, given the potential consequences of a missed true positive, the additional review time may be seen as worthwhile, especially in early stages of implementing this sort of classifier when concerns about missing actionable messages may limit adoption. Digital mental health interventions seek to ease burden on providers while delivering care to patients, but adoption requires faith in system performance among those on the “front lines.”

LIWC may also be preferable given its easier real-time implementation. Our experiments showed that LIWC features enabled faster training than BoW. LIWC may also have an

implementation advantage as the BoW approach involves calculating TF-IDF scores that reflect the occurrence rate of the word in a single document as well as that word's occurrence across all documents in a sample. This suggests that BoW may present computational challenges when applied in a "live" forum, as the overall occurrence for unigrams may change as new messages are posted and as new unigrams may emerge over time. On the other hand, as the LIWC dictionaries are broad and fixed, classifiers may work well in a system where messages are continually added.

We found that our classifiers were flexible enough to capture numerous circumstances that present problems in recovery, including interpersonal conflicts, job and housing instability, feelings of hopelessness, and encountering triggers. Despite the variety of problems described, classifiers relied heavily on particular ways of talking about drinking, affect, and context, as evident from the important features extracted for each method. Decision rules using the BoW approach were sometimes based on weights of words explicitly linked to drinking ("drink," "relapse," "sobriety," and so on), but decision trees also revealed the use of certain location and context-related terms ("stay" and "where") in decision rules. Decision rules from the LIWC approach were rarely based on explicit topics of discussion, but instead reflected characteristics such as tone, affect, insight, and presence of quantifiers and time references, as well as pronouns. In a tree-based approach, it is not simply using words within these categories that matters but co-occurrence with words in other categories.

Comparison With Prior Work

Notably, in all cases, we achieved good performance relative to Huh and colleagues [12], who also attempted to detect appropriate messages for moderator intervention, and who achieved F-scores up to 0.54. This may in part reflect the difference in machine learning approach, as we used boosted decision trees rather than the Naive Bayes technique they report. The improvement may also reflect the labeling process for training data. Specifically, their training approach labeled messages according to whether they actually received a moderator response, presuming these to be messages of greatest concern, but we implemented a reliable human coding process that we thought would minimize error, as moderators' responses are actually driven by a number of factors beyond the level of concern a message produces.

Indeed, our results are more closely in line with studies that have used hand-labeled data for training. F-scores for our hybrid model are comparable with the best results achieved in a shared task challenge to flag messages for elevated suicide risk in a forum for Australian youth [50] and slightly lower than a follow-up study from the same forum that utilized an ensemble of feature extraction approaches (LIWC, topic modeling, meta-data, etc) [51]. However, it is important to note our more conservative approach of testing our model in a separate iteration of the forum with a separate patient population. Like Conan et al [51], we also obtained better results for boosted decision trees relative to SVM.

Implications for System Design

Moderators can play a pivotal role in digital forums for at-risk populations but face difficulties keeping up with new content. Recently, scholars have called for improving digital health interventions by emphasizing *efficiency* of human support: the level of increased engagement and intervention effectiveness relative to the effort expended by staff [13]. Our findings demonstrate an opportunity to improve efficiency through automatically identifying, in real time, when participants disclose pressing concerns. Resulting classifications could be easily used to populate an interface to display high-priority content to moderators (see [Multimedia Appendix 1](#) for an example of how our classifier has been applied in our live mobile-based recovery support platform). The interface may also provide moderators with an opportunity to dispute message classifications they view as erroneous, generating data to refine classifiers in the future (See [Multimedia Appendix 2](#)). Upon review of flagged messages, moderators might choose to intervene in a number of ways, such as through providing emotional support, directing participants to intervention elements that might suit their needs, or connecting participants with mentors or services.

Although our present solution requires human review and response, it is worth noting an alternate approach of fully automating responses. For instance, flagged messages could prompt the system to provide immediate contact information for treatment providers or emergency services, thus offering support even late at night and early in the morning. Some systems have also used machine learning methods to match newly posted content to semantically similar earlier content, displaying these older messages alongside the responses they generated in case they are useful to the current poster [70]. In addition, more complex dialogue systems have been applied to further reduce the human labor behind digital health interventions, including interactive "conversational agents," software programs that mimic human conversation, and that may further display human-like cues through voice or visual representations [71,72]. Such techniques are promising but involve trade-offs relative to trained staff who develop personal relationships with participants and can exercise expert judgment [73]. For instance, unlike software programs, moderators can choose to ignore messages they believe are "false positives," not warranting their expression of concern. Of course, moderators also vary in personal and professional qualities that make them effective. For instance, staff may be particularly successful through conveying a combination of trustworthiness, benevolence, and expertise [74].

In the future, efficient just-in-time support may involve judicious use of both human support and automated messages. Short of full automation, efficiency could be enhanced through providing moderators with a drop-down list of common responses that may be appropriate after a problem is disclosed, with an editor allowing optional personalization. Information about a given participant (eg, risk score from the last completed survey) could also indicate whether a flagged message should be sent to the moderator for a personalized response or managed through automation.

Future Research Directions

Findings from this study suggest promising areas for future research. First, a number of additional optimizations of our classifiers may be possible. For instance, additional dictionaries have also been developed in the realm of electronic medical records and these could prove promising in capturing recovery-related concepts [75]. Conditional Random Fields methods also work well in classifying natural language [76]. In future, we may also improve our BoW-based model through attention to dimensionality reduction, latent semantic analysis, and potentially extracting bigrams (or trigrams, etc) in addition to unigrams. As far as our hybrid approach is concerned, we might further optimize performance by giving further consideration to the number of features pulled from each component method. Specifically, to determine the number of important features from LIWC and BoW to include in the hybrid model, we tested cut-off points at 5% intervals (10%, 15%, 20%, etc) and found the best results for 10%, but more fine-grained adjustments could be tested, including plotting F-score relative to the number of features, and perhaps allowing for different cut-offs for LIWC and BoW.

Although our models were robust regardless of type of substance of abuse (which varied across Studies 1 and 2) and by gender, our leave-one-out experiments suggest that further research may also be needed to understand if older adults use similar language to convey recovery problems. We also did not test our model across differences such as race or education, leaving it unclear whether our models would work well in populations of different compositions.

Models might also take additional data into account. This analysis was conducted at the message level, but it may be possible to improve our models by considering each individual's pattern of messaging. Those who habitually post recovery problems may require a different level and style of response than those who escalate posting of worrisome messages. Other system use or sensor data may also inform our model, such that patterns of reading messages, interacting with intervention features (eg, pressing a "panic button"), or moving to new geographic locations may be integrated into decision rules around moderator involvement [77]. Similar work in the domain of suicide risk has incorporated additional features reflecting metadata from the discussion forum (eg, How many usernames are referenced in a message? Where does a message fall in sequence within a thread?) [51].

Ultimately, the efficiency of our approach to flagging concerning messages should be addressed empirically, such as through a trial randomizing some participants to a system where moderators manually review the forum and others to a system where moderators rely on text-based classification. Outcomes may include moderators' workload as well as patients' satisfaction and health outcomes. Further research is also needed to establish how to best intervene after a recovery problem message, including through personalized responses from moderators or automated messages.

A final future direction relates to privacy. Our surveillance approach offers opportunities to intervene early to help those in need, but introduces an important trade-off as far as privacy.

Specifically, we use passively collected data to infer underlying risk levels that patients may not even be aware of, with these data being highly sensitive [50]. Future research is needed to clarify how patients understand uses of their data for surveillance, how they balance surveillance and privacy concerns, and the contexts under which they find surveillance acceptable. In this study, it is possible that we allayed some privacy concerns by recruiting patients through trusted treatment providers and clinicians and obtaining informed consent, but patients may have greater privacy concerns in the domain of commercial mental health platforms.

Limitations

This study has limitations. First, our approach would not allow us to assist participants who do not post on a discussion forum. Furthermore, we do not look at private messages, where participants potentially disclose even more sensitive information [50]. In addition, as we did not label subtypes of recovery problems, it is possible that our classifier may be biased toward recognizing certain types of common problematic messages over others. Future work should consider coding subtypes of recovery problems. For instance, relatively rare problems that are nonetheless highly concerning may include mentions of suicide risk or solicitations to buy or sell drugs. Finally, one of the core strengths of our dataset is also tied to one of our study limitations. Specifically, we have access to a dataset of anonymous messages exchanged in a system restricted to those who share a SUDs diagnosis (a condition of study eligibility). These factors mean that discussion may be particularly candid and may offer unusual insight into mental health risk. At the same time, these considerations imply that existing labeled datasets cannot easily be adapted to train classifiers within our dataset. Our model leverages a relatively small set of training messages, which has implications for the machine learning approaches available and the results obtained.

Conclusions

Digital interventions hold promise to offer cost-effective, constantly available support to those in recovery, and to reduce human workload relative to face-to-face SUDs interventions. However, human support still plays a vital role in many effective digital interventions. For interventions involving discussion forums, trained moderators can respond in real time to help participants who are facing challenges. Yet, these moderators must dedicate substantial time and effort to manually review newly posted messages to identify serious problems, and the process can be error-prone. Our results show that message content can be effectively leveraged toward facilitating just-in-time supportive intervention. Language-based classification models have potential for massive scalability as digital interventions for addiction support continue to expand.

Individuals' language use, both through its content and composition, offers a means of understanding psychological states and traits. Our work expands on the existing literature by combining and layering computational linguistics and machine learning techniques in the context of streamlining human support within digital substance abuse recovery interventions. Yet, this work also has theoretical and methodological value beyond this

specific context, suggesting useful directions for applying language classification to digital mental health more broadly.

Acknowledgments

This research was funded by the National Institute of Alcohol Abuse and Alcoholism (R01 AA017192) and the National Institute on Drug Abuse (R01DA034279, R01DA040449, and DP2DA042424). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funder had no role in any aspect of the development, conduct, analysis, or reporting of the study. The authors wish to thank Adam Maus, Ryan Westergaard, Alexandria Cull, Yan Liu, and Andrew Quanbeck for their important contributions to the project.

Conflicts of Interest

DHG has a shareholder interest in CHESS Mobile Health, a public benefit corporation that develops health care technology for patients struggling with addiction. DHG and the University of Wisconsin–Madison’s Conflict of Interest Committee manage this relationship. All other authors have no conflicts of interests to declare.

Multimedia Appendix 1

Interface for moderator to review messages classified as indicating recovery problems.

[PDF File (Adobe PDF File), 145KB - [jmir_v20i6e10136_app1.pdf](#)]

Multimedia Appendix 2

Interface for moderator to provide feedback on message classification.

[PDF File (Adobe PDF File), 138KB - [jmir_v20i6e10136_app2.pdf](#)]

References

1. Barak A, Grohol JM. Current and future trends in internet-supported mental health interventions. *J Technol Hum Serv* 2011 Jul;29(3):155-196. [doi: [10.1080/15228835.2011.616939](#)]
2. 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](#)]
3. Moore BA, Fazzino T, Garnet B, Cutter CJ, Barry DT. Computer-based interventions for drug use disorders: a systematic review. *J Subst Abuse Treat* 2011 Apr;40(3):215-223 [FREE Full text] [doi: [10.1016/j.jsat.2010.11.002](#)] [Medline: [21185683](#)]
4. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. *JMIR Ment Health* 2014;1(1):e5 [FREE Full text] [doi: [10.2196/mental.4004](#)] [Medline: [26543905](#)]
5. Substance Abuse and Mental Health Services Administration. SAMHSA. 2015. Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health URL: <https://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf> [accessed 2018-04-24] [WebCite Cache ID 6yvUm7BiF]
6. Quanbeck A, Chih M, Isham A, Gustafson D. Mobile delivery of treatment for alcohol use disorders: a review of the literature. *Alcohol Res* 2014;36(1):111-122 [FREE Full text] [Medline: [26259005](#)]
7. Litvin EB, Abrantes AM, Brown RA. Computer and mobile technology-based interventions for substance use disorders: an organizing framework. *Addict Behav* 2013 Mar;38(3):1747-1756. [doi: [10.1016/j.addbeh.2012.09.003](#)] [Medline: [23254225](#)]
8. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. *J Am Med Assoc Psychiatry* 2014 May;71(5):566-572 [FREE Full text] [doi: [10.1001/jamapsychiatry.2013.4642](#)] [Medline: [24671165](#)]
9. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions—a systematic review. *Internet Interv* 2014 Oct;1(4):205-215. [doi: [10.1016/j.invent.2014.08.003](#)]
10. Greidanus E, Everall RD. Helper therapy in an online suicide prevention community. *Br J Guid Coun* 2010 May;38(2):191-204. [doi: [10.1080/03069881003600991](#)]
11. Gilat I, Tobin Y, Shahar G. Responses to suicidal messages in an online support group: comparison between trained volunteers and lay individuals. *Soc Psychiatry Psychiatr Epidemiol* 2012 Dec;47(12):1929-1935. [doi: [10.1007/s00127-012-0508-7](#)] [Medline: [22491905](#)]
12. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. *J Biomed Inform* 2013 Dec;46(6):998-1005 [FREE Full text] [doi: [10.1016/j.jbi.2013.08.011](#)] [Medline: [24025513](#)]

13. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. *Clin Psychol Sci Pract* 2016 Nov 17;24(1):27-45. [doi: [10.1111/cpsp.12173](https://doi.org/10.1111/cpsp.12173)]
14. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2009 Dec 08;29(1):24-54. [doi: [10.1177/0261927X09351676](https://doi.org/10.1177/0261927X09351676)]
15. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. *Curr Opin Psychol* 2016 Jun;9:77-82 [FREE Full text] [doi: [10.1016/j.copsyc.2016.01.004](https://doi.org/10.1016/j.copsyc.2016.01.004)] [Medline: [27042689](https://pubmed.ncbi.nlm.nih.gov/27042689/)]
16. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. Austin, TX: The University of Texas at Austin; 2015. The development and psychometric properties of liwc2015 URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvUyNkq1]
17. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet* 2009 Jun 27;373(9682):2223-2233. [doi: [10.1016/S0140-6736\(09\)60746-7](https://doi.org/10.1016/S0140-6736(09)60746-7)] [Medline: [19560604](https://pubmed.ncbi.nlm.nih.gov/19560604/)]
18. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* 2013 Nov 9;382(9904):1575-1586. [doi: [10.1016/S0140-6736\(13\)61611-6](https://doi.org/10.1016/S0140-6736(13)61611-6)] [Medline: [23993280](https://pubmed.ncbi.nlm.nih.gov/23993280/)]
19. Dennis M, Scott CK. Managing addiction as a chronic condition. *Addict Sci Clin Pract* 2007 Dec;4(1):45-55 [FREE Full text] [Medline: [18292710](https://pubmed.ncbi.nlm.nih.gov/18292710/)]
20. Kelly JF, White WL. Recovery management and the future of addiction treatment and recovery in the USA. In: Kelly JF, White WL, editors. *Addiction Recovery Management: Theory, Research and Practice (Current Clinical Psychiatry)*. Totowa, NJ: Humana Press; 2010:303-316.
21. Pew Research Center. Demographics of Mobile Device Ownership and Adoption in the United States URL: <http://www.pewinternet.org/fact-sheet/mobile/> [accessed 2018-02-14] [WebCite Cache ID 6xEcZNiMF]
22. 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. *Br Med J* 2004 May 15;328(7449):1166 [FREE Full text] [doi: [10.1136/bmj.328.7449.1166](https://doi.org/10.1136/bmj.328.7449.1166)] [Medline: [15142921](https://pubmed.ncbi.nlm.nih.gov/15142921/)]
23. 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](https://doi.org/10.1080/15228830802094429)]
24. Heisler M, Vijan S, Makki F, Piette JD. Diabetes control with reciprocal peer support versus nurse care management: a randomized trial. *Ann Intern Med* 2010 Oct 19;153(8):507-515 [FREE Full text] [doi: [10.7326/0003-4819-153-8-201010190-00007](https://doi.org/10.7326/0003-4819-153-8-201010190-00007)] [Medline: [20956707](https://pubmed.ncbi.nlm.nih.gov/20956707/)]
25. DeAndrea DC, Anthony JC. Online peer support for mental health problems in the United States: 2004-2010. *Psychol Med* 2013 Nov;43(11):2277-2288 [FREE Full text] [doi: [10.1017/S0033291713000172](https://doi.org/10.1017/S0033291713000172)] [Medline: [23410539](https://pubmed.ncbi.nlm.nih.gov/23410539/)]
26. Cunningham JA, van Mierlo T, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. *Patient Educ Couns* 2008 Feb;70(2):193-198. [doi: [10.1016/j.pec.2007.10.003](https://doi.org/10.1016/j.pec.2007.10.003)] [Medline: [18022340](https://pubmed.ncbi.nlm.nih.gov/18022340/)]
27. Klaw E, Dearmin Huebsch P, Humphreys K. Communication patterns in an online mutual help group for problem drinkers. *J Community Psychol* 2000;28(5):535-546. [doi: [10.1002/1520-6629\(200009\)28:5<535::AID-JCOP7>3.0.CO;2-0](https://doi.org/10.1002/1520-6629(200009)28:5<535::AID-JCOP7>3.0.CO;2-0)]
28. Chuang KY, Yang CC. Interaction patterns of nurturant support exchanged in online health social networking. *J Med Internet Res* 2012;14(3):e54 [FREE Full text] [doi: [10.2196/jmir.1824](https://doi.org/10.2196/jmir.1824)] [Medline: [22555303](https://pubmed.ncbi.nlm.nih.gov/22555303/)]
29. Arnstein P, Vidal M, Wells-Federman C, Morgan B, Caudill M. From chronic pain patient to peer: benefits and risks of volunteering. *Pain Manag Nurs* 2002 Sep;3(3):94-103. [Medline: [12198640](https://pubmed.ncbi.nlm.nih.gov/12198640/)]
30. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. *J Med Internet Res* 2011;13(3):e62 [FREE Full text] [doi: [10.2196/jmir.1728](https://doi.org/10.2196/jmir.1728)] [Medline: [21846635](https://pubmed.ncbi.nlm.nih.gov/21846635/)]
31. Huh J, Patel R, Pratt W. Tackling dilemmas in supporting "the whole person" in online patient communities. *Proc SIGCHI Conf Hum Factor Comput Syst* 2012;2012:923-926 [FREE Full text] [doi: [10.1145/2207676.2208535](https://doi.org/10.1145/2207676.2208535)] [Medline: [24634893](https://pubmed.ncbi.nlm.nih.gov/24634893/)]
32. Mares M, Gustafson DH, Glass JE, Quanbeck A, McDowell H, McTavish F, et al. Implementing an mHealth system for substance use disorders in primary care: a mixed methods study of clinicians' initial expectations and first year experiences. *BMC Med Inform Decis Mak* 2016 Sep 29;16(1):126 [FREE Full text] [doi: [10.1186/s12911-016-0365-5](https://doi.org/10.1186/s12911-016-0365-5)] [Medline: [27687632](https://pubmed.ncbi.nlm.nih.gov/27687632/)]
33. De Vries RA, Zaga C, Bayer F, Drossaert CH, Truong KP, Evers V. Experts get me started, peers keep me going: comparing crowd- versus expert-designed motivational text messages for exercise behavior change. : EAI; 2017 Presented at: International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain p. 155-162. [doi: [10.1145/3154862.3154875](https://doi.org/10.1145/3154862.3154875)]
34. 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](https://doi.org/10.1097/NCN.0b013e3182343efa)] [Medline: [22240564](https://pubmed.ncbi.nlm.nih.gov/22240564/)]
35. Lieberman MA, Golant M, Winzelberg A, McTavish F, Gustafson DH. Comparisons: professionally-directed and self-directed internet groups for women with breast cancer. *Int J Self Help* 2004;2(3):219-235. [doi: [10.2190/GE85-J31W-XJV7-LB9L](https://doi.org/10.2190/GE85-J31W-XJV7-LB9L)]

36. Johnson K, Richards S, Chih MY, Moon TJ, Curtis H, Gustafson DH. A pilot test of a mobile app for drug court participants. *Subst Abuse* 2016;10:1-7 [FREE Full text] [doi: [10.4137/SART.S33390](https://doi.org/10.4137/SART.S33390)] [Medline: [26917964](https://pubmed.ncbi.nlm.nih.gov/26917964/)]
37. 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](https://doi.org/10.1371/journal.pone.0010939)] [Medline: [20544030](https://pubmed.ncbi.nlm.nih.gov/20544030/)]
38. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. *Transl Behav Med* 2015 Sep;5(3):335-346 [FREE Full text] [doi: [10.1007/s13142-015-0324-1](https://doi.org/10.1007/s13142-015-0324-1)] [Medline: [26327939](https://pubmed.ncbi.nlm.nih.gov/26327939/)]
39. Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend* 2007 May;88 Suppl 2:S24-S30 [FREE Full text] [doi: [10.1016/j.drugalcdep.2006.09.008](https://doi.org/10.1016/j.drugalcdep.2006.09.008)] [Medline: [17056207](https://pubmed.ncbi.nlm.nih.gov/17056207/)]
40. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. *Proc SIGCHI Conf Hum Factor Comput Syst* 2016 May;2016:2098-2110 [FREE Full text] [doi: [10.1145/2858036.2858207](https://doi.org/10.1145/2858036.2858207)] [Medline: [29082385](https://pubmed.ncbi.nlm.nih.gov/29082385/)]
41. Kornfield R, Toma CL, Shah DV, Moon TJ, Gustafson DH. What do you say before you relapse? How language use in a peer-to-peer online discussion forum predicts risky drinking among those in recovery. *Health Commun* 2017 Aug 09:1-10. [doi: [10.1080/10410236.2017.1350906](https://doi.org/10.1080/10410236.2017.1350906)] [Medline: [28792228](https://pubmed.ncbi.nlm.nih.gov/28792228/)]
42. Brancu M, Jobes D, Wagner BM, Greene JA, Fratto TA. Are there linguistic markers of suicidal writing that can predict the course of treatment? A repeated measures longitudinal analysis. *Arch Suicide Res* 2016 Jul 02;20(3):438-450. [doi: [10.1080/13811118.2015.1040935](https://doi.org/10.1080/13811118.2015.1040935)] [Medline: [26219609](https://pubmed.ncbi.nlm.nih.gov/26219609/)]
43. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France p. 3267-3276. [doi: [10.1145/2470654.2466447](https://doi.org/10.1145/2470654.2466447)]
44. Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno-Tempini ML, et al. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. 2014 Presented at: *ACL Workshop on Computational Linguistics and Clinical Psychology*; June 27, 2014; Baltimore, MD p. 27-36. [doi: [10.3115/v1/W14-3204](https://doi.org/10.3115/v1/W14-3204)]
45. Bak JY, Lin CY, Oh A. Self-disclosure topic model for classifying and analyzing Twitter conversations. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October 25-29, 2014; Doha, Qatar p. 1986-1996. [doi: [10.3115/v1/D14-1213](https://doi.org/10.3115/v1/D14-1213)]
46. Balani S, De Choudhury M. Detecting and characterizing mental health related self-disclosure in social media. : ACM; 2015 Presented at: *ACM Conference on Human Factors in Computing Systems*; April 18-23, 2015; Seoul, Korea p. 1373-1378. [doi: [10.1145/2702613.2732733](https://doi.org/10.1145/2702613.2732733)]
47. Wang YC, Burke M, Kraut R. Modeling self-disclosure in social networking sites. : ACM; 2016 Presented at: *19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*; February 27-March 02, 2016; San Francisco, CA, USA p. 74-85. [doi: [10.1145/2818048.2820010](https://doi.org/10.1145/2818048.2820010)]
48. Houghton DJ, Joinson AN. Linguistic markers of secrets and sensitive self-disclosure in Twitter. 2012 Presented at: *45th Hawaii International Conference on System Science (HICSS)*; Jan 4-7, 2012; Maui, Hawaii USA p. 3480-3489. [doi: [10.1109/HICSS.2012.415](https://doi.org/10.1109/HICSS.2012.415)]
49. De Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. In: *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. 2014 Presented at: *AAAI Conference on Weblogs and Social Media*; June 1-4, 2014; Ann Arbor, MI URL: <https://pdfs.semanticscholar.org/2db7/15a479c8961d3020fe906f7bedfa0311b937.pdf>
50. Milne DN, Pink G, Hachey B, Calvo RA. Clpsych 2016 shared task: Triaging content in online peer-support forums. In: *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 2016 Presented at: *Workshop on Computational Linguistics and Clinical Psychology*; June 16, 2016; San Diego, California, USA p. 118-127. [doi: [10.18653/v1/W16-0312](https://doi.org/10.18653/v1/W16-0312)]
51. Cohan A, Young S, Yates A, Goharian N. Triaging content severity in online mental health forums. *J Assoc Inf Sci Technol* 2017 Sep 25;68(11):2675-2689. [doi: [10.1002/asi.23865](https://doi.org/10.1002/asi.23865)]
52. Chuang KY, Yang CC. Informational support exchanges using different computer-mediated communication formats in a social media alcoholism community. *J Assoc Inf Sci Technol* 2013 Oct 23;65(1):37-52. [doi: [10.1002/asi.22960](https://doi.org/10.1002/asi.22960)]
53. Campbell SW, Kelley MJ. Mobile phone use among Alcoholics Anonymous members: new sites for recovery. *New Media Soc* 2008 Dec;10(6):915-933. [doi: [10.1177/1461444808096251](https://doi.org/10.1177/1461444808096251)] [Medline: [22973420](https://pubmed.ncbi.nlm.nih.gov/22973420/)]
54. Witkiewitz K. Predictors of heavy drinking during and following treatment. *Psychol Addict Behav* 2011 Sep;25(3):426-438 [FREE Full text] [doi: [10.1037/a0022889](https://doi.org/10.1037/a0022889)] [Medline: [21480681](https://pubmed.ncbi.nlm.nih.gov/21480681/)]
55. Chung C, Pennebaker JW. The psychological functions of function words. In: Fiedler K, editor. *Frontiers of Social Psychology. Social communication*. New York, NY, US: Psychology Press; 2007:343-359.

56. Zimmermann J, Wolf M, Bock A, Peham D, Benecke C. The way we refer to ourselves reflects how we relate to others: associations between first-person pronoun use and interpersonal problems. *J Res Pers* 2013 Jun;47(3):218-225. [doi: [10.1016/j.jrp.2013.01.008](https://doi.org/10.1016/j.jrp.2013.01.008)] [Medline: [25904163](https://pubmed.ncbi.nlm.nih.gov/25904163/)]
57. Anderson B, Goldin PR, Kurita K, Gross JJ. Self-representation in social anxiety disorder: linguistic analysis of autobiographical narratives. *Behav Res Ther* 2008 Oct;46(10):1119-1125 [FREE Full text] [doi: [10.1016/j.brat.2008.07.001](https://doi.org/10.1016/j.brat.2008.07.001)] [Medline: [18722589](https://pubmed.ncbi.nlm.nih.gov/18722589/)]
58. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: Detecting influenza epidemics using twitter. : Association for Computational Linguistics; 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 27-29, 2011; Edinburgh, UK p. 1568-1576.
59. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for Twitter data against established measures of suicidality. *JMIR Ment Health* 2016 May 16;3(2):e21 [FREE Full text] [doi: [10.2196/mental.4822](https://doi.org/10.2196/mental.4822)] [Medline: [27185366](https://pubmed.ncbi.nlm.nih.gov/27185366/)]
60. Quanbeck A, Gustafson DH, Marsch LA, Chih MY, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. *J Med Internet Res* 2018 Jan 30;20(1):e37 [FREE Full text] [doi: [10.2196/jmir.8928](https://doi.org/10.2196/jmir.8928)] [Medline: [29382624](https://pubmed.ncbi.nlm.nih.gov/29382624/)]
61. Sohn S, Wang Y, Wi C, Krusemark EA, Ryu E, Ali MH, et al. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. *J Am Med Inform Assoc* 2017 Nov 30:-. [doi: [10.1093/jamia/ocx138](https://doi.org/10.1093/jamia/ocx138)] [Medline: [29202185](https://pubmed.ncbi.nlm.nih.gov/29202185/)]
62. Cantú-Paz E, Newsam S, Kamath C. Feature selection in scientific applications. 2004 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 22-25, 2004; Seattle, WA, USA. [doi: [10.1145/1014052.1016915](https://doi.org/10.1145/1014052.1016915)]
63. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognit Lett* 2010 Oct;31(14):2225-2236. [doi: [10.1016/j.patrec.2010.03.014](https://doi.org/10.1016/j.patrec.2010.03.014)]
64. Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32. [doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)]
65. Tong S, Koller D. Support vector machine active learning with applications to text classification. *J Mach Learn Res* 2002;2:45-66. [doi: [10.1162/153244302760185243](https://doi.org/10.1162/153244302760185243)]
66. Apté C, Weiss SM. Data mining with decision trees and decision rules. *Future Gener Comput Syst* 1997 Nov;13(2-3):197-210. [doi: [10.1016/S0167-739X\(97\)00021-6](https://doi.org/10.1016/S0167-739X(97)00021-6)]
67. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825-2830 [FREE Full text]
68. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321-357. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]
69. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 Mar;33(1):159-174. [Medline: [843571](https://pubmed.ncbi.nlm.nih.gov/843571/)]
70. Wang Y, Mehrabi S, Mojarad MR, Li D, Liu H. Retrieval of Semantically Similar Healthcare Questions in Healthcare Forums. : IEEE; 2015 Presented at: International Conference on Healthcare Informatics (ICHI); October 21-23; Dallas, TX, USA p. 517-518. [doi: [10.1109/ICHI.2015.97](https://doi.org/10.1109/ICHI.2015.97)]
71. Bickmore T, Giorgino T, Green N, Picard R. Special issue on dialog systems for health communication. *J Biomed Inform* 2006 Oct;39(5):465-467 [FREE Full text] [doi: [10.1016/j.jbi.2006.02.002](https://doi.org/10.1016/j.jbi.2006.02.002)] [Medline: [16546453](https://pubmed.ncbi.nlm.nih.gov/16546453/)]
72. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *J Am Med Assoc Intern Med* 2016 May 01;176(5):619-625 [FREE Full text] [doi: [10.1001/jamainternmed.2016.0400](https://doi.org/10.1001/jamainternmed.2016.0400)] [Medline: [26974260](https://pubmed.ncbi.nlm.nih.gov/26974260/)]
73. Kasckow J, Zickmund S, Rotondi A, Mrkva A, Gurklis J, Chinman M, et al. Development of telehealth dialogues for monitoring suicidal patients with schizophrenia: consumer feedback. *Community Ment Health J* 2014 Apr;50(3):339-342. [doi: [10.1007/s10597-012-9589-8](https://doi.org/10.1007/s10597-012-9589-8)] [Medline: [23306676](https://pubmed.ncbi.nlm.nih.gov/23306676/)]
74. 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](https://doi.org/10.2196/jmir.1602)] [Medline: [21393123](https://pubmed.ncbi.nlm.nih.gov/21393123/)]
75. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. *J Biomed Inform* 2018 Jan;77:34-49 [FREE Full text] [doi: [10.1016/j.jbi.2017.11.011](https://doi.org/10.1016/j.jbi.2017.11.011)] [Medline: [29162496](https://pubmed.ncbi.nlm.nih.gov/29162496/)]
76. Nakagawa T, Inui K, Kurohashi S. Dependency tree-based sentiment classification using CRFs with hidden variables. : Association for Computational Linguistics; 2010 Presented at: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics; June 2-4, 2010; Los Angeles, California, USA p. 786-794.
77. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas L, Wingood G, et al. A computational future for preventing HIV in minority communities: how advanced technology can improve implementation of effective programs. *J Acquir Immune Defic Syndr* 2013 Jun 1;63 Suppl 1:S72-S84 [FREE Full text] [doi: [10.1097/QAI.0b013e31829372bd](https://doi.org/10.1097/QAI.0b013e31829372bd)] [Medline: [23673892](https://pubmed.ncbi.nlm.nih.gov/23673892/)]

Abbreviations

AUC: area under curve

BoW: Bag-of-Words

LIWC: Linguistic Inquiry and Word Count

ROC: receiver operating characteristics

SUD: substance use disorder

SVM: support vector machines

TF-IDF: Term Frequency-Inverse Document Frequency

Edited by G Wadley, R Calvo, M Czerwinski, J Torous; submitted 15.02.18; peer-reviewed by Y Wang, J Colditz, K Loveys; comments to author 22.03.18; revised version received 04.04.18; accepted 05.04.18; published 12.06.18.

Please cite as:

Kornfield R, Sarma PK, Shah DV, McTavish F, Landucci G, Pe-Romashko K, Gustafson DH

Detecting Recovery Problems Just in Time: Application of Automated Linguistic Analysis and Supervised Machine Learning to an Online Substance Abuse Forum

J Med Internet Res 2018;20(6):e10136

URL: <http://www.jmir.org/2018/6/e10136/>

doi: [10.2196/10136](https://doi.org/10.2196/10136)

PMID: [29895517](https://pubmed.ncbi.nlm.nih.gov/29895517/)

©Rachel Kornfield, Prathusha K Sarma, Dhavan V Shah, Fiona McTavish, Gina Landucci, Klaren Pe-Romashko, David H Gustafson. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 12.06.2018. 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.

Original Paper

Comparing Approaches to Mobile Depression Assessment for Measurement-Based Care: Prospective Study

Amy M Bauer¹, MS, MD; Scott A Baldwin², PhD; Joaquin A Anguera³, PhD; Patricia A Areán¹, PhD; David C Atkins¹, PhD

¹Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

²Department of Psychology, Brigham Young University, Provo, UT, United States

³Department of Neurology, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:

Amy M Bauer, MS, MD

Department of Psychiatry and Behavioral Sciences

University of Washington

1959 NE Pacific Street

Box 356560

Seattle, WA, 98195 6560

United States

Phone: 1 2062218385

Email: abauer1@uw.edu

Abstract

Background: To inform measurement-based care, practice guidelines suggest routine symptom monitoring, often on a weekly or monthly basis. Increasingly, patient-provider contacts occur remotely (eg, by telephone and Web-based portals), and mobile health tools can now monitor depressed mood daily or more frequently. However, the reliability and utility of daily ratings are unclear.

Objective: This study aimed to examine the association between a daily depressive symptom measure and the Patient Health Questionnaire-9 (PHQ-9), the most widely adopted depression self-report measure, and compare how well these 2 assessment methods predict patient outcomes.

Methods: A total of 547 individuals completed smartphone-based measures, including the Patient Health Questionnaire-2 (PHQ-2) modified for daily administration, the PHQ-9, and the Sheehan Disability Scale. Multilevel factor analyses evaluated the reliability of latent depression based on the PHQ-2 (for repeated measures) between weeks 2 and 4 and its correlation with the PHQ-9 at week 4. Regression models predicted week 8 depressive symptoms and disability ratings with daily PHQ-2 and PHQ-9.

Results: The daily PHQ-2 and PHQ-9 are highly reliable (range: 0.80-0.88) and highly correlated ($r=.80$). Findings were robust across demographic groups (age, gender, and ethnic minority status). Daily PHQ-2 and PHQ-9 were comparable in predicting week 8 disability and were independent predictors of week 8 depressive symptoms and disability, though the unique contribution of the PHQ-2 was small in magnitude.

Conclusions: Daily completion of the PHQ-2 is a reasonable proxy for the PHQ-9 and is comparable to the PHQ-9 in predicting future outcomes. Mobile assessment methods offer researchers and clinicians reliable and valid new methods for depression assessment that may be leveraged for measurement-based depression care.

(*J Med Internet Res* 2018;20(6):e10001) doi:[10.2196/10001](https://doi.org/10.2196/10001)

KEYWORDS

Patient Health Questionnaire; depression; mobile health; symptom assessment

Introduction

Background

Practice guidelines for depression treatment call for systematic symptom monitoring to drive treatment adjustment, known as measurement-based care, an approach that improves patient outcomes [1-3]. In addition, recent value-based payment reforms have led to the creation of incentives for measurement-based care for depression by major payors such as the Centers for Medicare and Medicaid services and several large private insurers [3]. Yet, the majority of mental health providers do not use symptom rating scales, and in the few settings where scales are used routinely, they may be administered too infrequently to inform clinical decision making [3,4]. Chief among the reasons that psychiatrists and psychologists report that they do not use symptom measures is that they consider it too time-consuming or burdensome to administer measures [4,5].

For symptom measures to be clinically actionable and drive measurement-based care, the measures need to be reliable, current, interpretable, and sensitive to change [6]. In addition, experts have called for improving symptom measures by making them more brief [6]. Evidence suggests that measurement-based care is most effective when measures are completed frequently by patients in the outpatient setting, feedback is provided to both patients and clinicians, and progress is monitored over time [7]. However, these guidelines have not been fully implemented in real-world practice. Among large-scale programs identified as exemplars of measurement-based care, symptom data are often infrequently collected by providers—only when patients present to clinic or when ordered by a provider. Furthermore, feedback is usually available only to clinicians [6]. Many of these limitations can be overcome by mobile assessment tools on consumer devices that allow individuals to track their own symptoms. Wide-scale adoption of mobile mood assessment has the potential to alleviate the time burden on clinicians, support real-time symptom monitoring on a daily or more frequent basis even for people who are not presenting for clinic-based services, and may have intrinsic therapeutic value in activating patients [8,9]. However, it is unclear how brief daily ratings relate to established clinical measures and, thus, whether they may constitute useful tools for driving measurement-based care. Likewise, it is not known whether the additional information provided in daily ratings offers better prognostic information on patient outcomes that may be useful in guiding treatment decisions.

Several small-scale studies exploring associations between daily depression measures and standardized scales have yielded inconsistent results. For example, 2 small studies among patients in specialty care for depression found that brief daily measures were associated with the Patient Health Questionnaire-9 (PHQ-9) [10,11]. Likewise, a small study of patients with bipolar disorder in specialty care found that daily mood ratings were associated with a clinician-administered Hamilton Depression Rating Scale [12]. In contrast, in a community-based sample, PHQ-9 scores were not associated with 15 items indicative of depression administered by ecological momentary assessment twice daily [13]. Limitations of earlier work include

the small scale of studies and the restricted range of depression symptoms in clinic-based samples.

Objectives

In this study, the 2 primary aims were (1) to examine the association between the PHQ-9, which is the most widely adopted self-report measure that is typically given at most every 2 weeks, and a daily depressive symptom measure; and (2) to compare how these 2 assessment methods perform in predicting patient outcomes. This study, which serves as a *proof of concept* for the reliability and predictive ability of daily depression symptom measurement, is a secondary analysis of an existing dataset from a large national community-based sample that includes daily depression measures.

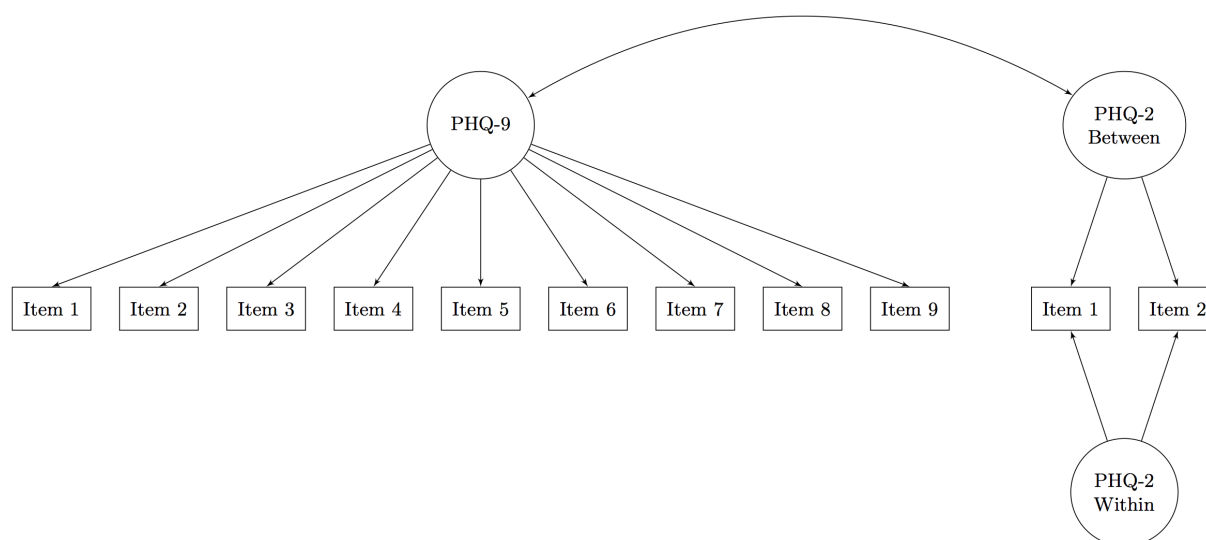
Methods

Data Source and Measures

Data came from a fully remote trial of smartphone-based depression apps in which individuals with a PHQ-9 score of 5 or greater were recruited through a Web-based interface (NCT00540865). Participants were recruited across the United States via 4 approaches: traditional approaches (written advertisements in city buses, newspapers, and Craigslist throughout the United States: 88.98%, 2601/2923); social networking methods (regular postings on sites such as Facebook and Twitter, and contextual-targeting methods to identify and directly push recruitment advertisements to potential participants, based on their Twitter and other social media comments: 0.41%; 12/2923); search engine–based methods (Google Adwords: 0.31%, 9/2923); and unanticipated means (eg, referral or own search: 10.30%, 301/2923). Of those recruited, 1098 met eligibility criteria for the trial and were enrolled. Ethical approval was granted by the University of California San Francisco Committee for Human Research, and details of the trial, including methods for obtaining participants' informed consent, have been previously published [14,15].

Participants completed the PHQ-9 and Sheehan Disability Scale at specified timepoints (eg, weeks 4 and 8) and received US \$20 in Amazon gift vouchers for completing the assessment at each of these timepoints. The PHQ-9 is a valid measure of depression symptoms that is widely used for depression screening and treatment monitoring [16,17]. Individuals report how often over the last 2 weeks they have experienced each of 9 core symptoms of major depression. The Sheehan Disability Scale is a 3-item measure of functioning in work, social, and health domains that has been validated in medical and psychiatric populations [18,19]. Participants also self-reported depressed mood on a daily basis via a modified Patient Health Questionnaire-2 (PHQ-2). The PHQ-2, which consists of the first 2 items of the PHQ-9, is a valid screening tool for detecting depressive disorders and is sensitive to change over time [20,21]. The PHQ-2 was modified for daily administration by changing the timeframe to “yesterday” and the response options to a 5-point Likert-type scale anchored at “not at all” (1) to “most of the day” (5). Participants did not receive an incentive for completion of daily measures.

Figure 1. Path diagram for the factor analysis of the Patient Health Questionnaire-2 (PHQ-2) and Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 was measured at week 4. The PHQ-2 was measured daily for 14 days before the PHQ-9. Rectangles represent observed variables and ovals represent latent variables. Single-headed arrows are factor loadings and double-headed arrows are correlations. Because participants responded to the PHQ-2 repeatedly over a 2-week period, within-patient variability (ie, variability day-to-day) can be separated from between-patient variability (ie, variability in average PHQ-2 scores). Residual variances were estimated in the model but not included in the path diagram.



Participants were excluded from the sample if they were missing a PHQ-9 at week 4 or if they did not have any daily PHQ-2 scores within 14 days preceding their week 4 PHQ-9 score. In addition, participants were excluded from the regression analysis if they were missing the PHQ-9 or Sheehan score at week 8 (see below for a description of the analyses).

Data Analysis

To address our first aim, we compared week 4 PHQ-9 with all PHQ-2 scores within 14 days preceding the week 4 PHQ-9 (the timeframe for PHQ-9 response). Specifically, we fit a multilevel confirmatory factor analysis [22] for the PHQ-2 repeated measures and single-level confirmatory factor analysis for the PHQ-9 to evaluate the reliability and correlation of latent depression across these measures (see Figure 1 for a path diagram). The PHQ-2 and PHQ-9 models were estimated in a single analysis, producing a single model and set of fit statistics. We computed reliability of the PHQ-9, between-persons reliability of the PHQ-2 (ie, average depression over a given time period for a specific person), within-persons reliability of the PHQ-2 (ie, reliability of day-to-day variability in depression for a specific person), and the correlation between the 2 latent depression scores (similar to comparing the PHQ-9 with the average of the PHQ-2 over 2 weeks).

We also used a multiple group confirmatory factor analysis to examine differences in reliability and the correlation between the PHQ-2 and PHQ-9 for groups based on age, gender, and racial or ethnic minority status. Separate multiple group models were used to compare the following groups: racial or ethnic minority versus white, men versus women, and age 55+ years versus <55 years. The multiple group models compared an unconstrained model with a constrained model. In the unconstrained model, all factor loadings, intercepts, and residuals were uniquely estimated in each group. In the constrained model, all factor loadings, intercepts, and residuals

were constrained to be equal across groups. Given that reliability is a function of factor loadings and residual variances [23] and that it is generally inappropriate to constrain residuals but not intercepts across groups [23], we chose to constrain all parameters.

For our second aim, we evaluated how well each depression measure (PHQ-9 at week 4 and daily PHQ-2 from weeks 2 to 4) predicted PHQ-9 and the Sheehan Disability scale at week 8. One potential advantage of repeatedly administering the PHQ-2 is that the repeated measures can be summarized with a variety of indicators. For example, in addition to the mean PHQ-2, we can examine the predictive ability of parameters such as the score trajectory over time (slope) or highest PHQ-2 score (maximum). Therefore, to construct predictors, we summarized the daily PHQ-2 using the mean daily value, SD, minimum, maximum, and linear slope. The first primary model included the PHQ-9 and PHQ-2 mean as predictors and the second included the PHQ-9; PHQ-2 mean; and the PHQ-2 SD, minimum, maximum, and slope as predictors. We used seemingly unrelated regression [24] to estimate the models, which is a multivariate regression approach that allowed to simultaneously fit the regression model for both week 8 PHQ-9 and the Sheehan Disability scale.

Results

Participants

Among 1098 adult participants, 545 individuals completed the PHQ-9 at week 4 and at least one PHQ-2 in the preceding 2 weeks. Of these, 3 individuals were missing demographic data and therefore could not be included in factor analyses with demographic data. The final dataset has 2992 observations on 545 participants. Participants completed an average of 5.5 daily measurements (see Table 1 for distribution of responses). Among the final sample, 78.3% (427/545) were women and

38.1% (208/545) were identified as racial or ethnic minorities, which is similar to national rate of 39% [25]. The average age was 33 years (SD 11) and 6.1% (33/545) participants were 55 years of age or older. The average PHQ-9 score at baseline was 13.9 (SD 5.0) and 77.8% (424/545) had a PHQ-9 score of 10 or greater. Participants were predominantly employed (68.3%, 372/545) and most (60.0%, 327/545) had at least a bachelor's degree. Income was widely distributed (\leq \$20,000: 22.6% [123/545]; \$20,000-\$50,000: 18.0%, [98/545]; \$50,000-\$80,000: 10.0% [55/545]; $>$ \$80,000: 4.4% [24/545]; missing: 44.4% [242/545]). A minority (27.3%, 149/545) were married or partnered.

Reliability

Table 2 presents the results of the factor analysis of the week 4 data (see Figure 1 for a path diagram). To identify the model, we constrained the mean of each latent variable to 0 and the variance to 1. Given that the PHQ-2 had only 2 indicators, we constrained the loadings at the between level to be equal and the loadings at the within level to be equal. Overall fit of the model was good as indicated by a comparative fit index=0.94, Tucker-Lewis index=0.93, and root mean square error of approximation=0.04. Factor loadings were generally strong for the PHQ-9, with the weakest loadings being for items 8 and 9 (psychomotor symptoms and suicidal ideation). Loadings for the PHQ-2 were strong at both levels.

The PHQ-9 and between-persons PHQ-2 were strongly correlated, $r=.8$ ($P<.001$; see Figure 2). Reliability was also high. Specifically, reliability for the PHQ-2 was estimated as 0.82 (within-persons) and 0.86 (between-persons). Consequently, day-to-day variability in PHQ-2 and the PHQ-2 average over 2 weeks were reliable. Reliability for the PHQ-9 was estimated as 0.88 (see Figure 3 and Table 2).

For all but the minority versus white group comparisons, the unconstrained model fits better than the constrained model: men

versus women, $\chi^2_{32}=84.0$, $P<.001$; 55+ years versus <55 years, $\chi^2_{32}=57.8$, $P=.004$; minority versus white, $\chi^2_{32}=38.0$, $P=.21$. This suggests that men and women and participants who are aged 55 years and above and participants who are <55 years have a statistically different factor loadings, intercepts, residuals, and reliabilities. In contrast, participants from racial or ethnic minority groups and white participants did not statistically differ.

Figure 2 shows the estimated correlation by group between the PHQ-9 and PHQ-2 based on the unconstrained model. Figure 3 shows the estimated reliabilities by group based on the unconstrained model. The differences in the correlations and reliabilities were generally small and likely of little practical significance. The most notable difference is between men and women for the within-person PHQ-2 reliability. Specifically, men had a lower estimated reliability but also more uncertainty (ie, wider CI) than women, which is likely because of the fact that there were fewer men than women in the sample.

Predicting Future Functioning

Week 4 PHQ-2 and PHQ-9 were statistically significant predictors of both the PHQ-9 and the Sheehan Disability Scale (SDS) at week 8 (Table 3). This was true when the week 4 predictors were entered into the regression alone and together (compare Models 1 and 2 with Model 3 in Table 3). R^2 for Model 1, where PHQ-9 at week 4 was the only predictor, was 0.49 (PHQ-9 week 8) and 0.37 (SDS week 8). R^2 for Model 3, where both week 4 PHQ-9 and PHQ-2 were predictors, was 0.50 (PHQ-9 week 8) and 0.41 (SDS week 8). This suggests that PHQ-2 does not provide much predictive information above and beyond the PHQ-9 when predicting future PHQ-9 values. When predicting future functional disability with the SDS, a more global measure of outcomes, the PHQ-2 does add somewhat to the predictive information of the PHQ-9.

Table 1. Distribution of daily Patient Health Questionnaire-2 (PHQ-2) ratings completed.

Number of daily ratings submitted	Participants, n (%)	Cumulative, n (%)
1	40 (7.3)	40 (7.3)
2	22 (4.0)	62 (11.4)
3	27 (5.0)	89 (16.3)
4	45 (8.3)	134 (24.6)
5	79 (14.5)	213 (39.1)
6	109 (20.0)	322 (59.1)
7	199 (36.5)	521 (95.6)
8	18 (3.3)	539 (98.9)
9	2 (0.4)	541 (99.3)
10	2 (0.4)	543 (99.6)
11	1 (0.2)	544 (99.8)
12	1 (0.2)	545 (100.0)
13	0 (0.0)	545 (100.0)
14	0 (0.0)	545 (100.0)

Table 2. Factor analysis at week 4.

Measure	Loadings ^a	Intercept	Variance ^b	Reliability
PHQ-9^c		0.0 ^d	1.0 ^d	0.88
Item 1	0.60	1.07	0.20	
Item 2	0.63	1.12	0.22	
Item 3	0.59	1.25	0.53	
Item 4	0.65	1.39	0.40	
Item 5	0.65	1.07	0.58	
Item 6	0.70	1.03	0.44	
Item 7	0.56	0.90	0.45	
Item 8	0.34	0.34	0.32	
Item 9	0.37	0.34	0.36	
PHQ-2^c between		0.0 ^d	1.0 ^d	0.86
Item 1	0.85 ^d	2.20	0.05	
Item 2	0.85 ^d	2.20	0.05	
PHQ-2 within		0.0 ^d	1.0 ^d	0.82
Item 1	0.66 ^d	—	0.20	
Item 2	0.66 ^d	—	0.20	
Participants (N)	545			
Comparative fit index	0.94			
Tucker-Lewis index	0.93			
Root mean square error of approximation	0.04			

^aAll estimated coefficients were statistically significant.

^bVariances on rows with items are residual variances and variances on other rows are variances.

^cPHQ-9: Patient Health Questionnaire-9.

^dConstrained for identification of latent variables.

^ePHQ-2: Patient Health Questionnaire-2.

Figure 2. Correlation between Patient Health Questionnaire-2 and Patient Health Questionnaire-9. Interval estimates are 95% CIs. Point estimates are rounded to 2 digits. PHQ: Patient Health Questionnaire; <55 refers to participants under the age of 55 years; 55+ refers to participants aged 55 years or older.

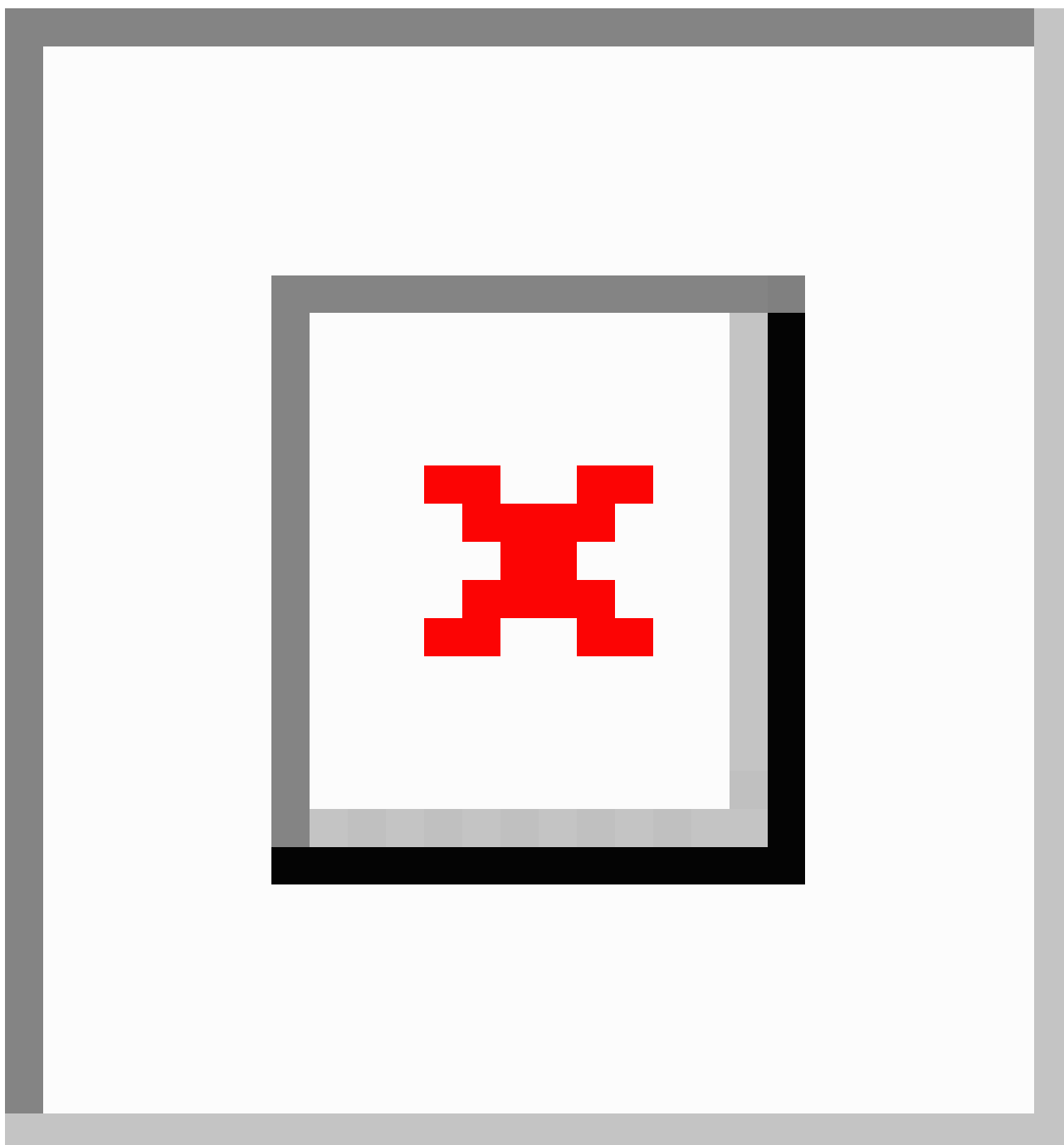


Figure 3. Reliability estimates. Interval estimates are 95% CIs. Point estimates are rounded to 2 digits. PHQ: Patient Health Questionnaire; <55 refers to participants under the age of 55 years; 55+ refers to participants aged 55 years or older.

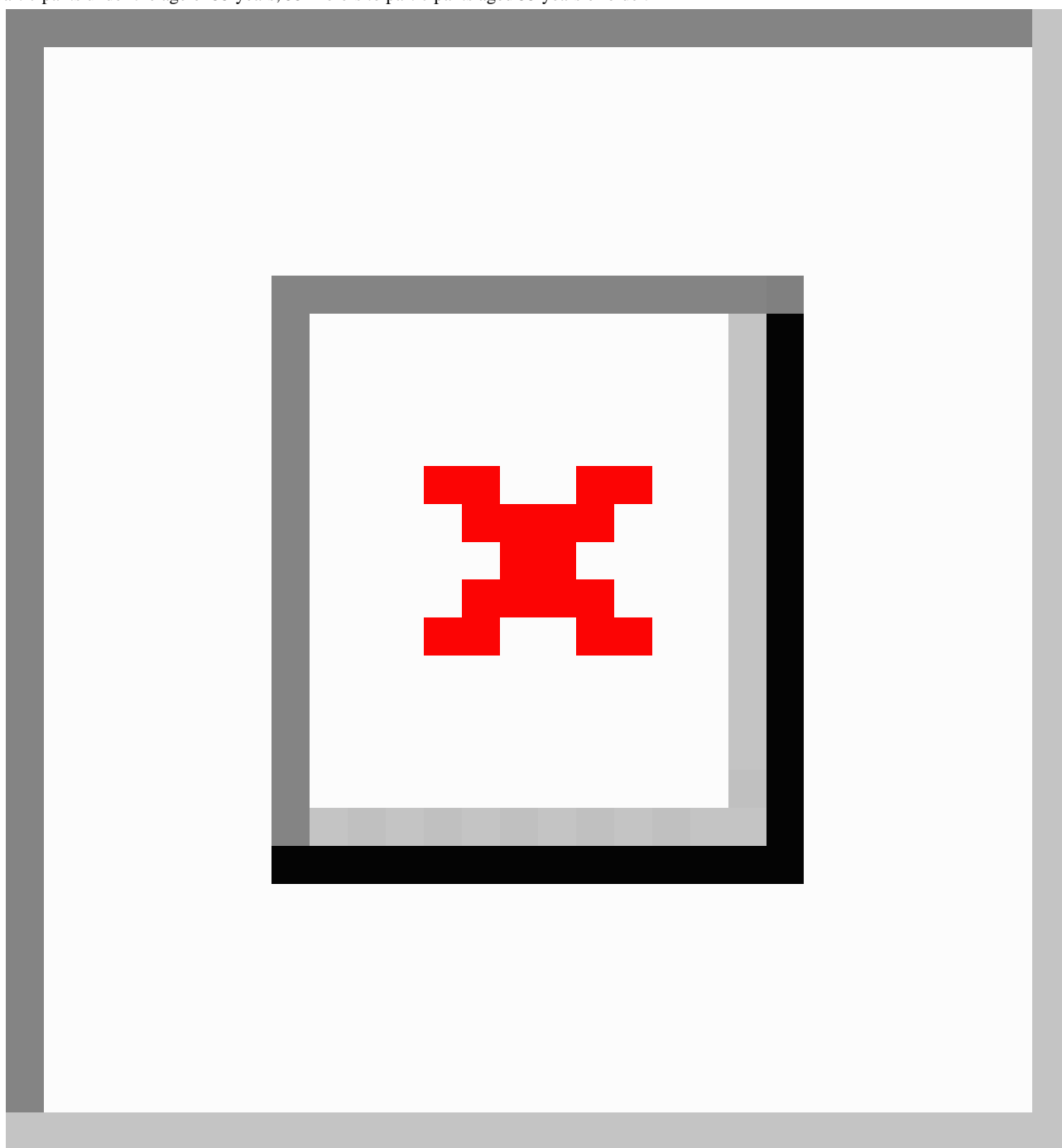


Table 3. Prediction of week 8 outcomes based on depression scores at week 4.

Outcome ^a and predictor ^{b,c}	Model 1		Model 2		Model 3		Model 4	
	Coefficient ^d	P value	Coefficient ^d	P value	Coefficient ^d	P value	Coefficient ^d	P value
PHQ-9^e								
PHQ-9 mean	0.7	<.001	—		0.6	<.001	—	
PHQ-2 ^f mean	—		0.4	<.001	0.1	.001	0.3	.002
PHQ-2 slope	—		—		—		0.5	.43
PHQ-2 max ^g	—		—		—		−0.1	.22
PHQ-2 min ^h	—		—		—		0.2	.11
PHQ-2 SD	—		—		—		0.5	.04
R ²	0.49		0.34		0.50		0.37	
Sheehan Disability Scale								
PHQ-9 mean	2.7	<.001	—		1.9	<.001	—	
PHQ-2 mean	—		1.7	<.001	0.8	<.001	1.5	.001
PHQ-2 slope	—		—		—		0.8	.77
PHQ-2 max	—		—		—		0.6	.16
PHQ-2 min	—		—		—		−0.4	.43
PHQ-2 SD	—		—		—		−1.2	.26
R ²	0.37		0.32		0.41		0.34	
N ⁱ	352		362		348		351	

^aOutcomes were all measured at week 8.^bPredictors were all measured at week 4.^cEach model has a different set of predictors. The coefficient and P values are listed only for those predictors that were included in the Model.^dUnstandardized regression coefficients are reported.^ePHQ-9: Patient Health Questionnaire-9.^fPHQ-2: Patient Health Questionnaire-2.^gMax: maximum.^hMin: minimum.ⁱN: Number of participants.

One potential advantage of daily symptom monitoring is that this allows patients' scores to be characterized in more nuanced ways in addition to the daily average, such as the slope over 2 weeks or the minimum or maximum value. Model 4 shows that slope, maximum, minimum, and SD of the PHQ-2 over the 2 weeks preceding week 4 were not significant predictors of either the PHQ-9 or the SDS at week 8, suggesting no additional predictive ability of these features beyond the daily average PHQ-2.

Discussion

Principal Findings

Our findings establish that both the daily PHQ-2 mean and daily PHQ-2 variability are reliable measures of depressive symptoms. The daily PHQ-2 mean is closely correlated with the PHQ-9, the most commonly used measure for assessing depressive symptom severity. These findings hold across a range of demographic groups. As such, we have identified that a brief

daily measure can provide current, accurate information on depressive symptom status, thus fulfilling several of the qualities needed for a measure to be clinically actionable and inform measurement-based care [6].

Our findings further demonstrate that daily PHQ-2 mean between weeks 2 and 4 is a strong predictor of depressive symptoms and of overall functioning at week 8 that is independent of the PHQ-9. Although it is an independent predictor of week 8 outcomes, the magnitude of the independent contribution of the PHQ-2 is small when a PHQ-9 is obtained at week 4. Likewise, despite demonstrating that daily PHQ-2 variability is a reliable measure, none of the daily indicators we examined (minimum, maximum, slope, and SD) improved prediction of week 8 outcomes above and beyond the daily PHQ-2 mean. Therefore, it appears that the predictive value of the PHQ-2 in this sample is related to the stable information obtained from the average of the PHQ-2 across the 2 weeks and not the repeated assessments. The daily PHQ-2 is somewhat less strongly predictive of week 8 depressive symptoms than

the week 4 PHQ-9, although this is not surprising given that week 8 depressive symptoms were measured with the PHQ-9. Importantly, when we examined a more global outcome measure at week 8, the Sheehan Disability Scale, the daily PHQ-2 performs comparably to the week 4 PHQ-9 in predicting the week 8 outcomes. Therefore, the daily PHQ-2 may serve as a proxy measure that can reasonably substitute for the PHQ-9. Given the favorable psychometric properties and strong predictive ability of each measure, researchers and clinicians may wish to consider response rate when selecting a mobile depression measure. In this sample, the response rate for the daily PHQ-2 was greater than the response rate for the PHQ-9 at all timepoints [26], further supporting the utility of the daily PHQ-2. However, it is possible that the response rates in this sample may differ from those that would be obtained in a clinical sample.

Limitations

Our findings are based on a large national sample of community-dwelling individuals who reported a range of depressive symptom scores, which represent a strength for the psychometric analyses we conducted. However, certain limitations also apply. The sample consisted of moderately depressed individuals, 78.3% (427/545) of whom score 10 or above on the PHQ-9. This restricted the response range of participants; however, it is also similar to the scores that would be observed among patients in clinical settings, and the reliability of the PHQ-9 in this study (0.88) is similar to reported values ranging from 0.79 to 0.86 [27]. The factor analysis models were based only on PHQ-9 scores at a single timepoint (week 4) and not all participants completed all daily PHQ-2 measures. Data came from a broad community-based sample that increased generalizability; however, detailed clinical or diagnostic information was not available to characterize participants. Although it is likely that individuals who volunteered to participate in this study differed from nonparticipants in certain characteristics, it is unknown to what extent these differences would have affected the psychometric properties of the measures or the relationship between the daily PHQ-2 measure and the PHQ-9 or disability scores. Similarly, we were unable to separate variability due to completing an assessment from the variability due to responding to these specific questions about depression. Future research that seeks to separate these sources of variability could have participants respond to depression items as well as neutral items to determine which responses are predictive of future symptoms.

We analyzed the modified PHQ-2 as the daily measure based on the availability of this measure in the existing dataset. Although our findings support the use of a daily PHQ-2 for monitoring depressive symptoms, these findings may not entirely translate to other daily mood measures. Our findings can serve as a *proof of concept* for brief daily depression symptom measurement; however, future studies should assess other measures. Despite receiving daily notifications,

participants completed an average of 5.5 daily ratings. This study did not assess patients' experiences with daily depression monitoring; however, previous research does support the acceptability of monitoring depressive symptoms on a daily basis [28]. In our ongoing work with a weekly symptom monitoring app, one of the most common requests from patients has been for the addition of a daily mood measure [29]. Although, on average, participants completed fewer than half of the daily ratings, this represents completion of a symptom measure every 2 to 3 days, which is considerably more frequent than clinic-based assessment. Treatment dropout is a serious concern with clinic-based services, given that the modal number of visits for depressed patients receiving behavioral interventions is 1 [30,31]. Daily mobile symptom monitoring may serve to promote better engagement in care, and future research in clinical samples should examine its potential to promote retention among patients who may otherwise discontinue services prematurely.

Future Directions

Our findings support the use of a daily PHQ-2 as an alternative to the PHQ-9 for monitoring depressive symptoms. Our finding that daily average PHQ-2 is an independent predictor of week 8 depressive symptoms and functioning demonstrates that a daily measure does provide additional information. Although we found that daily depressive symptom variability as measured by the PHQ-2 did not substantially improve prediction of week 8 outcomes, it is plausible that indicators of daily variability in depressive symptoms may have greater utility at certain phases of treatment, for example, as indicators of incomplete treatment response that may be associated with higher risk of relapse. Future research should examine other types of brief mobile mood measures and passively collected behavioral indicators to evaluate how such measures can be combined with or substitute for self-report measures and whether novel assessment (eg, Photographic Affect Meter [32]) may be more engaging to patients and facilitate long-term self-monitoring. This study serves as a model for researchers examining other brief mobile assessment methods, and future work should link daily mood assessments with clinical information to determine the clinical utility of new assessment methods and to determine the optimal frequency of measurement.

Conclusions

This study represents an important step in establishing a daily depressive symptom measure to drive measurement-based care. Mobile assessment methods that are convenient to administer, reliable, and predict functioning may facilitate the adoption of measurement-based depression care and improve the quality of care and health outcomes. Because optimal outcomes from measurement-based care are achieved when results of such measures are provided to clinicians and incorporated in a structured manner into clinical encounters [7,33], future research should explore the incorporation of daily mood measures within the context of comprehensive measurement feedback systems.

Acknowledgments

This project was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2TR000421) through the Clinical and Translational Science Awards Program and the National Institute of Mental Health (MH100466). The authors thank M Brander (University of California, San Francisco, UCSF) and A Bodepudi (UCSF) for assistance in data monitoring; M Gross (UCSF) and J Camire (Ekho.me) for their help in participant recruitment; J Steinmetz (Ekho.me) for database architecture; D Ziegler (UCSF) for helping with app deployment; D Albert (UCSF) for assistance in Web design; C Catledge (UCSF) for administrative oversight; I Elson, L Kaye, S Goobich, and the rest of the Ginger.io team for helping with data collection and partnering on this project. The authors also would especially like to thank all the participants whose time and efforts made this work possible.

Conflicts of Interest

AMB, SAB, JAA, and PAA report no competing interests. DCA is a cofounder with equity stake in a technology company, Lyssn.io, focused on tools to support training, supervision, and quality assurance of psychotherapy and counseling. No product or service from Lyssn.io was used in this study.

References

1. American Psychiatric Association. 2010. Practice Guideline for the Treatment of Patients with Major Depressive Disorder, third edition URL: https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf [accessed 2018-05-10] [WebCite Cache ID 6zIuWN2Y6]
2. Guo T, Xiang YT, Xiao L, Hu CQ, Chiu HF, Ungvari GS, et al. Measurement-Based Care Versus Standard Care for Major Depression: A Randomized Controlled Trial With Blind Raters. *Am J Psychiatry* 2015 Oct;172(10):1004-1013. [doi: [10.1176/appi.ajp.2015.14050652](https://doi.org/10.1176/appi.ajp.2015.14050652)] [Medline: [26315978](https://pubmed.ncbi.nlm.nih.gov/26315978/)]
3. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A Tipping Point for Measurement-Based Care. *Psychiatr Serv* 2017 Feb 01;68(2):179-188. [doi: [10.1176/appi.ps.201500439](https://doi.org/10.1176/appi.ps.201500439)] [Medline: [27582237](https://pubmed.ncbi.nlm.nih.gov/27582237/)]
4. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry* 2008 Dec;69(12):1916-1919. [Medline: [19192467](https://pubmed.ncbi.nlm.nih.gov/19192467/)]
5. Hatfield DR, Ogles BM. Why some clinicians use outcome measures and others do not. *Adm Policy Ment Health* 2007 May;34(3):283-291. [doi: [10.1007/s10488-006-0110-y](https://doi.org/10.1007/s10488-006-0110-y)] [Medline: [17211715](https://pubmed.ncbi.nlm.nih.gov/17211715/)]
6. Fortney J, Sladek R, Unutzer J. The Kennedy Forum. 2015. Fixing Behavioral Health Care in America: A National Call for Measurement-Based Care in the Delivery of Behavioral Health Services URL: http://thekennedyforum-dot-org.s3.amazonaws.com/documents/KennedyForum-BehavioralHealth_FINAL_3.pdf [accessed 2018-01-31] [WebCite Cache ID 6wsCCiC34]
7. Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009 Jul;195(1):15-22 [FREE Full text] [doi: [10.1192/bjp.bp.108.053967](https://doi.org/10.1192/bjp.bp.108.053967)] [Medline: [19567889](https://pubmed.ncbi.nlm.nih.gov/19567889/)]
8. Eisen SV, Dickey B, Sederer LI. A self-report symptom and problem rating scale to increase inpatients' involvement in treatment. *Psychiatr Serv* 2000 Mar;51(3):349-353. [doi: [10.1176/appi.ps.51.3.349](https://doi.org/10.1176/appi.ps.51.3.349)] [Medline: [10686242](https://pubmed.ncbi.nlm.nih.gov/10686242/)]
9. Valenstein M, Adler DA, Berlant J, Dixon LB, Dulit RA, Goldman B, et al. Implementing standardized assessments in clinical care: now's the time. *Psychiatr Serv* 2009 Oct;60(10):1372-1375. [doi: [10.1176/ps.2009.60.10.1372](https://doi.org/10.1176/ps.2009.60.10.1372)] [Medline: [19797378](https://pubmed.ncbi.nlm.nih.gov/19797378/)]
10. 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](https://doi.org/10.1016/j.jad.2015.01.033)] [Medline: [25679202](https://pubmed.ncbi.nlm.nih.gov/25679202/)]
11. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder. *JMIR Ment Health* 2015 Mar 24;2(1):e8 [FREE Full text] [doi: [10.2196/mental.3889](https://doi.org/10.2196/mental.3889)] [Medline: [26543914](https://pubmed.ncbi.nlm.nih.gov/26543914/)]
12. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. *Psychiatry Res* 2014 Jun 30;217(1-2):124-127. [doi: [10.1016/j.psychres.2014.03.009](https://doi.org/10.1016/j.psychres.2014.03.009)] [Medline: [24679993](https://pubmed.ncbi.nlm.nih.gov/24679993/)]
13. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr DC. The Relationship between Clinical, Momentary, and Sensor-based Assessment of Depression. *Int Conf Pervasive Comput Technol Healthc* 2015 Aug;2015:229-232 [FREE Full text] [doi: [10.4108/icst.pervasivehealth.2015.259034](https://doi.org/10.4108/icst.pervasivehealth.2015.259034)] [Medline: [26640739](https://pubmed.ncbi.nlm.nih.gov/26640739/)]
14. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ Innov* 2016 Jan;2(1):14-21 [FREE Full text] [doi: [10.1136/bmjinnov-2015-000098](https://doi.org/10.1136/bmjinnov-2015-000098)] [Medline: [27019745](https://pubmed.ncbi.nlm.nih.gov/27019745/)]
15. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. *J Med Internet Res* 2016 Dec 20;18(12):e330 [FREE Full text] [doi: [10.2196/jmir.6482](https://doi.org/10.2196/jmir.6482)] [Medline: [27998876](https://pubmed.ncbi.nlm.nih.gov/27998876/)]

16. 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](#)]
17. Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry* 2010;32(4):345-359. [doi: [10.1016/j.genhosppsych.2010.03.006](#)] [Medline: [20633738](#)]
18. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *Int Clin Psychopharmacol* 1996 Jun;11(Suppl 3):89-95. [Medline: [8923116](#)]
19. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *Int J Psychiatry Med* 1997;27(2):93-105. [Medline: [9565717](#)]
20. Kroenke K, Spitzer RL, Williams JBW. 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](#)]
21. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res* 2005 Feb;58(2):163-171. [doi: [10.1016/j.jpsychores.2004.09.006](#)] [Medline: [15820844](#)]
22. Geldhof GJ, Preacher KJ, Zyphur MJ. Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychol Methods* 2014 Mar;19(1):72-91. [doi: [10.1037/a0032138](#)] [Medline: [23646988](#)]
23. Brown TA. Confirmatory factor analysis for applied research. Second edition. New York: The Guilford Press; 2015.
24. Zellner A. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *J Am Stat Assoc* 1962 Jun;57(298):348-368.
25. United States Census Bureau. QuickFacts URL: <https://www.census.gov/quickfacts/fact/table/US/PST045216> [accessed 2018-05-01] [WebCite Cache ID 6z64YnCrw]
26. Pratap A, Anguera JA, Renn BN, Neto EC, Volponi J, Mooney SD, et al. The feasibility of using smartphones to assess and remediate depression in Hispanic/Latino individuals nationally. In: *UbiComp'17. 2017 Sep 11 Presented at: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers; September 11 - 15, 2017; Maui, Hawaii p. 854-860.* [doi: [10.1145/3123024.3127877](#)]
27. Chen TM, Huang FY, Chang C, Chung H. Using the PHQ-9 for depression screening and treatment monitoring for Chinese Americans in primary care. *Psychiatr Serv* 2006 Jul;57(7):976-981. [doi: [10.1176/ps.2006.57.7.976](#)] [Medline: [16816282](#)]
28. Bauer AM, Iles-Shih M, Ghomi RH, Rue T, Grover T, Kincler N, et al. Acceptability of mHealth augmentation of Collaborative Care: A mixed methods pilot study. *Gen Hosp Psychiatry* 2017 Nov 24;51:22-29. [doi: [10.1016/j.genhosppsych.2017.11.010](#)] [Medline: [29272712](#)]
29. Bauer A, Hodsdon S, Bechtel J, Fortney J. Applying the Principles for Digital Development: Case Study of a Smartphone App to Support Collaborative Care for Rural Patients With Posttraumatic Stress Disorder or Bipolar Disorder. *J Med Internet Res* 2018 Jun 06;20(6):e10048 [FREE Full text] [doi: [10.2196/10048](#)] [Medline: [29875085](#)]
30. Olfson M, Marcus SC, Druss B, Elinson L, Tanielian T, Pincus HA. National trends in the outpatient treatment of depression. *JAMA* 2002 Jan 09;287(2):203-209. [Medline: [11779262](#)]
31. Simon GE, Ludman EJ. Predictors of early dropout from psychotherapy for depression in community practice. *Psychiatr Serv* 2010 Jul;61(7):684-689. [doi: [10.1176/ps.2010.61.7.684](#)] [Medline: [20592003](#)]
32. Pollak JP, Adams P, Gay G. PAM: A photographic affect meter for frequent, in situ measurement of affect. 2011 May Presented at: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI); 2011; Vancouver, BC, Canada p. 725-734. [doi: [10.1145/1978942.1979047](#)]
33. 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](#)]

Abbreviations

PHQ-2: Patient Health Questionnaire-2
PHQ-9: Patient Health Questionnaire-9
SDS: Sheehan Disability Scale
UCSF: University of California, San Francisco

Edited by J Torous, R Calvo, M Czerwinski, G Wadley; submitted 05.02.18; peer-reviewed by M Valenstein, R Calvo; comments to author 12.03.18; revised version received 26.03.18; accepted 27.03.18; published 19.06.18.

Please cite as:

Bauer AM, Baldwin SA, Anguera JA, Areán PA, Atkins DC

Comparing Approaches to Mobile Depression Assessment for Measurement-Based Care: Prospective Study

J Med Internet Res 2018;20(6):e10001

URL: <http://www.jmir.org/2018/6/e10001/>

doi: [10.2196/10001](https://doi.org/10.2196/10001)

PMID: [29921564](https://pubmed.ncbi.nlm.nih.gov/29921564/)

©Amy M Bauer, Scott A Baldwin, Joaquin A Anguera, Patricia A Areán, David C Atkins. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 19.06.2018. 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.

Original Paper

The Burden of a Remote Trial in a Nursing Home Setting: Qualitative Study

Susie Donnelly¹, PhD; Brenda Reginatto¹, MSc; Oisín Kearns¹, MA; Marie Mc Carthy², MBA; Bill Byrom², PhD; Willie Muehlhausen², DVM; Brian Caulfield¹, PhD

¹Applied Research for Connected Health, University College Dublin, Dublin, Ireland

²ICON plc, Dublin, Ireland

Corresponding Author:

Susie Donnelly, PhD

Applied Research for Connected Health

University College Dublin

NexusUCD Belfield Office Park

Block 9/10 Clonskeagh

Dublin, D04 V2N9

Ireland

Phone: 353 1 716 5403

Email: susie.donnelly@ucd.ie

Abstract

Background: Despite an aging population, older adults are typically underrecruited in clinical trials, often because of the perceived burden associated with participation, particularly travel associated with clinic visits. Conducting a clinical trial remotely presents an opportunity to leverage mobile and wearable technologies to bring the research to the patient. However, the burden associated with shifting clinical research to a remote site requires exploration. While a remote trial may reduce patient burden, the extent to which this shifts burden on the other stakeholders needs to be investigated.

Objective: The aim of this study was to explore the burden associated with a remote trial in a nursing home setting on both staff and residents.

Methods: Using results from a grounded analysis of qualitative data, this study explored and characterized the burden associated with a remote trial conducted in a nursing home in Dublin, Ireland. A total of 11 residents were recruited to participate in this trial (mean age: 80 years; age range: 67-93 years). To support research activities, we also recruited 10 nursing home staff members, including health care assistants, an activities co-ordinator, and senior nurses. This study captured the lived experience of this remote trial among staff and residents and explored the burden associated with participation. At the end of the trial, a total of 6 residents and 8 members of staff participated in semistructured interviews (n=14). They reviewed clinical data generated by mobile and wearable devices and reflected upon their trial-related experiences.

Results: Staff reported extensive burden in fulfilling their roles and responsibilities to support activities of the trial. Among staff, we found eight key characteristics of burden: (1) comprehension, (2) time, (3) communication, (4) emotional load, (5) cognitive load, (6) research engagement, (7) logistical burden, and (8) product accountability. Residents reported comparatively less burden. Among residents, we found only four key characteristics of burden: (1) comprehension, (2) adherence, (3) emotional load, and (4) personal space.

Conclusions: A remote trial in a nursing home setting can minimize the burden on residents and enable inclusive participation. However, it arguably creates additional burden on staff, particularly where they have a role to play in locally supporting and maintaining technology as part of data collection. Future research should examine how to measure and minimize the burden associated with data collection in remote trials.

(*J Med Internet Res* 2018;20(6):e220) doi:[10.2196/jmir.9638](https://doi.org/10.2196/jmir.9638)

KEYWORDS

mHealth; patient burden; remote trial; clinical trial methodology; wearable technology; activity monitors

Introduction

Older adults, particularly those with impaired mobility and cognitive disorders such as dementia, are typically underrecruited in clinical trials, even though they have the greatest need for health care services [1-6]. Reasons for underrepresentation are unclear, but may relate to comorbidities, communication difficulties (due to hearing or vision impairments), and physical immobility that constrains transportation to a research site [7-9]. The perception among health care providers that the clinical research experience is stressful or burdensome can mean that researchers elect not to access patients [10]. Although remote clinical trials may resolve the burden associated with clinical trial participation and make clinical research participation more accessible for traditionally hard-to-reach groups, this has been largely unexplored, particularly in terms of understanding the role of mobile and wearable technology for supporting a partially or wholly remote trial.

Historically, efforts to ascertain and address the burden associated with clinical research have tended to focus on direct risks of clinical interventions or data collection procedures, such as survey participation [11-13]. Lingler et al argued that a complete conceptualization of research burden encompasses not only perceptions regarding direct risks involved in the study but also the indirect burdens [14]. These indirect burdens may vary based on the duration of research, intensity, and invasiveness of study procedures. Moreover, burden may be modulated by the perceived benefit of participation [14]; in other words, the value of participation. Ulrich et al suggested that although institutional review boards regularly evaluate risks and benefits of research on human subjects, how patients perceive burden and benefits of clinical trial participation remains to be extensively and qualitatively examined [15].

As yet, there is no clear and consistent definition of a remote trial in the existing literature. Rather, there are a range of interchangeable descriptive terms ranging from “web-based,” “virtual,” and “nonclinical” trials, to “place-shifted” and “remotely-monitored” trials. Fundamentally, a remote trial follows the underlying principle of placing individuals, rather than investigative sites, at the center of the research process using mobile and wearable technology to support data collection. Covington and Veley claim that a remote patient-centered model offers great potential to advance observational studies and randomized controlled trials [16]. It also offers an opportunity to increase research efficiency through remote patient recruitment and enrollment, retention programs, data collection, and long-term follow-up evaluation. At the same time, this operational approach intensifies patient-centeredness by directly engaging patients in research, overcoming geographic obstacles to connect stakeholders, and incorporating patient input into the research process [16].

There is evidence that some clinical trials have leveraged digital technology to varying extents [16-21]. A study published in 2014 involving researchers from Pfizer marked a clear departure from the traditional clinical model. It sought to purposefully

design and conduct a “virtual” trial. However, recruitment was problematic and of the 5157 individuals who registered on the trial website, only 18 were randomized to treat [20]. Although the online informed consent process was successful, for those patients progressing to that stage, the lack of human interaction in the recruitment process was a major barrier (particularly in population studies that predominantly comprise of older adult subjects) [22,23]. This study demonstrated the need for understanding potential sources of burden in a remote trial. It is proposed that the application of mobile and wearable technologies may help in innovating clinical trials toward a more patient-centered approach. However, this shift in location must consider the unintended consequences, namely, the relevant stakeholders who potentially absorb the burden associated with clinical trial participation. When targeting older adults with impaired mobility and cognitive disorders, such stakeholders may include family members, caregivers, or nursing home staff. Thus, to determine the feasibility of a remote trial, it is essential to understand the burden associated with remote data collection on the stakeholders. As our population ages, nursing homes and residential care settings are increasingly important study sites. Thus, in this study, we aimed to explore the burden associated with a remote trial in a nursing home for staff and residents.

Methods

Remote Trial

Between March 2017 and May 2017, we conducted a remote trial with a sample intervention at one nursing home site over 8 weeks. This trial explored the feasibility of remote data collection and transfer and the associated burden on participants in a novel setting. A falls prevention program acted as the intervention and was conducted by a private physiotherapy company, a third-party service provider. Their program involved weekly exercise classes and a daily exercise program over 6 weeks. The aim of the trial was not to assess the efficacy of this intervention—which was expected to be minimal in such a short period—rather, it was to explore the feasibility of remote data collection and transfer in a nursing home setting. Therefore, a number of technological solutions were selected for supporting activities of the trial (Table 1). These solutions were predominantly consumer devices, except for Quantitative Timed Up and Go (QTUG), a medical grade device for assessing fall risk. These solutions provided outcome measures for fall risk and mobility in older adults, offered a variety of data collection methods, offered an engagement tool, and were conducive to mobile data collection.

A participatory approach was used with the nursing home staff to codesign aspects of this study. Two focus groups were formed in the preliminary phase to explore current workflows and routines, existing assessment methods, and anticipated challenges associated with conducting a trial involving mobile and wearable technologies in this environment.

Ethical Approval

Approval for the remote trial was granted by the University College Dublin Research Ethics Committee in February 2017.

Table 1. Mobile and wearable technology in a remote trial.

Technology	Description	Users	Data collection period
QTUG ^a (Kinesis Health Technologies, Ireland) [24]	A sensor-based medical device that assesses gait, mobility, fall risk, and frailty while conducting a timed up and go performance test	Operated by physiotherapists and worn by residents	Pre- or postassessments conducted at weeks 1 or 8, respectively
Aging Research App (ICON Clinical Research, Ireland in partnership with mPROVE Health, US)	A tablet version of the Age-Related Muscle Loss Questionnaire adapted for self-assessing the impact of muscle loss on activities of daily living	Operated by residents and supported by the researcher	Conducted over weeks 2 and 3
Vivosmart HR (Garmin Ltd, US)	A wrist-worn watch that tracks daily activity, heart rate, and sleep patterns and acts as an engagement tool	Worn by residents, charged twice weekly by staff, and worn by 3× site administrators	Worn 24/7 from weeks 1-8
Galaxy J3 (Samsung, South Korea)	A mobile phone (smartphone) used for downloading data from Vivosmart HR and upload data to the cloud server	Operated by 3× site administrators	Worn 24/7 from weeks 1-8
Covalence (Big Cloud Analytics, US)	A platform for data analytics and visualization used for monitoring data collection and transfer	Operated by 3× site administrators and research team	Throughout the trial (from weeks 1 to 8)

^aQTUG: Quantitative Timed Up and Go.

Sampling and Recruitment

Using convenience sampling, we identified and enrolled a private nursing home in a middle-class suburb in Dublin. The nursing home had a population of 64 residents (as of February 24, 2017; see Table 2 for resident characteristics). A senior member of the nursing team was the first staff member enrolled in the trial. She was assigned the role of the trial co-ordinator (TC) and acted as a central point of contact between the research team and nursing home (see Table 3 for staff roles and responsibilities).

The TC and nursing home physiotherapist screened the resident population using the inclusion and exclusion criteria established by physiotherapists conducting the falls prevention program. Residents were eligible for inclusion if they exhibited (1) increased risk of falls (ie, experienced at least one fall in the last 6 months and/or scored moderate or high risk of falls in their last fall risk [FRASE] screening) and (2) were likely to benefit from taking part in a falls prevention program in the opinion of the TC or nursing home physiotherapist. Residents were excluded if (1) they were bed or chair bound; (2) required assistance of 2 people to walk; (3) had falls due to polypharmacy or unknown reasons; and (4) were clinically unstable according to TC; (5) were not likely to benefit from participating in the falls prevention program, for any reason, in the professional opinion of the TC or the nursing home physiotherapist (6) were fitted with a pacemaker.

The TC and nursing home physiotherapist identified 19 eligible residents (ie, 30% (19/64) of the nursing home population). The TC approached all eligible residents to invite them to meet the researchers. If they agreed, they were then brought to a private visitor's room in the nursing home to meet 2 members of the research team (SD and BR) for an information session. The information session was piloted prior to being rolled out. BR took responsibility of assisting the TC in bringing residents to and from the room. SD explained the study to the residents

using simple language. Researchers used the teach-back method for enhancing comprehension [24-27]. Visuals were used to indicate different components of the trial, featuring an image of the activity (eg, exercise class) and a picture of a person overseeing that activity (eg, the physiotherapists) as well as when and where the activity would take place. Physical devices, such as the Vivosmart HR device (hereafter referred to as the "watch"), were shown to the residents for them to touch and feel. Residents were provided with a summary and a long version of the participant information leaflets. They were given 5 days to review the material after which SD returned to the nursing home to answer any questions and collect written consents. In total, 11 residents were successfully enrolled in the trial (representing a response rate of 58%, 11/19). Of these residents, 5 provided independent consent and 6 provided consent by proxy. Reasons for nonparticipation varied from not being interested in the study to not taking part because fellow residents (ie, friends) were ineligible. Following resident recruitment, TC identified "reliable" health care assistants (HCA) ordinarily assigned to those residents. TC ensured that HCAs from each daytime shift were enrolled so that there was always a member of staff on hand to support trial activities. TC recruited two senior members of nursing staff as site administrators (SA) who were responsible for assisting with study activities and solving technical issues. Two physiotherapists from a private clinic were also enrolled to design and conduct the falls prevention program, and an activities co-ordinator (AC) was assigned for supporting the exercise classes.

In the first 2 weeks of the study, a member of the Big Cloud Analytics (BCA) team conducted 4 visits to the nursing home to provide education to residents on using the watch and to staff on charging the watch and using the mobile phone. They were also on site to set up the devices and address any problems. Prior to being rolled out, the educational process was piloted with an older female (aged 73 years) with low digital literacy.

Table 2. Characteristics of residents.

Characteristic	Value
Age (years), mean (SD)	80 (10)
Gender, n (%)	
Male	4 (36)
Female	7 (64)
Mobility	
Had a fall in past 6 months, n (%)	3 (27)
Uses mobility aid, n (%)	4 (36)
Timed up and go (seconds at baseline), mean (SD)	28 (17)
MMSE^a score, n (%)	
30-25 (normal)	1 (9)
24-21 (mild or early)	4 (36)
20-10 (moderate)	2 (18)
9-0 (severe)	4 (36)
Exercise Class, n (%)	
Weekly attendance (out of 6 classes)	4 (79)

^aMMSE: Mini-Mental State Examination.

Table 3. Staff recruitment, roles, and responsibilities in the remote trial.

Code	Role in the nursing home	Role in the trial	Responsibilities in the trial
TC (n=1)	Senior nursing staff	Trial co-ordinator and site administrator	<ul style="list-style-type: none"> Oversee trial activities daily Act as a central point of contact with the research team Recruit and enroll residents and staff Troubleshoot any technical issues
SA (n=2)	Nursing staff	Site administrator	<ul style="list-style-type: none"> Assist the trial co-ordinator with overseeing trial activities daily Troubleshoot any technical issues
HCA (n=6)	Health care assistant	Support	<ul style="list-style-type: none"> Charge the watch twice a week Assist residents with daily homework exercise program Assist residents with daily calendar entry
AC (n=1)	Activities co-ordinator	Support	<ul style="list-style-type: none"> Support the weekly falls prevention program by chaperoning residents to and from the exercise room Assist the Fit for Life physiotherapists when needed
Physiotherapists (n=2)	N/A ^a	Service provider	<ul style="list-style-type: none"> Design and conduct a fall prevention program with residents over 6 weeks (included weekly exercise classes and a daily homework exercise program) Conduct mobility assessments pre- or postfall prevention program (ie, QTUG^b and static balance tests)

^aN/A: not applicable.

^bQTUG: Quantitative Timed Up and Go.

Residents were issued watches and shown how to use them to tell the time and count their steps. In terms of staff, the TC and SAs were provided with support manuals and trained to use the Covalence dashboard and how to access online support. HCAs were shown how to use the watch, complete a daily diary, and record the home work exercise program. Support manuals for

staff were reviewed by the research team for clarity and accessibility before being issued to staff.

It should be noted that compensation—in terms of time or payment—was not offered to staff or residents taking part in the study. Although the research team and industry collaborators (Kinesis, ICON, and BCA) provided support to the study site,

the trial predominantly relied upon existing care providers within the nursing home setting.

Qualitative Research

A range of qualitative methods were embedded throughout the study to capture the lived experience of trial participants. During the 8-week trial, regular visits (1-2 times per week) were made by the research team (SD, OK, and BR) to conduct research activities, conduct observations, and troubleshoot on technical issues. Researchers also maintained regular phone and email contact. An issues log was kept by the research team throughout the trial to document interactions with staff relating to device troubleshooting. All participants were provided with a diary (described as an “experience calendar”) to complete daily or as often as they wished. It featured a calendar with stickers they could assign to any particular day to reflect the feelings they associated with participation. The stickers were adapted from an original faces scale [25]. These diaries were intended to act as a prompt during the semistructured interviews at the end of the trial. Results showing the residents’ activity data (heart rate, sleep patterns, and step count) and static balance were reviewed during the interviews.

Participants were then asked to consider the value of the trial in the context of the burdens they identified. At the end of the trial, semistructured interviews were conducted with staff (n=8) and residents (n=6). Staff interviews were conducted with 4 HCAs, 1 AC, 2 SAs, and the TC. Despite attempts, in some cases severe cognitive impairment and dementia meant resulted in 3 residents being unable to engage in exit interviews.

A team of experienced qualitative researchers, including a research assistant (OK), research lead (BR), and postdoctoral research fellow (SD) with backgrounds in anthropology, gerontology, and sociology, respectively, conducted the interviews. A topic guide was developed that included two distinct sections to explore the perceived burden and value of participation. When discussing burden, diaries (experience calendars) were used as a prompt for reflecting on their experience and assisting with recall. Once burden was captured, the interviewer showed the participant data generated by the activity tracker (heart rate, sleep patterns, and step count) as well as results of the static balance tests. These were provided on hard copy print outs that were given to the residents to keep. The participant was finally asked to consider the value of the trial in the context of burdens they had identified.

Data Analysis

On average, interviews with staff lasted for 43 min, and interviews with residents lasted for 29 min. All interviews were digitally recorded and transcribed verbatim. Thematic analysis was conducted on these transcripts using NVivo 11 software package (QSR International Pty Ltd, Victoria, Australia) [28]. The analysis followed a largely grounded approach, with the exception of loosely imposed themes of “burden” and “value.” These themes provided an initial structure under which the interviews with residents were coded. First, by reviewing the resident interview transcripts, a comprehensive coding framework was generated. Analytical rigor was ensured using

interrater and intrarater reliability testing. Samples of transcripts were coded by SD and OK, after which they were compared, reviewed, and discussed to resolve coding issues. Miles and Huberman recommend minimum interrater reliability levels of 0.70 and minimum intrarater reliability levels of 0.80 [29]. Using this standard, reliability levels were in the acceptable range, at 0.73 and 0.81 respectively.

For the staff interview transcripts, the framework generated from the resident interviews was duplicated and expanded with additional codes that were relevant to the staff experience (eg, time emerged as a burden for staff but not residents). As part of this iterative process, the staff codes were then synthesized with the resident codes. Additionally, team members as well as 2 expert colleagues from the center for Applied Research for Connected Health were consulted to review and provide feedback.

Results

Thematic Analysis

The thematic analysis is summarized in Table 4 and includes the number of sources (people) who quoted the respective theme and the number of mentions (quotes). Among staff, we found that the following eight characteristics of burden emerged: (1) comprehension, (2) time, (3) communication, (4) emotional load, (5) cognitive load, (6) research engagement, (7) logistical burden, and (8) product accountability. The watch emerged as a consistent source of burden throughout the trial for HCAs who were responsible for charging it. Similarly, the Samsung mobile phone (hereafter referred to as the “phone”) was a source of burden for SAs responsible for syncing it with the watch to support data transfer. Among residents, we found only four sources of burden: (1) comprehension, (2) adherence, (3) emotional load, and (4) personal space. Comprehension and emotional load were shared burdens across the two groups.

Burden

Comprehension

Comprehension referred to the extent to which participants understood the research. Despite participants receiving detailed information about the study prior to providing informed consent, our results indicated that the trial was not consistently or fully understood by staff or residents at the nursing home. Among residents, there were various misinterpretations of the trial. One resident explained that she thought the daily exercises within the falls prevention program were being done “to keep your mind active” [Resident 5]. Other residents displayed an understanding of the purpose of the trial explaining that it was “preventing falls” [Resident 1]. However, the link between the watch and the intervention was unclear. When observing Resident 10 in conversation with another resident, she pointed to the watch and exclaimed, “How is a watch going to prevent falls?” [Resident 10]. However, some residents understood the basic function of the watch stating that “I understand [that it] was...counting steps...I think, to see how much we were walking. Is that it?” [Resident 9].

Table 4. Codes applied, number of participants quoting the respective topic, and number of quotations within each theme.

Key themes	Participants, n	Quotes, n
Thematic analysis of nursing home staff (n=8)		
Burden		
Comprehension	8	39
Time	7	57
Communication	7	43
Emotional load	7	34
Cognitive load	7	28
Research engagement	7	26
Logistical burden	6	20
Product accountability	5	7
Thematic analysis of nursing home residents (n=6)		
Burden		
Comprehension	6	33
Adherence	5	7
Emotional load	1	1

Residents' cognitive impairments were potentially a factor which limited their understanding and imposed additional burden. At times, residents struggled to retain information about the research. When asked about her experience of taking part in the 8-week trial, a resident with mild cognitive impairment reported that "I don't know because I don't know what this thing is all about." She also added, "What study? This study? Well I've only [learned] this for a few minutes, so I don't know anything about it" [Resident 10].

Similarly, staff expressed difficulty understanding what the research was trying to achieve and what the devices were for as well as their roles and responsibilities within the trial. A senior nurse explained how her understanding shifted over time: "In the beginning, I understood that it was with regard to falls in the nursing home, but then I said that it has [little] to do with falls, it's just for the technology to see how the technology works, I think"? [SA2]. HCAs' understanding of the trial revolved around the watch—the device with which they had the most interaction. HCA 1 explained that "I knew it was obviously about reading the steps and the movement of residents, that's what I understood [what] it was about. Just the kind of steps" [HCA1].

For staff, understanding their roles and responsibilities was also a burden. Staff in the dementia unit initially struggled with understanding how to use the experience calendar with their patients. HCA4 explains that although the study was explained to her:

I didn't fully understand until you and your colleague [the researchers] came the following day or whatever day, until you explained fully. I didn't get completely what I was supposed to do. Filling in their mood and stuff like that because they [dementia patients] couldn't reply, they can't tell me, you know that way? But, you then told me that I could answer for them. I

was thinking "ah, this is never going to work," if I was to ask them "are they OK with taking part?" because they have no understanding, you know? [HCA4]

The burden associated with comprehending the role and tasks of site administration was evident in the comments of SA1. She expressed her frustrations:

It was a nightmare (laughs), no it was okay. So, when I started off, first probably through my own fault, I probably didn't have enough of an insight as to what my role was. Kind of throughout the study I was kind of unsure what it was that I was meant to be doing. So, was I just meant to be overlooking the whole thing? Was it only meant to charge the watches? I was little bit unsure of what to do then with the [phone] in terms of how to sync it, but I asked questions, and I asked my colleague to show me. [SA1]

She added that she "felt a bit stupid sometimes because I wanted to give you guys [research team] an answer but because I wasn't sure [of the exact issue with the activity kits] myself...I felt a bit silly sometimes" [SA1]. Overall, the burden associated with comprehension—encompassing the trial, their roles, and the technology—was found to be the most pervasive aspect of burden in this study, as it emerged in some form in every interview.

Time

Staff reported lack of time to conduct research activities. Aside from the duration of tasks, time was constituted by many factors, for example disruption of workflow, interruption of tasks, and a lack of prioritization. Most notably, charging the watches was perceived as time-consuming. HCA5 explains that:

...It is time-consuming. By the time you go to the resident, get the watch off them, go up, plug it in...you get called on the way to do other things. And you have to put the residents' needs before charging a watch, you know? [HCA5]

HCA5 recognized that charging the watch was relatively quick but finding the time to do so was challenging: "Ah, it was easy enough to do [charge the watch]. It was just getting the time to do it. I know it sounds silly because it only takes like a minute or two." [HCA5] The issue of having insufficient time was inherently explained in terms of being interrupted (ie, called to do other things) and prioritizing needs of the residents above the needs of the trial.

Completion of the daily homework exercise program with residents was also perceived to be burdensome. These exercises were part of the fall prevention program. HCA5 explained her lack of time to complete these exercises with residents: "We just don't have time. We have other residents as well, upstairs we have 34 residents. And you just don't have time to do the exercises with them every day." [HCA5]

Overall, there was a lack of protected time for staff to complete trial-related tasks, and they simply had to add them to their current workload. When asked if she would take part in this type of research again, HCA4 replied:

...(Laughs) Not on top of my work, I wouldn't, no. But, if you're just doing like that, when you assign separate time to be able to...fully kind of focus on it, but no, I am just too busy in my work day. [HCA4]

Communication

TC and two SAs provided local technical support to the research team. During the trial, communication between the nursing home and research team was largely one-way. The staff did not contact the research team for technical support at any point. However, the research team frequently contacted SAs. Continuous technical support was required throughout the trial to ensure data were transferred from the activity tracking kits to a remote server. BCA provided onsite tech support on two occasions between weeks 1 and 2 of the trial, and the research team provided technical support between weeks 3 and 6. In that time, 37 issues were logged with SAs at the nursing home and it took, on average, 4 days to resolve each issue. The communication and time burdens were enormously demanding on staff.

The trial appeared to increase the burden associated with communication between staff internally. Staff explained that when they would forget to charge the watch, they would need to contact one of their colleagues. HCA1 explained that:

I called one of the night staff to just tell one of the girls upstairs because it was one of the nurses downstairs that picked up. I asked her to tell one of the girls [upstairs] to put [the resident's watch] in one of the chargers for me. The girls from the other team would know to take it out the next day, so it's fine. [HCA1]

In the interviews, staff reported difficulties with family members who did not understand the presence of the phone in the resident's rooms. For example, HCA1 recalls one incident:

A family member came in for [Resident1] and I don't know, well he didn't really know that it was going on, but I presume it wasn't him that they would have told about [the study]. So, he came in and was like 'oh, what is this phone?' as he thought it was a different phone. I said to him it was about the watch. It's a case study that's going on here and he was like "I didn't know about it." I was just like "oh my god, I'm so sorry about that" but somebody would have been told because the family has to know about it. So, it was just a bit of concern about the phone but I said it was for the watch and then that was about it, like there wasn't much of a big deal about it, but he didn't know about it. [HCA1]

Emotional Load

Emotional load is based on the concept of "emotional labor" [30]. The term captures feelings of shame, embarrassment, and stress are often related to noncompletion of research tasks. In relation to charging the watch, one HCA remarked, "I was afraid to forget about them." She explains, "...I [would] forget about the watches and I was always worried about [them]" [HCA6]. Another HCA reflected on the guilt she felt as a result of the watch not being charged and therefore not collecting data, "I felt bad because obviously it wasn't recording anything for them...So it's obviously going to have gaps here and there" [HCA4].

When discussing aspects of the trial that she deemed unsuccessful, TC attributed some personal blame:

The daily exercise thing was a massive flop, and I feel I am partly to blame for that. Because, um...I'd forgotten it. So, I know some were doing them and some weren't, but I didn't push on that. [TC]

She explained her perception of staff feelings about the trial:

I did feel that they [staff] felt...like they'd kind of let the study down...because they just didn't get around to doing [tasks]. I said look this is all learning, don't worry if you don't...get time to do it. Its fine, so don't sweat about it...But I know there was a bit of that as well, you know? [TC]

This sense of letting the study down was echoed by 1 resident. This resident had severe visual impairment and needed assistance of a staff member to fill out her experience calendar and complete her exercises. She explained how she was not inclined to interrupt the staff: "I hadn't the courage to say would you do this with me when they were coming in for just about 10 minutes you know? So, I didn't help you much there" [Resident 6]. This resident felt that she may have not been able to "help" data collection as a result.

Cognitive Load

Staff discussed the mental effort and burden associated with needing to remember to complete research tasks. This included charging the watch twice a week and assisting residents to

complete daily exercises and their experience calendar. In this context, charging the watch was reported as particularly challenging. HCA4 reflected on the least enjoyable part of the trial as charging the watch, explaining that she was frustrated with "...trying to remember it, then forgetting it, and then remembering it" [HCA4]. Similarly, HCA1 also reflected on the burden associated with forgetting to charge the watch as the least enjoyable part of her experience stating that "Charging it (laughs). Charging it. Not that I don't like [it], it's just that I never remember to do it" [HCA1].

HCA1 outlined the context in which she was trying to complete the task of charging the watch. She explained that:

There were a few times when a good few hours would have past because I wouldn't be getting them all at the same time to charge. It could be different, at the start [Resident 4, dementia patient], up until recently, would sleep until later. So, I wouldn't be able to get his watch until later in the day to try charging, and maybe someone else would be earlier in the morning. Trying to remember was the thing, when you are busy. [HCA1]

Research Engagement and Adherence

Although there was initial excitement and interest in the study, engaging in research activities over time became a burden for staff:

I didn't mind it at the start...as the weeks went on I was getting a bit "ughh," I have to keep doing it... then [I kept] forgetting the watches and then you're like "ughh," [if the watch is] obviously not on them...it's not going to be recording [data] for you. And, just toward the end, I was like "I'm over it now" (laughs). [HCA4]

Staff also perceived that residents disengaged with research activities over time and became nonadherent. HCA5 reflects that "At the start, for the first week or two, I did [Resident 5's experience calendar] with her, but after that... I think some of the resident's kind of lost interest in it, especially [Resident 5]" [HCA5]. Indeed, this resident reported that she had initially filled out her experience calendar but then stopped explaining, "I just never thought of it really" [Resident 5]. She reported that she also gradually forgot to complete the daily exercises, "I did it at first, with one of the girls [HCA]...After that, then I didn't do it anymore...I didn't think about it" [Resident 5]. Another resident explained that she did not engage with the experience calendar because she could not see the purpose of it:

I don't think I did [fill it out]. I kind of said, "what's this for." I pushed it aside and kind of said, "I think I'm jogging along reasonably well." It was something...I kind of pushed aside. I paid no attention to it. [Resident 12]

Another resident remarked that "I didn't do it...because I didn't know what I was doing" [Resident 1]. Residents who did perform the homework exercise program did not report any associated burden. However, a member of staff reported some discontent: "No they [the residents] never protest [about exercising], but there were sometimes they weren't happy.

They'd say 'oh not again,' but it depends on the day for everybody" [HCA6]. AC experienced some difficulty in motivating residents from the dementia unit to attend the weekly exercise class:

Some residents didn't attend [the exercise class] particularly people with dementia. So one time [Resident 14] ...and I don't know if he remembered...when I said "we go for exercises," he didn't want to go, but I think it's kind of dementia you know. [AC]

Logistical Burden

The nursing home contained 4 wards spread over 2 floors. Mobile phones and chargers for watches were set up in the participating resident's bedrooms. Residents were recruited from every ward; therefore, their bedrooms and the respective devices and equipment were dispersed across every ward. It appears that this presented a logistical burden for SAs who needed to move among different locations within the nursing home to check and sync devices. This was regarded as highly inconvenient:

You need to go take the watch to the phone, take the phone to the watch, person to the phone...when it wasn't syncing. It took a long time to get through them. And it was never just one, there was always more than one, you know if it was one, it takes 5 minutes, but...there was upstairs, downstairs. Residents are here, residents are in the dining room. [TC]

Each watch was identical, but had a unique identifying code (UIC) placed on its underside. As is standard practice in research, for data protection purposes staff, did not have the key between the UIC and the names of the residents. However, this created a burden for staff when attempting to collect and charge multiple watches at once. HCA6 explains "I never took all of [the watches] together. I only took them one-by-one...I never wanted to mix them up" [HCA6]. She explains:

I think one time I took two...and I said left (hand) is [Resident 9] and the right one is [Resident 10] (laughs). Or I look at the number on the code and say okay, this is (910) or (374). So, (910) is [Resident 10] and (374) is [Resident 9]. [HCA6]

Product Accountability

There was a burden associated with responsibility to keep the devices safe and secure so that they were not damaged or broken. Staff appeared to be conscious of the cost of the devices. In their supervisory role as SAs, senior nurses took it upon themselves to ensure that the activity trackers issued to residents throughout the nursing home were kept safe. SA1 was conscious that these products were not their property and was apologetic in instances where they were damaged:

In terms of the residents, I would always be keeping an eye; there were two watches, as you know, that were a little bit damaged by the residents for whatever reasons. I did say to [the researcher] that they were damaged and I apologized for that but, in general,

we, the staff, would make sure that they were being looked after and weren't thrown around. So, yeah we would be conscious as they weren't our property. [SA1]

The phones were often placed on the floor under the resident's nightstand to keep them out of the way. SA2 reported awareness that the products may become damaged by routine cleaning:

I was more aware of the cleaning staff because some of the phones were on the floor, so I was more aware not to wet them or not to damage them in that way. But, the HCA were very careful and I knew they were going to take care of it, just cleaning staff more so than anyone else. I made sure they were on the skirting boards and not on the actual floor. [SA2]

One staff member reported an incident where a phone went missing:

One day, one of the families unplugged the whole phone and everything because they thought it was a staff member charging their phone. Even though they had been told about the study, they just didn't want them charged in the room. Then, I had to go, plug it all back in, and make sure it was working. [HCA5]

This HCA described this encounter with the family member as "awkward." Locating the device was challenging and stressful for HCA: "I had to go and find it because her daughter had put it away, and then [the resident] had moved it...at first, I thought part of it was missing and I was like 'Oh God, where did it go?'" At this point, she described an emotional burden related to her responsibility and accountability for the device. She reported feeling panicked and thinking "God, if it is gone, what are we gonna do?" [HCA5]. However, the phone was ultimately found and the presence of the device was re-explained to the family member. HCA felt this resolved the issue.

Personal Space

Conducting research in a residential setting has the potential of reduce privacy for residents in their home. During the 8 weeks of the trial, their phones were frequently checked by SAs and the researcher in the resident's rooms. Sometimes, this happened when the resident was not in his or her room and may have created unease if the resident thought someone had been in their personal space. One resident explained:

Funny I thought that someone was mooching around my clothes, affairs, and things like that now...don't know whether I lost anything or not...I thought that someone was coming into the place. [Resident 12]

Relationship Between Burden and Value

While participants reported to find value in data, this was found to be somewhat superficial. For example, one member of staff reported that "I just find it interesting, I don't know [why]" [HCA6]. Value was largely expressed in terms of its absence rather than its presence. Staff struggled to understand how tracking heart rate, step count, and sleep patterns could be useful to them in their everyday life and how it related to preventing falls. One staff member explained "I think it's useful (dubious tone). Yeah, no I do, but I would have to look at it a little bit

more just to see how this can help us to further prevent falls" [SA1]. Similarly, abstract relationship between the watch and falls preventions was a barrier for residents. When the interviewer explained that the watch was counting steps, one resident exclaimed, "What the hell difference does that make? Sorry for my language" [Resident 10].

TC felt they captured much of this data informally anyway and would communicate it in handover meetings between day and night shifts:

Generally, our communication systems around our hand over are that...those type of things are our main highlighters. In terms of a hand over, [it] would be, "Did they sleep well?" "Did they have any falls today?" "Did they eat?" "Did they eat well?" "Are they drinking well?" All that type of thing is basic. [So we are] doing that anyway. [TC]

The perceived value of data may have been comprised by skepticism over the integrity of data collection. A staff member reflects:

Well I don't think you got the full information that you wanted. Did you get the full information that you wanted? I'm still not sure if I helped you enough to have your study fully, fully completed. [SA2]

TC explained, "I don't know the relevance of it. Is it going to be, will it be accurate? You know, in terms of feedback, just because there are so many gaps and holes in it" [TC]. Overall, analysis suggested that the value of data appeared to be limited for participants, relative to the burden they encountered.

Discussion

Principal Findings

Mobile and wearable technologies offer the potential to conduct clinical trials remotely in a nonclinical setting, thereby reducing the associated burden on patients. However, this study suggests that the burden associated with the remote trial was not modulated by the value of participation. Staff burden emerged to be multifaceted and interrelated in practice. For example, a cognitive burden (ie, forgetting to charge the watch), may have prompted an emotional burden (ie, anxiety about the battery dying) and posed a logistical burden (ie, phoning a colleague on the ground floor to instruct a colleague on the first floor to charge the watch). Thus, burden should be considered as multifaceted in practice.

Comprehension emerged as the most commonly shared burden among residents and staff in the trial. Inability to comprehend the purpose of the trial—particularly the relationship between the intervention and wearable device—was a substantial challenge and may have contributed to a lack of motivation and engagement to participate in trial activities. Thus, the need for participants to understand and "buy in" to the concept of the study is arguably paramount [31]. Indeed, there is a large amount of evidence to suggest that comprehension can pose a barrier to clinical trial participation and recruitment [32-34]. Our findings suggest this is also relevant in a remote trial where the

relationship between the intervention and devices used for supporting the intervention have the potential to be unclear.

Staff Experiences

A lack of adherence among staff is a recognized barrier to research in nursing homes [35]. Nonadherence among staff in this study may have been related to the experience of burden. Thus, it is useful to further analyze adherence to better understand how it is constituted in this setting.

Communication and Feedback

Communication was the most frequently cited burden. Although the research team made regular site visits, the informal nature of capturing feedback was not sufficient. The participatory approach enabling feedback and dialog between the staff and research team was front-loaded into the study design. However, communication may have been improved by formalizing a feedback loop with staff, residents, and their families throughout the trial. This could have enabled the research team to accurately capture and respond to emerging issues, thereby improving the experience for all stakeholders. For example, having a staff member assigned to the collection and charging of devices in the evening and another assigned to the return of devices each morning may have been more effective and may have created less logistical concerns. This would require careful labeling of devices to ensure consistency of use.

Workflow

Although the study protocol was developed with nursing home staff input, embedding tasks and activities into the work plans of staff might have helped in alleviating some communication and role or responsibility issues. Ensuring that study procedures are integrated into daily activities and task list of the care staff is important. Remuneration for the nursing home staff, as would be the case in a typical clinical trial, may have helped in achieving this. While the protocol was developed with staff input, protected time for TC and senior nursing staff to embed tasks and activities into work schedules might have helped in alleviating some communication and role or responsibility issues. Similarly, protected time for staff to execute these activities might also have reduced burden. Although efforts were made to build a collaborative design, there are known structural and cultural issues around the execution of participatory approaches that are not unique to nursing home settings [36-38]. Arguably, a more comprehensive, rather than a piecemeal approach to participatory research is required.

The job of integrating research activities into the routine workflow of the nursing home was the responsibility of TC. Although her knowledge of staff shifts and vast experience in the nursing home meant that she was best placed to oversee this, a lack of protected time to devote to her role and to train and work with the research team was challenging. Her responsibilities also shifted over time. For example, it was intended that TC would be responsible for recruitment of residents. However, in practice, managing this task alongside her senior role in the nursing home was not feasible. It is noteworthy that AC was the only staff member who did not identify time as a burden. She was not assigned with any tasks that fell outside of her typical workflow (ie, assisting completion

of the daily exercise program or experience calendar with residents or with charging the watch).

Exposure to Value

In this study, we explored the burden associated with a remote trial, and this study was not designed to explicitly measure values. Therefore, there were limited opportunities for participants to become exposed to data. It was assumed that there would be inherent value in participation in the form of resident's receiving data from the watch and that this would provide insight to staff. However, a report of the full clinical data (ie, step count, sleep patterns, and heart rate) was only shared with participants at the *end* of the study. Thus, staff and residents only had the opportunity to become exposed to the value of the data at the end of the trial, once the burden associated with participation was over. The only opportunity for feedback of data during the trial was to the residents in the form of the step count captured by the watch. Yet with the exception of one resident, it appears that the residents generally did not interact with the watch. Arguably, the conception of the trial was overly abstract and contributed to a lack of comprehension of the trial and study. The lack of exposure to value is problematic in that it is therefore unable to modulate the perceived burden and may indeed inflate the perception of burden.

Connectivity

For staff, burden was likely amplified by persistent problems with connectivity at the research site. This resulted in the research team frequently phoning and emailing the SAs to conduct localized troubleshooting. Arguably, this increased burden and restricted the potential value of the trial. This may have been compounded by a lack of protected time for the staff at the nursing home to conduct research activities and troubleshooting.

Residents' Experiences

Resident burden was found to be relatively limited. Where it was observed, it was mainly related to comprehension and adherence, with some evidence of burden relating to emotional load and personal space. There also appeared to be interdependency between comprehension and adherence, whereby a resident may not have understood why he or she needed to complete a research activity. Therefore, many residents did not engage with the research. Data security did not emerge as an issue in interviews with residents. This issue was also not raised by staff or families during the trial.

Engagement

It was assumed that the watch, which was a consumer device, would act as an engaging tool for residents (and staff), potentially acting as a talking point and generating competition among users. However, it was observed by staff and the research team that only one resident engaged with the watch as a step counter. Among this cohort, there is no evidence to suggest that inclusion of mobile and wearable technology improved participation, engagement, retention, or adherence. However, convenience of the remote trial did allow for the inclusion of typically hard-to-reach individuals, such as patients with dementia.

Limitations

Specificity of Site and Devices

The findings of this study refer to a remote trial conducted in a specific nursing home setting using specific health technology devices. This research was conducted at a single site; therefore, the extent to which unique aspects of this one setting—within this one organizational culture—contributed to the specific findings is unknown. The burden associated with technology for health care staff, however, is not necessarily unique to this setting. The problem of “real-world” deployment is recognized within the literature, particularly in terms of workflow integration and conceptualization [39,40]. Indeed, the evidence base to support clinical claims of eHealth technology has been debated, and it has been suggested that costs associated with commonly deployed eHealth technologies (ie, time and infrastructure) may in fact decrease organizational efficiency [39].

Thus, before we can overcome these burdens, qualitative exploratory studies—of the kind presented here—are crucial for understanding how and why burden manifests. The devices used in this study were chosen because they were widely available to the research team, which had worked with this technology in previous studies. It is possible that the choice of devices may have contributed to some of the burdens identified in this study, although the extent of this burden is unknown. Although comparing a range of devices and their limitations was beyond the scope of this study, this should be considered in future research.

Information Overload

In terms of comprehension, it is possible that in our efforts to be transparent and didactic during recruitment, we risked overloading the residents with information. Although there is no evidence to suggest that this interfered with their willingness to enroll in the study or their ability to retain information over the 8-week period, effectiveness of teach-back methods for information retention over time (ie, a clinical trial period) requires further study [41-43].

Opportunities for Feedback

As part of a participatory approach, focus group studies were conducted with staff prior to the trial being designed. However, this was not continued throughout the trial. Formalizing a feedback loop may have improved communication and eased the burden associated with comprehension. Although regular visits were made by the research team to the nursing home, whereby feedback was captured informally, more structured follow-ups with staff and residents may have enabled the research team to determine how much information was actually being retained, address issue or gaps in knowledge in a timely manner, and reinforce the purpose of the study.

Staff Training and Protected Time

This study involved leveraging the resource of research-naïve health care professionals. Increasing training resources available

at the start of and during the study may have been useful in alleviating some of the issues identified. In particular, comprehensive face-to-face training, which specifically addresses aims, objectives, and importance of the study, should be conducted for all study staff. Indeed, the need for more formalized staff training around the implementation of connected health technologies has been found outside of the nursing home setting, for example in community settings [38].

Similar to that observed with many clinical trials, finding approaches for effectively training personnel who are not able to attend face-to-face trainings because of operational commitments, shifts, or working patterns is an important consideration. A possible solution could be online or app-based study guides to provide more convenient and immediate methods to access training and troubleshooting materials. Progress feedback reports during the study identifying progress to date and detailing issues encountered and how to resolve them would provide an additional method to re-iterate study objectives and importance. Again, formalized meetings with staff should also be in place to support dialog between staff and researchers.

Staff were not remunerated for the additional work required by the study. Remuneration could have provided staff with protected time to engage in training and meetings. This in turn, may have improved communication during the trial. It is also possible that remuneration for the nursing home staff may have provided additional motivation.

In addition to the operation of the study, training should include defining roles and responsibilities. Clearer definition and communication of the roles and responsibilities of staff may have reduced the burden associated with comprehension. Again, this would have been better facilitated by ensuring protected time for staff.

Conclusions

The paper identifies and characterizes perceived burden associated with a remote trial in a particularly challenging patient population. It identifies eight aspects of staff burden and four aspects of resident burden that could be further explored and developed. The exploratory nature of this research meant that there were many unknowns before entering the research site. From the outset, we assumed that the staff burden associated with a remote trial would be related to increased workload and the time consuming nature of co-ordination. Although these burdens were indeed observed in the staff experience, their nuance, variation, and extent were relatively unexplored until now. This research, therefore, offers a novel understanding of the nature of staff burden in a remote trial and underlines the importance of the relationship between burden and value. The potential of remote clinical trials requires further examination to optimize and enhance the methodology. This study suggests the convenience of a remote trial that can aid inclusion of hard-to-reach patient groups; however, we need to comprehensively measure and minimize the associated burden on relevant stakeholders.

Acknowledgments

The study was funded by Applied Research for Connected Health, an Enterprise, Ireland, and IDA, Ireland, funded technology center led by the University College Dublin. The authors would like to acknowledge the collaboration of Dr Barry Greene at Kinesis Health Technologies and Dre Goheen, J Patrick Bewley, and Rob Cabral at Big Cloud Analytics. We are grateful for the contribution of Sam Fiens, Dr Marcella McGovern, Dr Shane O'Donnell, Dr Anushree Priyadarshini, Dr Maria Quinlan, and Dr Louise Rooney. We would like to thank Mark Sweeney and Ana Paunovic at Fit for Life. Finally, we express our gratitude to the staff, residents, and families at the nursing home for their support and participation in this research.

Authors' Contributions

SD, BR, and OK were involved in all stages of the project, including design, data collection, and analysis. MM, BB, WM, and BC were involved in setting research objectives and providing feedback throughout the project. All authors participated in reading and commenting on the manuscript.

Conflicts of Interest

None declared.

References

- Galvin JE, Meuser TM, Boise L, Connell CM. Predictors of physician referral for patient recruitment to Alzheimer disease clinical trials. *Alzheimer Dis Assoc Disord* 2009 Dec;23(4):352-356 [[FREE Full text](#)] [doi: [10.1097/WAD.0b013e31819e0cac](https://doi.org/10.1097/WAD.0b013e31819e0cac)] [Medline: [19561438](#)]
- Schneider LS, Olin JT, Lyness SA, Chui HC. Eligibility of Alzheimer's disease clinic patients for clinical trials. *J Am Geriatr Soc* 1997 Aug;45(8):923-928. [Medline: [9256842](#)]
- Shenoy P, Harugeri A. Elderly patients' participation in clinical trials. *Perspect Clin Res* 2015 Dec;6(4):184-189 [[FREE Full text](#)] [doi: [10.4103/2229-3485.167099](https://doi.org/10.4103/2229-3485.167099)] [Medline: [26623388](#)]
- Treves TA, Verchovsky R, Klimovitsky S, Korczyn AD. Recruitment rate to drug trials for dementia of the Alzheimer type. *Alzheimer Dis Assoc Disord* 2000 Dec;14(4):209-211. [Medline: [11186598](#)]
- Watson JL, Ryan L, Silverberg N, Cahan V, Bernard MA. Obstacles and opportunities in Alzheimer's clinical trial recruitment. *Health Aff (Millwood)* 2014 Apr;33(4):574-579 [[FREE Full text](#)] [doi: [10.1377/hlthaff.2013.1314](https://doi.org/10.1377/hlthaff.2013.1314)] [Medline: [24711317](#)]
- Watts G. Why the exclusion of older people from clinical research must stop. *BMJ* 2012 May 21;344:e3445. [Medline: [22613873](#)]
- Herrera AP, Snipes SA, King DW, Torres-Vigil I, Goldberg DS, Weinberg AD. Disparate inclusion of older adults in clinical trials: priorities and opportunities for policy and practice change. *Am J Public Health* 2010 Apr 01;100 Suppl 1:S105-S112. [doi: [10.2105/AJPH.2009.162982](https://doi.org/10.2105/AJPH.2009.162982)] [Medline: [20147682](#)]
- Marcantonio ER, Aneja J, Jones RN, Alsop DC, Fong TG, Crosby GJ, et al. Maximizing clinical research participation in vulnerable older persons: identification of barriers and motivators. *J Am Geriatr Soc* 2008 Aug;56(8):1522-1527 [[FREE Full text](#)] [doi: [10.1111/j.1532-5415.2008.01829.x](https://doi.org/10.1111/j.1532-5415.2008.01829.x)] [Medline: [18662204](#)]
- Townsend CA, Selby R, Siu LL. Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials. *J Clin Oncol* 2005 May 1;23(13):3112-3124. [doi: [10.1200/JCO.2005.00.141](https://doi.org/10.1200/JCO.2005.00.141)] [Medline: [15860871](#)]
- Eardley A, Cribb A, Pendleton L. Ethical issues in psychosocial research among patients with cancer. *Eur J Cancer* 1991;27(2):166-169. [Medline: [1827282](#)]
- Kim M, Dahlberg A, Hagell P. Respondent burden and patient-perceived validity of the PDQ-39. *Acta Neurol Scand* 2006 Feb;113(2):132-137. [doi: [10.1111/j.1600-0404.2005.00549.x](https://doi.org/10.1111/j.1600-0404.2005.00549.x)] [Medline: [16411975](#)]
- Sharp L, Frankel J. Respondent burden: A test of some common assumptions. *Public Opin Q* 1983 Jan 1983;47(1):A [[FREE Full text](#)] [doi: [10.1086/268765](https://doi.org/10.1086/268765)]
- Ulrich CM, Wallen GR, Feister A, Grady C. Respondent burden in clinical research: when are we asking too much of subjects? *IRB* 2005;27(4):17-20. [Medline: [16220630](#)]
- Lingler JH, Schmidt K, Gentry A, Hu L, Terhorst L. A New Measure of Research Participant Burden: Brief Report. *J Empir Res Hum Res Ethics* 2014 Oct;9(4):46-49 [[FREE Full text](#)] [doi: [10.1177/1556264614545037](https://doi.org/10.1177/1556264614545037)] [Medline: [26125079](#)]
- Ulrich CM, Knafl KA, Ratcliffe SJ, Richmond TS, Grady C, Miller-Davis C, et al. Developing a Model of the Benefits and Burdens of Research Participation in Cancer Clinical Trials. *AJOB Prim Res* 2012;3(2):10-23 [[FREE Full text](#)] [doi: [10.1080/21507716.2011.653472](https://doi.org/10.1080/21507716.2011.653472)] [Medline: [24748992](#)]
- Covington D, Velez K. The remote patient-centered approach in clinical research. *Applied Clinical Trials* 2015;24(2/3):30-34 [[FREE Full text](#)]
- Bates M. New Trends in Clinical Trials: A Conversation with Pfizer's Craig Lipset. *IEEE Pulse* 2016;7(2):34-36. [doi: [10.1109/MPUL.2015.2513725](https://doi.org/10.1109/MPUL.2015.2513725)] [Medline: [26978850](#)]
- Dorsey ER, Venuto C, Venkataraman V, Harris DA, Kiebert K. Novel methods and technologies for 21st-century clinical trials: a review. *JAMA Neurol* 2015 May;72(5):582-588 [[FREE Full text](#)] [doi: [10.1001/jamaneurol.2014.4524](https://doi.org/10.1001/jamaneurol.2014.4524)] [Medline: [25730665](#)]

19. McAlindon T, Formica M, Kabbara K, LaValley M, Lehmer M. Conducting clinical trials over the internet: feasibility study. *BMJ* 2003 Aug 30;327(7413):484-487 [[FREE Full text](#)] [doi: [10.1136/bmj.327.7413.484](https://doi.org/10.1136/bmj.327.7413.484)] [Medline: [12946971](#)]
20. Orri M, Lipset CH, Jacobs BP, Costello AJ, Cummings SR. Web-based trial to evaluate the efficacy and safety of tolterodine ER 4 mg in participants with overactive bladder: REMOTE trial. *Contemp Clin Trials* 2014 Jul;38(2):190-197. [doi: [10.1016/j.cct.2014.04.009](https://doi.org/10.1016/j.cct.2014.04.009)] [Medline: [24792229](#)]
21. Rosa C, Campbell ANC, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. *Contemp Clin Trials* 2015 Nov;45(Pt A):41-54. [doi: [10.1016/j.cct.2015.07.007](https://doi.org/10.1016/j.cct.2015.07.007)] [Medline: [26176884](#)]
22. Gossen R. Is Pfizer's virtual clinical trial model a failure? *Rebar Interactive* 2012 Mar:n.p-n.p [[FREE Full text](#)]
23. Koberstein W. Pfizer perseveres in pioneering virtual clinical trials. *Clinical Leader* 2012:n.p-n.p [[FREE Full text](#)]
24. Dantic D. A critical review of the effectiveness of 'teach-back' technique in teaching COPD patients self-management using respiratory inhalers. *Health Education Journal* 2013 Jan 25;73(1):41-50. [doi: [10.1177/0017896912469575](https://doi.org/10.1177/0017896912469575)]
25. Doak C, Doak L, Root J. Teaching patients with low literacy skills. *AJN. Am J Nurs* 1996;96(12):16.
26. Schillinger D, Piette J, Grumbach K, Wang F, Wilson C, Daher C, et al. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med* 2003 Jan 13;163(1):83-90. [Medline: [12523921](#)]
27. Sudore RL, Schillinger D. Interventions to Improve Care for Patients with Limited Health Literacy. *J Clin Outcomes Manag* 2009 Jan 01;16(1):20-29 [[FREE Full text](#)] [Medline: [20046798](#)]
28. NVivo QAS. NVivo [computer program]. QSR International Pty Ltd, Version 11 2002.
29. Miles M, Huberman M. *Qualitative Data Analysis*. London: SAGE; 1994.
30. Hochschild A. *The managed heart: commercialization of human feeling*. Berkeley: University of California Press; 1983.
31. Shaw RJ, Kaufman MA, Bosworth HB, Weiner BJ, Zullig LL, Lee SD, et al. Organizational factors associated with readiness to implement and translate a primary care based telemedicine behavioral program to improve blood pressure control: the HTN-IMPROVE study. *Implement Sci* 2013 Sep 08;8:106 [[FREE Full text](#)] [doi: [10.1186/1748-5908-8-106](https://doi.org/10.1186/1748-5908-8-106)] [Medline: [24010683](#)]
32. Barrett R. Quality of informed consent: measuring understanding among participants in oncology clinical trials. *Oncol Nurs Forum* 2005 Jul 01;32(4):751-755. [doi: [10.1188/05.ONF.751-755](https://doi.org/10.1188/05.ONF.751-755)] [Medline: [15990904](#)]
33. Bergenmar M, Molin C, Wilking N, Brandberg Y. Knowledge and understanding among cancer patients consenting to participate in clinical trials. *Eur J Cancer* 2008 Nov;44(17):2627-2633. [doi: [10.1016/j.ejca.2008.08.013](https://doi.org/10.1016/j.ejca.2008.08.013)] [Medline: [18818068](#)]
34. Stryker JE, Wray RJ, Emmons KM, Winer E, Demetri G. Understanding the decisions of cancer clinical trial participants to enter research studies: factors associated with informed consent, patient satisfaction, and decisional regret. *Patient Educ Couns* 2006 Oct;63(1-2):104-109. [doi: [10.1016/j.pec.2005.09.006](https://doi.org/10.1016/j.pec.2005.09.006)] [Medline: [16242898](#)]
35. Beck AM, Damkjaer K, Tetens I. Lack of compliance of staff in an intervention study with focus on nutrition, exercise and oral care among old (65+ yrs) Danish nursing home residents. *Aging Clin Exp Res* 2009 Apr;21(2):143-149. [Medline: [19448386](#)]
36. Ahmed SM, Beck B, Maurana CA, Newton G. Overcoming barriers to effective community-based participatory research in US medical schools. *Educ Health (Abingdon)* 2004 Jul;17(2):141-151. [doi: [10.1080/13576280410001710969](https://doi.org/10.1080/13576280410001710969)] [Medline: [15763757](#)]
37. Casswell S. A decade of community action research. *Subst Use Misuse* 2000 Jan;35(1-2):55-74. [Medline: [10677875](#)]
38. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Implementing a telehealth service: nurses' perceptions and experiences. *Br J Nurs* 2014;23(21):1133-1137. [doi: [10.12968/bjon.2014.23.21.1133](https://doi.org/10.12968/bjon.2014.23.21.1133)] [Medline: [25426527](#)]
39. Black A, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med* 2011 Jan 18;8(1):e1000387 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1000387](https://doi.org/10.1371/journal.pmed.1000387)] [Medline: [21267058](#)]
40. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med* 2011 Jan 18;8(1):e1000387 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1000387](https://doi.org/10.1371/journal.pmed.1000387)] [Medline: [21267058](#)]
41. Griffey R, Shin N, Jones S, Aginam N, Gross M, Kinsella Y, et al. The impact of teach-back on comprehension of discharge instructions and satisfaction among emergency patients with limited health literacy: A randomized, controlled study. *J Commun Healthc* 2015 Mar;8(1):10-21 [[FREE Full text](#)] [doi: [10.1179/1753807615Y.0000000001](https://doi.org/10.1179/1753807615Y.0000000001)] [Medline: [26617669](#)]
42. Kandula NR, Malli T, Zei CP, Larsen E, Baker DW. Literacy and retention of information after a multimedia diabetes education program and teach-back. *J Health Commun* 2011;16 Suppl 3:89-102. [doi: [10.1080/10810730.2011.604382](https://doi.org/10.1080/10810730.2011.604382)] [Medline: [21951245](#)]
43. Ha DT, Bonner A, Clark R, Ramsbotham J, Hines S. The effectiveness of the teach-back method on adherence and self-management in health education for people with chronic disease: a systematic review. *JBHI Database System Rev Implement Rep* 2016 Jan;14(1):210-247. [doi: [10.11124/jbisrir-2016-2296](https://doi.org/10.11124/jbisrir-2016-2296)] [Medline: [26878928](#)]

Abbreviations

AC: activities co-ordinator

BCA: Big Cloud Analytics, Inc
HCA: health care assistant
MMSE: Mini-Mental State Examination
QTUG: Quantitative Timed Up and Go
SA: site administrator
TC: trial co-ordinator

Edited by G Eysenbach; submitted 12.12.17; peer-reviewed by A Campbell, C Lipset, E Dove; comments to author 18.01.18; revised version received 14.03.18; accepted 14.04.18; published 19.06.18.

Please cite as:

Donnelly S, Reginatto B, Kearns O, Mc Carthy M, Byrom B, Muehlhausen W, Caulfield B
The Burden of a Remote Trial in a Nursing Home Setting: Qualitative Study
J Med Internet Res 2018;20(6):e220
URL: <http://www.jmir.org/2018/6/e220/>
doi: [10.2196/jmir.9638](https://doi.org/10.2196/jmir.9638)
PMID: [29921563](https://pubmed.ncbi.nlm.nih.gov/29921563/)

©Susie Donnelly, Brenda Reginatto, Oisin Kearns, Marie Mc Carthy, Bill Byrom, Willie Muehlhausen, Brian Caulfield. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 19.06.2018. 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.

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

<https://www.jmirpublications.com/>